HAL
archives-ouvertes.fr

La décision de l'expérimentation à l'interprétation : l'apport de Donald Davidson

Pôl-Vincent Harnay

▶ **To cite this version:**

Pôl-Vincent Harnay. La décision de l'expérimentation à l'interprétation : l'apport de Donald Davidson. Économies et finances. Université Panthéon-Sorbonne - Paris I, 2008. Français. <tel-00363905>

HAL Id: tel-00363905

https://tel.archives-ouvertes.fr/tel-00363905

Submitted on 24 Feb 2009

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

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

UNIVERSITE PARIS I – PANTHEON – SORBONNE
SCIENCES ECONOMIQUES – SCIENCES HUMAINES – SCIENCES JURIDIQUES ET POLITIQUES

LA DÉCISION, DE L'EXPÉRIMENTATION À L'INTERPRÉTATION : L'APPORT DE DONALD DAVIDSON

Thèse pour le Doctorat en Sciences Economiques
(arrêté du 30 mars 1992)

Présentée et soutenue publiquement par
Pôl-Vincent Harnay

Directeur de recherche :
André Lapidus

Membres du jury :

Damien Besancenot, Professeur à l'Université de Paris Nord (rapporteur)
Pascal Engel, Professeur à l'Université de Genève
Pierre Garrouste, Professeur à l'Université de Lyon 2 (rapporteur)
André Lapidus, Professeur à l'Université Paris I Panthéon-Sorbonne
Louis Lévy-Garboua, Professeur à l'Université Paris I Panthéon-Sorbonne
Christian Schmidt, Professeur à l'Université Paris Dauphine.

8 décembre 2008

L’UNIVERSITE PARIS I – PANTHEON-SORBONNE n’entend donner aucune approbation ni improbation aux opinions émises dans cette thèse ; ces opinions doivent être considérées comme propres à leur auteur.

Remerciements

Je tiens, tout d'abord, à remercier mon directeur de thèse, André Lapidus et lui témoigner toute ma gratitude pour avoir cru en moi tout au long de ces quatre années de thèse. Ses conseils et remarques m'ont permis de mener à bien cette thèse.

Je tiens aussi à remercier toutes les personnes qui m'ont soutenu et aidé au sein du laboratoire PHARE ainsi qu'au CEPN. Mes remerciements vont aussi à Marc-Arthur Diaye pour sa disponibilité et ses conseils, Bernard Guerrien pour sa pédagogie et son soutien ainsi Jean-Yves Jaffray pour m'avoir permis d'avoir accès à des textes et ouvrages très utiles.

Ma rencontre avec Marcia Cavell fut pour moi inoubliable et je la remercie pour l'entretien qu'elle a bien voulu m'accorder, sa disponibilité et pour m'avoir permis d'avoir accès aux archives de Davidson.

Je voudrais aussi remercier mes amis proches qui m'ont soutenu et supporté : Williams Macias et Thibault Lafarie.

Je remercie mes parents et mes grands-parents ainsi que tous les membres de ma famille qui, de près ou de loin, m'ont apporté leur aide.

Enfin, je tiens à remercier Pétronille, ma femme, pour être un ange de patience et pour m'avoir permis de réaliser mon rêve.

Introduction générale

Donald Davidson (1917-2003) est connu à travers le monde, depuis les années 1960-1970 pour ses travaux en philosophie de l'action et du langage[1] qui s'inscrivent dans la tradition analytique[2]. Pourtant, il est aussi un auteur majeur de la théorie de la décision et, plus précisément, de l'économie expérimentale de la décision telle qu'elle s'est développée, aux Etats-Unis dans les années 1950. La séparation historique entre ses travaux de philosophe et ses travaux d'économiste ne reflète pas, toutefois, l'interconnexion entre ces deux domaines que Davidson a défendue tout au long de son parcours et, tout particulièrement, au cours des années 1980 lorsqu'il a tenté de les unir, au sein d'une même théorie.

Le point de départ du Davidson économiste pourrait être décrit dans des termes familiers du point de vue de la théorie de la décision d'aujourd'hui. C'est un paradigme désirs-croyances, où les désirs s'exprimeraient au travers d'utilités et les croyances par des probabilités. Lorsque Davidson aborde cette question au milieu des années 1950, elle a déjà pris la forme de la théorie de l'utilité espérée, dans ses deux versions majeures, celle de von Neumann et Morgenstern [1944, 1947] (deuxième édition où se trouve l'appendice sur l'axiomatisation de l'utilité espérée) où les probabilités sont objectives, et celle de Savage [1954] où les probabilités sont subjectives – en ce sens qu'elles émergent, en même temps que la fonction d'utilité, des préférences que les agents construisent sur leurs actes, c'est-à-dire des relations qu'ils établissent entre l'espace des états du monde et celui des conséquences.

Au commencement, comme on le verra, la contribution de Davidson peut se comprendre comme une tentative de donner une densité expérimentale à cette approche qui n'était pas sérieusement remise en cause. Si l'on revient sur ses premiers travaux des années 1950 à Stanford, on y rencontre un Davidson engagé dans ce qui allait apparaître plus tard comme la genèse de l'expérimentation

---

[1] Comme en témoignent par exemple, les deux ouvrages édités par Ernest Lepore, *Actions and Events, Perspectives on the philosophy of Donald Davidson* [1985], et *Truth and Interpretation, Perspectives on the philosophy of Donald Davidson* [1986] ou encore les ouvrages qui lui sont consacrés en Europe, comme celui de Pascal Engel [1994], *Davidson et la philosophie du langage*.

[2] Parmi les auteurs majeurs de cette tradition, on peut citer par exemple Frege (1848-1925) et Russell (1872-1970).

économique sur des problèmes de décision. L'auteur évoque non seulement des « effets de mémoire » et « d'apprentissage » mais aussi « des effets de présentation ». Ces effets du comportement de choix des sujets de l'expérience peuvent conduire, explique Davidson, à des résultats qui ne sont pas conformes à ce que la théorie de l'utilité espérée prévoit. Ces effets procéderaient au moins partiellement de l'impasse sur les significations dans la théorie de la décision standard. La spécificité de l'approche qui progressivement voit le jour chez Davidson et qui semble l'éloigner du territoire habituel des économistes, alors même que leurs préoccupations ne le quittent jamais complètement, c'est qu'il va essayer de mettre à jour ce à quoi le désir et la croyance font l'un et l'autre appel : la signification. Or, le rôle des significations découle de ce que les choix s'expriment en général verbalement. Dès lors que l'expérimentateur propose, comme ce fut souvent le cas dans les années 1950, une solution behavioriste au problème de la mesure des utilités et des probabilités, il impose un langage. Ce langage se retrouve dans la formulation des options sur lesquelles portent les choix. Si bien qu'un décalage peut s'installer entre les significations que les sujets accordent à ces options et à leurs résultats et les significations que leur attribue l'expérimentateur lorsqu'il élabore un protocole expérimental.

La critique est féconde : l'analyse des significations offrirait une donnée mentale supplémentaire à l'expérimentateur, et cette donnée lui permettrait en retour d'avoir une représentation plus globale et fidèle des motifs de la décision.

C'est cette fécondité qui caractérise l'introduction des significations dans l'œuvre de Davidson. Pour construire une théorie qui prenne en compte ces dernières, il choisit de s'appuyer sur Richard Jeffrey (1926-2002) et, plus précisément, sur son modèle de décision dont l'ontologie porte uniquement sur des propositions. Jeffrey [1965,1983] avait en effet conçu une théorie de l'utilité espérée dont les objets étaient des propositions, c'est-à-dire un ensemble de phrases dotées de significations sur lesquelles portent les préférences des agents ainsi que leurs jugements de probabilité. En utilisant les axiomes d'Ethan Bolker (1967) et la logique propositionnelle, Jeffrey était en mesure de formuler un théorème de représentation (c'est-à-dire un théorème permettant de représenter la relation de préférences par une fonction d'utilité) pour cette nouvelle théorie. Davidson reprend le modèle de Jeffrey

dans sa structure générale en substituant toutefois aux propositions des phrases non interprétées afin de ne pas postuler à l'avance les significations. Ainsi, tout comme Ramsey proposait de mesurer les utilités cardinales et les probabilités subjectives à partir d'une donnée minimale directement ouverte à l'observation (les préférences ordinales) – méthode que le Davidson économiste avait utilisé dans les années 1950 –, Davidson propose de ne considérer qu'une donnée minimale, l'attitude consistant à tenir pour vraie une phrase à un certain moment et dans un certain contexte, et ce afin de déterminer simultanément, à la manière de Ramsey, les croyances et les significations.

Ce que l'on vient d'évoquer rapidement s'est pourtant déroulé pendant plus de vingt ans. Cette évolution va conduire Davidson à placer au premier plan l'interdépendance entre désirs, croyances et significations. L'opération est complexe puisqu'elle conduit à réunir deux théories initialement distinctes : celle de la décision et celle de l'interprétation. C'est ainsi qu'en 1980, la théorie dite « unifiée » apparaît dans l'œuvre de Davidson. Du point de vue de la théorie de la décision, le changement de perspective est important puisqu'il conduit à élargir sa base, le couple désirs-croyances, au triplet désirs-croyances-significations. L'initiative de Davidson pourrait rassembler à la fois celle d'économistes qui considèreraient que la théorie standard de la décision ne suffit pas à la rationaliser, et celle de philosophes pour qui une analyse philosophique de la décision ou de l'action ne suffirait pas à embrasser le problème dans son ensemble.

La voie que Davidson choisit peut toutefois paraître surprenante : les critiques adressées à la théorie de l'utilité espérée[3] ne le conduisent pas à choisir la voie de la théorie de l'utilité non espérée ou celle de ce qui deviendra l'économie comportementale. L'issue pour lui réside dans l'introduction de ce troisième élément, la signification, voie qui n'est pourtant pas jusque là envisagée par les économistes. Pour un économiste, cette introduction peut paraître déroutante puisqu'elle conduit à faire apparaître dans l'analyse quelque chose qui semble en dehors de sa compétence. Pourtant les outils mêmes que Davidson utilise pour construire sa théorie unifiée sont issus des théories économiques de Jeffrey et Ramsey, puis combinés avec d'autres

[3] Comme par exemple les effets de présentation ou les effets de récence.

inspirés des théories philosophiques de Tarski et de Quine. Dans cette opération, l'économiste n'oublie donc en rien sa compétence et ses méthodes ; il ne donne pas une réponse externe, oublieuse de ses acquis. Il élargit seulement son dispositif pour enrichir l'analyse de la décision économique standard.

Pourtant, le caractère apparemment hétéroclite des travaux de Davidson a généralement incité les commentateurs (Tversky [1970], Camerer [1995]) à séparer les recherches économiques en théorie de la décision et les recherches philosophiques portant sur la théorie causale de l'action et la sémantique pour les langues naturelles. Mais si cette séparation est fautive, ce n'est pas seulement eu égard aux impératifs d'une théorie unifiée de la décision. Elle revient, plus généralement, à négliger le « holisme » omniprésent dans l'ensemble de la théorie de Davidson. Ce dernier postule en effet dès le départ une imbrication entre les concepts mentaux et les méthodes permettant de les mesurer. Ce holisme relatif au mental révèle d'une autre manière l'interdépendance des états mentaux représentés par les désirs, les croyances et les significations. Les désirs et les croyances s'expriment en général verbalement et Davidson considère que les énonciations offrent une information immédiatement connectée aux désirs et aux croyances : on ne peut pas comprendre ce qu'une personne dit si nous ne comprenons pas ce qu'elle croit et ce qu'elle désire. Autrement dit, les désirs, les croyances et les significations sont reliés par une interdépendance forte qui fournit une image significative du mental, et cette image est à la fois plus riche que l'image behavioriste standard – dans laquelle l'esprit est une boîte noire qui reçoit des informations par des stimuli et qui en produit d'autres par l'intermédiaire de réponses à ces stimuli – et plus significative puisque sa portée explicative est plus large du fait de la prise en compte d'une donnée supplémentaire, les significations.

Ces trois pôles de la « théorie unifiée de la pensée, de la signification et de l'action » constituent, comme l'explique Pascal Engel [1994], les trois sommets d'un même triangle permettant de construire une théorie de l'interprétation au sein de laquelle la théorie de la décision a une place centrale. Alors que Pascal Engel avait privilégié l'analyse à la fois des relations entre les croyances et les significations, et celles qui unissent les désirs et les significations, il s'agira ici de compléter ce tableau en déplaçant l'accent vers le couple désirs-croyances.

Plus précisément, il s'agira de montrer que la théorie de la décision est non seulement présente dans la théorie unifiée de Davidson et qu'elle en est même le mode d'expression puisqu'elle est axiomatisée et utilise ses outils (plus précisément, ceux de l'utilité espérée) ; mais aussi qu'appréhender les travaux de Davidson sous l'angle de la théorie de la décision permet une mise en perspective directe de ses idées avec celles d'économistes qui participent aux débats relatifs à l'utilité espérée et, plus généralement, à la mesure des préférences. On s'efforcera donc à la fois de décrire les étapes permettant d'intégrer un élément supplémentaire – les significations – au cœur de la théorie de la décision, et de fournir une évaluation d'une telle démarche en tentant d'estimer ses apports et ses limites.

Ces étapes ne sont pas sans liens avec celles qui jalonnent le parcours chronologique de Davidson dont il nous appartient ici de montrer qu'il ne peut être définitivement classé du côté de la philosophie :

- Davidson est d'abord l'un des fondateurs de la théorie de la décision économique tout en étant peu reconnu au sein de notre discipline.

- Il est aussi l'un des pionniers de l'économie expérimentale. Et c'est le rôle décisif qu'il y a joué en présentant des modèles et contributions expérimentaux, à notre connaissance, jusqu'ici restés dans l'ombre, que l'on s'efforcera de mettre en évidence.

- Davidson a ensuite été l'un des plus sévères critiques de ses propres modèles, notant durant vingt ans leurs insuffisances et anomalies, sur le plan tant analytique qu'expérimental. Il questionne donc, en ce sens, une théorie centrale pour notre discipline.

- D'ailleurs, non content de remettre en cause la théorie canonique de la décision à laquelle il a lui-même contribué, il propose d'en construire une alternative qui prendrait en compte ses dimensions philosophiques (en particulier le rôle du langage, de l'interprétation) au lieu de produire une théorie alternative du type utilité non-espérée. En proposant une nouvelle théorie qualifiée d' « unifiée », Davidson caresse l'espoir de construire une théorie plus représentative des attitudes et des motifs qui déterminent la décision, et dont l'objet est supposé être plus large que celui de la théorie de la décision classique.

- Enfin, et cela pourrait paraître une curiosité épistémologique, la théorie *a priori* philosophique qui en résulte, emprunte à la théorie économique une partie de ses concepts et hypothèses, voire reprend presque à l'identique l'un de ses modèles (Jeffrey [1965, 1983]).

Chez Davidson, il n'existe donc ni méthode, ni objet qui puisse être qualifié de « strictement » philosophique ou « strictement » économique. On l'aura compris, dans l'œuvre de Davidson, c'est à cette « théorie unifiée », qu'en tant qu'économiste, on s'est plus spécifiquement intéressé. Car si la théorie de la décision, et en particulier la théorie de l'utilité espérée, est au cœur de la théorie économique, elle fut l'objet de nombreuses critiques depuis les années 1940. Les diverses modifications qui lui ont été apportées depuis la théorie de von Neumann et Morgenstern (comme l'introduction des probabilités subjectives, et les tentatives d'expérimentations) n'ont pas résolu les problèmes d'interprétation, de signification des phrases qui permettent d'exprimer les préférences.
Notre objectif ici est de sortir de l'ombre la proposition de Davidson – qui choisit précisément d'emprunter cette voie – puis de tenter d'en évaluer la portée pour l'économiste.

Nous commencerons par présenter la première version de la théorie de la décision de Davidson en lui consacrant la **première partie** de ce travail (**Comment mesurer l'utilité ? La théorie de la décision face aux tentatives expérimentales (Davidson, 1957)**). La mise en perspective historique et biographique de l'œuvre de Davidson nous permettra d'abord de faire apparaître la logique interne qu'il y voit et l'importance de ses rencontres avec des théoriciens de la décision comme John C. McKinsey (1908-1955), Patrick Suppes (1922-) ainsi que et des spécialistes de la philosophie du langage et de la logique comme Willard von Quine (1908-2000) et Alfred Tarski (1902-1983).

Consacrant la suite de cette partie aux premiers travaux de Davidson en théorie de la décision, nous présenterons le modèle de décision canonique qui prévalait avant ses premiers écrits. Nous retraçons les origines et les débats économiques dans lesquels il s'inscrit en insistant en particulier sur le travail

fondateur de von Neumann et Morgenstern (1944,1947). Outre ces deux auteurs, c'est aussi à Ramsey (1926) et Savage (1954) que Davidson fait référence et nous préciserons les éléments théoriques qu'il leur emprunte pour bâtir son modèle. Nous chercherons ensuite à éclairer les raisons pour lesquelles Davidson choisit de proposer son propre modèle expérimental et en détaillerons à la fois les principales hypothèses, la structure et les conclusions théoriques qu'il en tire en matière de décision économique. Il s'agira, en particulier, de montrer son insatisfaction à l'égard des expérimentations de la théorie de l'utilité espérée réalisée avant les siennes (comme celle de Mosteller et Nogee, 1951) et d'examiner en détail la procédure complexe qu'il choisit de mettre en place pour en pallier les défauts. Dans notre tentative de discussion de son travail, nous montrerons d'une part que Davidson est le premier à proposer un test de l'hypothèse d'utilité espérée avec des probabilités subjectives et, d'autre part, qu'il propose une axiomatique et une procédure de test, minutieuse, détaillée et, en ce sens, riche d'enseignements pour l'économiste. Mais nous soulignerons aussi les limites de l'analyse qu'il propose en 1957, et notamment les critiques qu'il adresse déjà à son propre travail.

La **seconde partie** (**Surmonter les défaillances théoriques et expérimentales de la théorie de la décision : l'introduction de la signification dans la théorie unifiée (Davidson, 1980)**) est d'abord l'occasion d'ajouter aux critiques précédentes, celles du philosophe dans les années 1970. En choisissant pendant cette période, de travailler en philosophie de l'action, Davidson semble avoir bâti un programme de recherche tout à fait différent et qui lui a octroyé l'essentiel de sa célébrité (on peut citer par exemple ses articles « Actions, raisons et causes » [1963], ou encore « Sémantique pour les langues naturelles » [1970]). Cela ne l'empêche guère de garder présent son intérêt pour la théorie de la décision sur laquelle il n'hésite pas à revenir pour mieux la critiquer. Il renforce ainsi progressivement les objections qu'il adresse à la théorie de la décision des années 1950. Celles-ci le conduisent à proposer une nouvelle version de celle-là. Toutefois, le statut de la théorie qui en résulte ne sera pas le même. Davidson propose en effet, comme nous l'avons dit, d'intégrer la théorie de la décision à une théorie unifiée des croyances, désirs et significations. Nous nous attacherons à rendre compte de ce projet et de l'axiomatisation du modèle qui lui est associé. Nous montrerons en particulier que la structure axiomatisée mise

en place dans les années 1950 n'est en rien abandonnée mais utilisée et détournée pour étudier le triplet croyances-désir-significations. Ainsi, la méthode opérationnelle de Ramsey est-elle combinée avec le modèle de Jeffrey qui intègre, à côté des désirs et des croyances, des propositions. Mais Davidson ne se contente pas de combiner les modèles existants, car aucun ne propose d'analyser les significations sans les postuler en amont. Nous montrerons qu'il remplace les « propositions » introduites par Jeffrey par des phrases non interprétées et complexifie encore le modèle de référence. L'analyse des phrases fait alors appel à la philosophie du langage, la logique, et la sémantique et notamment aux travaux de Tarski et Quine.

Nous tenterons enfin de déterminer l'impact des modifications analytiques et méthodologiques opérées par Davidson en matière de théorie de la décision pour discuter l'apport de cette nouvelle version. Il s'agira en particulier de s'interroger sur le statut épistémologique de cette théorie unifiée, sur les effets qui résultent de la combinaison d'arguments issus de disciplines différentes et, surtout, sur cet effort intellectuel auquel nous invite Davidson lorsqu'il propose d'immerger ces utilités et ces probabilités, qui nous offrent le confort d'une quantification numérique envisageable, dans des significations qui semblent la refuser.

Première partie :
Comment mesurer l'utilité ?
La théorie de la décision face aux tentatives expérimentales (Davidson, 1957)

Introduction

La première partie de notre thèse porte sur la théorie de la décision présentée par Donald Davidson au cours des années 1950. Il s'agit ici après avoir resitué le parcours du philosophe économiste qui fait apparaître la fécondité de ses rencontres avec divers théoriciens de la décision, philosophes du langages, logiciens etc. (Chapitre 1), d'évaluer la pertinence de son initiative. Pour ce faire, nous cherchons à mettre en lumière l'inscription de la théorie de Davidson, construite avec Siegel et Suppes (1957) au sein des débats économiques autour de la théorie de l'utilité espérée telle qu'elle se présente à l'époque (Chapitre 2). La présentation des modèles canoniques de von Neumann et Morgenstern[4] (1947), Friedman et Savage, (1948 et 1952) et de Savage (1954) et de la méthode proposée par Ramsey en 1926 nous permettent de montrer les emprunts multiples que Davidson fait aux théoriciens de la décision qui le précède pour bâtir sa propre théorie. En détaillant l'axiomatique, les procédures de tests et les expériences réalisées par Davidson, Suppes et Siegel en 1957 (Chapitre 3), nous chercherons à apprécier l'apport, et les limites, d'une proposition qui consiste à expérimenter une théorie de l'utilité espérée avec probabilités subjectives

[4] Noté vNM par la suite.

Chapitre 1. Parcours de Donald Davidson

L'œuvre de Davidson relève à la fois de la théorie économique de la décision et de la philosophie. Les deux disciplines sont en réalité imbriquées dans l'ambition de bâtir une théorie unique que l'on pourrait représenter, comme le suggère Pascal Engel [1994] par un triangle dont les sommets sont le langage, la pensée et l'action (comprenant les décisions). En mobilisant ces trois pôles, l'idée est d'étudier ensemble et simultanément trois entités mentales : les significations, les croyances et les désirs.

Chacun des sommets de ce triangle est intimement lié à l'autre de telle sorte qu'ils se déterminent mutuellement. C'est pourquoi, dans le travail de Davidson, on ne peut s'intéresser à l'un des sommets sans que les autres ne soient impliqués.

Ainsi les désirs ou les préférences d'un individu ne peuvent-ils être compris sans faire appel à leurs croyances et aux significations qu'ils leur attribuent.

Il vient que, outre chacun des sommets, chacune des arêtes du triangle, chaque connexion – désirs-croyances, désirs-significations et croyances-significations – fait l'objet d'une réflexion de la part de Davidson. Tout au long de cette thèse, nous insisterons sur les imbrications entre ces trois sommets. Il s'agira, d'une part, d'identifier sa conception, sa description de l'imbrication entre des entités mentales (désirs, croyances, significations). D'autre part, à un niveau plus épistémologique, nous chercherons à montrer comment la construction d'une théorie de l'une des entités implique celle d'une autre.

Nous insisterons plus particulièrement sur la manière dont Davidson tente d'inclure une théorie de l'interprétation du langage (des significations) au sein de la théorie de la décision standard en économie (théorie bâtie sur la structure désirs-croyances).

Si dans les chapitres 2 et 3 de cette partie, nous nous intéressons à la connexion entre désirs et croyances dans les travaux de jeunesse de Davison en théorie

expérimentale de la décision (pour faire le lien avec la signification dans la seconde partie), dans ce première chapitre, nous nous efforçons de reconstituer les raisons d'un tel travail triangulaire. Il s'agit de comprendre la manière dont le triangle s'est formé dans la réflexion de l'auteur.

Pour ce faire, nous identifions trois grands moments dans le parcours de Davidson. Le premier est celui de sa thèse de doctorat portant sur le *Philèbe* de Platon (1949,1990). Notre hypothèse est que là s'est sans doute construite sa représentation du désir (1.1).

Les expériences effectuées avec Patrick Suppes (1957), notamment, où Davidson s'intéresse aux probabilités représentent ensuite pour nous une première analyse détaillée des relations entre désirs et croyances en théorie de la décision (1.2). Nous considérons en effet que les probabilités sont des représentations numériques des croyances et nous montrerons comment cette assimilation est possible. Ces travaux correspondent en outre aux premières mises en œuvre d'un schéma méthodologique qui sera suivi jusque dans les années 1990 (prémisses, axiomes, expériences, résultats).

Enfin le troisième moment que nous identifions correspond aux travaux de Davidson en philosophie du langage où celui-ci s'intéresse plus précisément à la sémantique – l'étude des significations (1.3).

1.1. Premiers intérêts philosophiques et thèse à Harvard

Le parcours de Donald Davidson est de bout en bout animé d'intérêts intellectuels multiples, d'une curiosité couvrant plusieurs domaines du savoir et d'un optimisme inébranlable. Pour s'en apercevoir, il suffit de lire attentivement son autobiographie publiée dans *The Philosophy of Donald Davidson*[5].

Les diverses interviews qu'a données Davidson, de 1988 à 1991[6], à Ernest Lepore[7] puis, en 1993[8] à propos de la philosophie du langage, de l'écriture et de la

[5] Lewis Edwin Hahn (ed.), *The Philosophy of Donald Davidson*, Chicago, 1999, The Library of living philosophers, Open court, Intellectual autobiography, pages 2-70.

[6] Interview publiée notamment dans Davidson [2004].

lecture avec Thomas Kent, et enfin celle de mai 2000[9] au journal de Stanford (son ancienne université) sont autant d'autres sources qui nous éclairent aussi sur le parcours de Davidson, et plus précisément sur les raisons de son passage de la philosophie à l'économie expérimentale puis son retour à la philosophie.

En apportant un éclairage historique sur l'œuvre de Davidson et en suivant son parcours personnel, nous espérons donc reconstituer les raisons de son intérêt pour ce triangle, et plus généralement son cheminement intellectuel. Comme il le décrit en filigrane dans son autobiographie, ses rencontres (et en particulier celle avec John Goheen (1906-1994) ou Patrick Suppes) et ses expériences à Stanford ont alimenté de manière significative la réflexion de l'auteur dans tous les domaines.

1.1.1. L'éveil à la philosophie

Donald Davidson est né le 6 mars 1917 à Springfield dans le Massachusetts. Sa famille a pris l'habitude de déménager au gré des différents emplois qu'occupe le père, Clarence Herbert Davidson, ingénieur. Ainsi, après Springfield, partent-ils pour les Philippines pour trois ans, pour rejoindre ensuite la banlieue de Philadelphie.

Ces déplacements successifs ne permettent guère au jeune Davidson d'être scolarisé. Lecture et écriture lui sont apprises par sa mère – Grace Cordelia Anthony - jusqu'à l'âge de 9 ans, lors de l'emménagement de la famille à Staten Island, âge auquel il rejoint les bancs de l'école. Il y reçoit une « éducation progressiste » selon ses propres termes par le biais de son professeur, Madame Wilcox, qui suit les préceptes de John Dewey (1859-1952) en matière d'éducation.

Son initiation à la musique, d'abord via le piano puis plusieurs autres instruments comme le violon et la trompette, et finalement la clarinette à laquelle il s'attachera, traduit en outre à la fois une capacité d'apprentissage extrêmement développée puisqu'après deux semaines d'enseignement intensif, il sait déchiffrer, comme il le souligne, la musique sur une partition, mais aussi un premier intérêt pour la

---

[7] Philosophe et ami et Davidson, il est notamment l'éditeur des Blackwell series « Philosophers and Their Critics ».

[8] Disponible sur le site http://jac.gsu.edu/jac/13.1/Articles/1.htm.

[9] Disponible sur le site http://www.stanford.edu/group/dualist/vol7/pdfs/davidson.pdf.

traduction d'un langage en un autre ; il est en effet fasciné par le fait que la musique puisse se lire.

Son éveil à la philosophie et en particulier à celle du langage est d'ailleurs intimement lié à ce qu'il vit comme le montre l'exemple suivant, qu'il donne lui-même.

Davidson raconte qu'une nuit, par temps de neige, alors qu'il raccompagne une amie après une fête, celle-ci est renversée par une voiture. Davidson demande au conducteur de rester à l'endroit de l'accident et tente de ramener son amie, inconsciente chez elle. La voiture n'est plus là lorsque Davidson revient sur les lieux, mais peu de temps après le conducteur est rattrapé, et poursuivit en justice. Lors du procès le juge demande à Davidson si celui-ci a eu une conversation avec le conducteur. Il rétorque qu'il lui a bien parlé mais sans obtenir de réponse. Alors que le juge le somme de ne répondre que par « oui » ou « non », Davidson explique qu'il ne sait pas si des propos énoncés à un interlocuteur muet constituent une conversation. Davidson précisera que sa position était sous tendue par un précepte philosophique cohérent : quand l'application d'un concept, dans un cas particulier, dépend d'un nombre de facteurs qui le sous-tendent et que ces facteurs ont été entièrement spécifiés, il n'est pas utile de demander, en plus, si le concept original s'applique (Davidson [1999], p.9).

Mais l'éveil à la philosophie, par les textes, a véritablement lieu au lycée lorsque Davidson se passionne pour Nietzsche et Platon ainsi que pour Kant. L'auteur raconte d'ailleurs qu'il voit dans les dialogues de Platon tous les éléments d'une pièce de théâtre tant les dialogues lui semblent vivants.

Si la nature curieuse, interrogative et passionnée du jeune Davidson laisse deviner son élan philosophique, ce sont surtout ses intérêts pour les sciences sociales qui vont le révéler dans la pratique.

1.1.2. L'éveil aux sciences sociales et à la politique

Au lycée, Davidson se lie d'amitié avec Hume Dow qu'il suivra à Harvard. Ce dernier lui fait découvrir la littérature française et russe du XIXème siècle ainsi que

la philosophie et les doctrines politiques. Davidson s'intéresse notamment au socialisme et se rallie aux côtés des mouvements pacifistes et antifascistes.
Cet intérêt théorique est renforcé par une prise de conscience née d'une confrontation la réalité sociale. En 1934, en effet, son père lui trouve un emploi saisonnier comme commis sur un cargo appartenant à la Bethlehem Steel Company. L'équipage prend la route du canal de Panama puis rejoint le port de San Pedro à Los Angeles.
La vie et la nature des tâches qui lui sont confiées sur le navire lui paraissent particulièrement rudes et peu attrayantes et à son arrivée à San Pedro, il rejoint le syndicat des dockers qui avait appelé à la grève du fait de mauvaises conditions de travail, menaçant les non-grévistes de représailles.
L'implication au sein du syndicat (il assiste par exemple au « Bloody Thursday » du 5 juillet 1934) mais aussi les résultats obtenus - une victoire des grévistes et l'amélioration des conditions de vie sur le bateau - lui donnent un aperçu de ce qu'une union solide peut accomplir. C'est aussi là que, selon lui, le goût du voyage lui est venu.
Davidson est alors dans sa dernière année de lycée et il choisit Harvard pour la suite de ses études.

1.1.3. Harvard : l'éveil à la logique

C'est à l'automne 1935 que Davidson entre à Harvard. Plus tard, il soulignera la chance qu'il eut, en tant qu'étudiant, de pourvoir y côtoyer des professeurs de premier ordre qui eurent sur lui une influence primordiale.
Davidson étudie la littérature anglaise, plus précisément Shakespeare et la poésie des 17ème et 18ème siècles ainsi que la Bible et se passionne pour l'histoire de la philosophie et l'histoire des idées en particulier (Davidson [1999], p.14). Cet intérêt est nourri par sa rencontre essentielle avec Alfred Whitehead (1861-1947), spécialiste de la discipline, qui exerce une grande influence sur lui. Whitehead le séduit en effet par sa vivacité, le prend sous son aile, l'invitant par exemple chez lui à l'heure du thé (Davidson [1999], p.13). La fécondité de cette rencontre repose surtout sur la stimulation ressentie par Davidson au contact de Whitehead ; tel est par exemple le cas lorsque le premier, en deuxième année, demande au second s'il peut

assister au séminaire réservé aux étudiants relatif à l'ouvrage *Process and Reality* (Whitehead, 1929). Davidson excellera dans ce cours pour lequel il écrira un mémoire accompagné son article de poèmes et d'images illustrant son propos.

Outre ce cours d'histoire des idées, l'éducation que reçoit Davidson est très riche et variée. Très vite, il suit des cours de grec et découvre l'architecture, la philosophie ainsi que l'histoire grecques mais ce n'est qu'à la fin de son parcours universitaire à Harvard (et profitant d'une bourse universitaire) que Davidson se plonge véritablement dans la philosophie en suivant des cours d'éthique, de logique, de métaphysique et d'histoire de la philosophie.

Comme il l'explique dans son autobiographie, sa première véritable épreuve en philosophie fut la demande de son tuteur, David Prall : écrire sur le thème « Free will and determinism ». Davidson penche du côté du déterminisme pensant que toute chose doit avoir une explication parfaite (Davidson [1999], p.15). Il ne changera d'ailleurs jamais de position et se refusera constamment à croire en d'éventuels miracles pour expliquer des phénomènes. Tel sera en particulier le cas lors de ses réflexions notamment relatives à l'action. Ainsi, lorsque l'auteur explique les liens entre intention et action, il précise que « partout où il y a comportement l'intention intervient d'une façon ou d'une autre » (Davidson [1974,1993a], p.305). Toute forme de comportement peut donc être expliquée intentionnellement.

Une autre rencontre importante est celle de Clarence Irving Lewis (1883-1964) qui l'a incité à écrire sur le *Philèbe* de Platon pour sa thèse de doctorat et qui eut aussi, selon Davidson, une grande influence sur l'épistémologie de Quine.

Lewis est notamment l'auteur d'une une étude critique des *Principia Mathematica* de Whitehead mais aussi et surtout l'un des représentants influents du pragmatisme dans les années 1930-1940. Plus précisément, Lewis tente d'appliquer le principe pragmatiste aux nouvelles questions de la logique, la linguistique, et la politique (Deledalle [1998], p.187). Mais l'influence sans doute la plus féconde de Lewis sur Davidson est relative à la théorie de la valeur. Lewis considère le processus d'évaluation comme la réponse d'un organisme à différents types d'expériences[10]. La

[10] Comme le souligne Roger Pouivet ([2003], p.336) concernant la théorie de la valeur de Lewis, « une chose est une bonne chose, autrement dit elle a une valeur, si elle possède une propriété grâce à laquelle elle est un potentiel de réalisation d'expériences positives futures. La signification d'un énoncé qui attribue une valeur

valeur ou l'absence de valeur (ou de manière plus précise la valorisation et la non valorisation – *value/disvalue*) donne accès, selon Lewis, au désir et à l'aversion (Meckler [1950]). Cette conception imbriquée de la valeur et du désir sera utilisée par Davidson lorsque celui-ci écrira son premier article majeur en 1955 en collaboration avec McKinsey et Suppes.

Mais la rencontre fondamentale pour Davidson est celle avec W.V.O. Quine (1908-2000)[11], dont l'influence sur les travaux ultérieurs sera essentielle comme on le verra dans les chapitre 3 et 4 de la partie II notamment. Le premier est en effet très frappé par l'absence de formation philosophique caractérisant Quine à son arrivée à Harvard, qui ne l'a pourtant pas empêché de décrocher son doctorat en deux ans. Selon Davidson très peu de personnes pouvaient à l'époque se prévaloir de connaitre sérieusement la logique.

Davidson assiste dès l'automne 1939 au cours de logique de Quine et, selon ses propres termes, c'est sous la « tutelle » de Quine qu'il découvre le plaisir de combiner des preuves élémentaires formelles (Davidson [1999], pp. 22-23). Davidson et Quine ne cesseront d'interagir.

Très vite, Davidson se passionne par exemple pour les débats que Quine mène avec Carnap notamment à propos des objections adressées par Quine à la fameuse « distinction analytique-synthétique »[12] – ce qui l'amène à évoluer dans sa pensée et dans sa compréhension de l'histoire des idées et s'atteler avec rigueur à la philosophie (Davidson [1999], p.19).

Davidson aime d'ailleurs à relater les origines de ses discussions avec Quine (Davidson [1999], p.22) : lorsqu'ils se rencontrent à Harvard, l'une de leurs premières conversations est relative à la vérité, en tant qu'objet théorique. Quine demande à Davidson s'il pense que la phrase « la neige est blanche » est vraie si et

---

est une certaine propriété que possède une chose, propriété dont la présence est manifestée dans l'expérience que nous en faisons […] La valeur en tant que propriété est une potentialité d'expérience ayant une certaine qualité. »

[11] Avant que celui-ci ne parte pour l'Europe où il rencontrera Carnap (1891-1970), ce qui constituera pour lui, un vrai bouleversement.

[12] « Le dualisme du synthétique et de l'analytique est un dualisme entre des phrases dont certaines sont vraies (ou fausses) à la fois en vertu de ce qu'elles signifient et en vertu de leur contenu empirique, alors que d'autres sont vraies (ou fausses) en vertu de leur signification seulement, et n'ont pas de contenu empirique » (Davidson [1974, 1993b], p. 275).

seulement si la neige était blanche. Cette assertion renvoie à une conception particulière de la vérité, la vérité comme correspondance. Ce type de problème sera la pierre de touche d'une série de débats entre Quine et Davidson notamment, plusieurs années après (Quine [1999], p.74).

Toutefois Davidson ne commencera sa thèse qu'en mars 1946 après s'être engagé comme volontaire dans la réserve navale, lui qui pourtant voulait être pilote mais avait dû se raviser du fait d'une vue médiocre[13]. Sur les conseils de Lewis, Davidson écrit alors sa thèse sur le Philèbe de Platon, qu'il terminera en 1949, bénéficiant pendant ces années, par l'intermédiaire de John Goheen, le plus vieil ami philosophe de Davidson, d'un poste au Queen's college de New York alors qu'il n'a pas encore fini sa thèse.

Ces années d'études à Harvard mènent Davidson à s'interroger sur la notion de désir. Comme il le souligne, il fut surpris de la manière dont les philosophes, de Platon à Mill en passant par Hume, « semblent tous d'accord sur un point : ce que quelqu'un désire est véritablement désirable » (Davidson [1999], p.31). Autrement dit, désirer quelque chose est la preuve que cette chose est désirable. Cette question de la nature du désir émerge sans doute chez Davidson à la suite de sa thèse de doctorat sur le *Philèbe* comme en témoignent les articles ultérieurs publiés par l'auteur dans les années 1980-1990[14]. En effet, Platon, dans le *Philèbe*, évoque la genèse des affections comme l'espoir et la crainte, le plaisir et la peine. Parmi les plaisirs psychiques, on trouve le désir. L'une des idées centrales de ce dialogue est, plus précisément, la coexistence constante entre le niveau psychique et le niveau physique du désir. Par exemple, un état physiologique comme la faim ou la soif sera suivi par un état psychique mettant en jeu des peines présentes et des plaisirs futurs si la faim et la soif ne tardent pas à être assouvies. Même si ces deux niveaux ne sont pas identiques, leurs relations ont une influence sur la conception du désir et ce type de positionnement théorique ne sera pas sans effet, comme on le verra, sur les réflexions ultérieures de Davidson.

---

[13] Lorsque les Etats-Unis décidèrent de s'engager dans la seconde guerre mondiale, Davidson ne voulait pas y prendre part car il considérait que ce conflit était une lutte pour les marchés. Mais lorsque l'Allemagne envahit les pays alliés, son attitude changea et il s'engagea. De 1942 à 1946, il est instructeur, en charge d'un programme destiné à préparer les pilotes à discriminer entre les navires et avions amis ou ennemis.

[14] Voir par exemple Davidson [1985b], [1997].

1.2. Arrivée à Stanford, la théorie de la décision

En janvier 1951, Goheen devient président de Stanford et offre un poste à Davidson en philosophie. Ce dernier explique n'avoir alors aucune spécialité, se refusant à choisir parmi les domaines de la philosophie.

Son intérêt pour la théorie de la décision nait de plusieurs rencontres. C'est en effet d'abord à ce moment qu'il rencontre Patrick Suppes, arrivé quant à lui en septembre 1950. La chaire de philosophie est encore peu développée à Stanford et sa réputation quasi inexistante mais elle est notamment portée par un autre personnage important, John Charles McKinsey, un logicien auteur notamment d'un ouvrage de théorie des jeux (McKinsey [1952]), qui sensibilise Davidson et Suppes à la théorie de la décision.

Les suggestions de Suppes et McKinsey permettront l'écriture de l'article « Outlines of a formal theory of value » [1955], dont l'objectif est de mêler à la fois les débats philosophiques relatifs à la valeur (chez Kant par exemple), la théorie de la mesure développée notamment dans les sciences physiques ainsi que l'axiomatique de l'utilité espérée proposée par vNM (Davidson, McKinsey, Suppes [1955], p.140). L'ouvrage *Decison making* publié en 1957 fut le résultat de ces tentatives avec l'objectif de « déterminer les probabilités subjectives d'une personne et son classement de préférences simultanément » (Davidson [1999], p.32).

L'auteur mentionne les différentes lectures qu'il est amené à faire dans les années 1950 en théorie de la décision lui permettant de comprendre que plusieurs expériences de la théorie de l'utilité espérée ont déjà eu lieu mais que la plupart du temps, celles-ci faisaient usage de probabilités objectives. Or, comme le souligne Davidson, il était possible dès les années 1950 de proposer un modèle permettant de déterminer à la fois les utilités cardinales et les probabilités subjectives et ceci n'avait jamais été fait. Davidson prétend avoir trouvé, dès le milieu des années 1950, un moyen de déterminer les utilités et probabilités subjectives en fixant au départ les probabilités à ½. Mais comme Herbert Bohnert avant lui, Davidson a été « victime » (Davidson [1999], p.32) de l'effet Ramsey puisque ce dernier avait déjà trouvé cette méthode dans les années 1920.

Désappointé par cette découverte, et prenant conscience que celle-ci n'était pas exploitée par vNM et Friedman et Savage, il choisit de se consacrer au versant expérimental de la théorie de l'utilité espérée.

La rencontre d'une théorie économique aux méthodologies (axiomatisées) et objets (décisions individuelles des acteurs) spécifiques, eut un impact fort sur la démarche philosophique de Davidson. L'utilisation d'hypothèses et de conditions formelles permettant la construction d'une axiomatique cohérente et précise et d'une procédure de test pouvait en effet, selon lui, constituer une méthode alternative « aux tentatives généralement futiles réalisées pour définir et analyser des concepts » (*ibid*. p. 32). Cette méthode inspirée de la théorie de la décision (hypothèses, conditions formelles, axiomes) avait, selon lui, l'avantage d'établir des liens et des corrélations précis entre des concepts jusque là difficiles à définir et analyser de manière absolue.

Cette théorie de la décision comportait néanmoins un certain nombre de défauts qu'il contribua à rendre saillants d'abord au travers de l'article de 1959 écrit avec Marschak puis tout au long des années 1970 et 1980 : si la théorie de la décision était capable de produire une structure puissante et solide « contenant des termes comme « préférer à », elle ne disait rien de ce que signifiait « la préférence » » (*ibid*. p. 32).

Davidson gardera donc la structure formelle, le schéma méthodologique de ses premiers modèles de décisions bâtis en 1955, 1957 et 1959 dans sa nouvelle théorie des années 1980.

1.3. La philosophie du langage : un nouveau questionnement de la théorie de la décision

En plus de l'initier à cette théorie de la décision, McKinsey incite Davidson à renouer avec ses premiers intérêts en lui proposant de coécrire un article sur « la méthode extension/intension de Carnap », objet que ce Davidson semblait mal connaitre (Davidson [1999], p. 33). La mort de McKinsey peu avant la publication de l'article de 1955 va le forcer à investir le champ par lui-même.

La rédaction de l'article lui demande plus d'un an. Mais cet article aura plus d'une conséquence positive sur la suite de son travail. Il constitue d'abord pour lui «

expérience éclairante » (*ibid.*) et plus précisément un tournant vers l'analyse de deux thèmes centraux : « qu'est ce que la signification ? » et « la sémantique » telle que traitée par Gottlob Frege (1848-1925), Bertrand Russell (1872-1970), Willard von Orman Quine (1908-2000) et Rudolph Carnap (1891-1970).

Une fois cet article envoyé à Carnap, une discussion s'engage entre les deux philosophes à propos de sémantique mais aussi plus curieusement des premiers travaux en théorie de la décision de Davidson dont Carnap avait pris connaissance. Davidson se dit alors obsédé par « le problème d'une sémantique satisfaisante pour le discours indirect, les phrases rapportant des croyances et les auteurs phrases rapportant des contenus intensionnels » (Davidson [1999], p.34).

Toutes ces questions relatives à la philosophie de l'action et à la philosophie du langage lui suggèrent de nouvelles interrogations quant à la fécondité et la validité de la théorie de la décision et l'incitent à y revenir, comme on le verra, dans les années 1980.

Enfin, cet article portant sur le débat intension/extension chez Carnap, est présenté lors d'une conférence à Berkeley et cette occasion lui permettra de rencontrer Tarski (a) et de retrouver Quine (b).

(a) La rencontre avec Tarski

Avec Tarski, Davidson trouve une première solution au problème qui l'occupe : « une analyse de la forme logique des phrases ne peut être satisfaisante que si elle peut être incorporée dans une définition de la vérité » (*ibid.*, p. 35).

Les travaux de Tarski sont en fait à l'origine des premiers travaux de Davidson en philosophie du langage : « Truth and Meaning » (1967) constitue l'un des premières tentatives de l'auteur d'adapter la théorie de la vérité de Tarski pour les langages formels, aux langages naturels. Cette tentative sera plus tard appelé le « programme de Davidson ». Cette idée d'analyser la structure logique des phrases – mêlée à celle de construire une théorie de la vérité à la Tarski pour les langages naturels – incitera Davidson à introduire ce type d'analyse en philosophie de l'action. En effet, c'est en cherchant ce que la modification des adverbes modifie dans une phrase, qu'il lui

apparut que cela ne pouvait se déterminer qu'en référence à des évènements particuliers, des actions.

(b) Word and object, Quine (1960)

C'est alors que Davidson retrouve Quine au « Center for Advanced Studies in Behavorial » sciences à Stanford en 1958-1959. L'ouvrage *Word and Object* alors en préparation va changer la vie de Davidson selon ses propres termes (Davidson [1999], p. 41).

En effet, depuis le début des années 1950, Quine avait proposé à plusieurs reprises à Davidson de lire ses manuscrits avant même qu'ils ne soient publiés. Davidson mentionne par exemple l'été 1950 où Quine l'avait rejoint à Nice pour lui faire découvrir le manuscrit dans lequel il fustigeait les deux dogmes de l'empirisme.

Vers la fin des années 1950, Davidson est à un tournant de son parcours intellectuel : il est perplexe face aux résultats expérimentaux issus de *Decision Making* ainsi que ceux de l'article publié en 1959 en collaboration avec Jacob Marschak. Il adresse alors toute une série de critiques à la théorie de la décision : des critiques internes relatives au caractère statique de la théorie ainsi qu'à l'impossibilité de prédire de manière fiable le comportement des sujets (nous analyserons ces critiques dans le dernier chapitre de la partie I ; des critiques externes qui recouvrent essentiellement l'impasse faite sur les significations comme on le verra au début de la partie II).

Lorsque Davidson découvre *Word and Object*, (1960) l'ouvrage de Quine, une réponse à ces objections à l'égard de la théorie de la décision se présente à lui. Ce qui frappe d'emblée Davidson c'est l'usage combiné de la logique et de la sémantique. L'ouvrage est complexe mais il place comme thème central l'idée « qu'il n'existe pas de signification autre que celle qui peut être apprise quand elle se manifeste dans le comportement verbal des locuteurs » (Davidson [1999], p.41). Plus précisément, l'attribution de significations est la clé de l'apprentissage du langage. Ce « tournant linguistique » marquera Davidson jusqu'à la fin de sa carrière et il constitue le premier accès au troisième sommet du triangle, les significations. Dès lors, il n'aura de cesse de s'interroger sur les imbrications entre les trois sommets. Ce tournant linguistique fera partie intégrante des travaux de l'auteur à partir des années 1970.

Dans l'intervalle, Davidson réinvestit le couple désirs-croyances, en philosophie cette fois.

Il écrit alors nombre d'articles célèbres en philosophie de l'action. Parmi ceux-ci, on peut citer « Actions, raisons et causes » [1963], « La forme logique des phrases d'action » [1967a], ou encore « Comment la faiblesse de la volonté est-elle possible ? » [1970a].

Sur la période allant de la fin des années 1960 jusqu'au début des années 1970, les travaux de Davidson portent à la fois sur la théorie de l'action et sur la théorie de l'interprétation du langage. En effet, les discussions avec Quine l'amènent à s'intéresser aux liens entre significations et croyances. Toutefois, comme nous le montrerons, Davidson n'adhère pas au réductionnisme de Quine faisant des stimulations sensorielles et des réponses verbales les données de base d'une théorie de l'interprétation. Davidson propose dans plusieurs articles (« Truth and Meaning » [1967b], « Semantics for Natural Languages » [1970b]) une théorie de l'interprétation faisant usage non seulement des travaux de Quine mais aussi ceux de Tarski en logique.

Il construit peu à peu sa propre théorie portant sur le triangle désirs-croyances-significations avec l'objectif de conserver une structure formelle qui supposent de respecter des conditions de rationalité sous forme de définitions et hypothèses, pour définir des axiomes et théorèmes. Alors qu'il arrive à l'université de Chicago[15], il commence un nouveau cycle de conférence sur « les paradoxes de l'irrationalité » à partir de 1977. L'idée centrale défendue par Davidson est que l'irrationalité n'a de sens que sur fond de rationalité. Autrement dit, l'irrationalité n'est attribuable qu'à des créatures rationnelles. Toutefois une fois cette structure mise en place, il faut analyser l'interprétation des significations.

A partir des années 1980, commencent les conférences sur l'interprétation radicale, notamment à Oslo avec Stig Kanger, où Davidson explique « chercher à extraire une

[15] Après plusieurs années passées à Princeton, Davidson se rend en 1970 à l'université Rockefeller de New York où il participe à des groupes interdisciplinaires (physiciens, logiciens, mathématiciens, biologistes…) : « J'aimais le contact avec des sciences sérieuses » (Davidson [1999], p. 51) dira Davidson mais aussi la liberté qui lui est laissée de faire autant de déplacements qu'il le souhaite. Il en profitera pour retrouver Quine en 1973 et 1974 lors d'années passées à Oxford.
Il se voit plus tard offrir un nouveau poste à l'université de Chicago vers. La vie là bas est particulièrement plaisante pour lui notamment parce que les groupes interdisciplinaires qu'il a jusqu'à alors fréquentés ailleurs ont à Chicago un succès particulier.

théorie de la signification d'une version de la théorie de la décision » ([1999], p. 57), travail auquel il réfléchit depuis des années. C'est alors que Davidson publie son article « A unified theory of thought, meaning and action » où celui-ci tente de réunir, en utilisant les travaux de Jeffrey notamment, la théorie de la décision telle qu'il l'a testé expérimentalement et la théorie de l'interprétation du langage.
Davidson rejoint l'université de Californie en 1981[16]. Richard Rorty organise en 1983 une session de la société allemande de Hegel regroupant Quine et Hilary Putnam où Davidson parle d'une « A Coherent Theory of Truth and Knowledge » et surtout où s'engage entre Quine et Davidson un débat autour de l'accès aux croyances.

En 1993, c'est une succession de conférences sur Spinoza qu'il donne. Pour Davidson, le monisme anomal (la relation qu'il suppose entre le physique et le mental) s'avère être tout à fait proche de la conception spinoziste. Cette découverte tardive de sa proximité avec Spinoza témoigne, selon Davidson, de son trop peu de connaissance de l'histoire de la philosophie ou de ses oublis sauf Aristote, Platon, Kant, et Hume qui furent toujours au cœur de ses influences. Davidson se voit attribué des prix (Hegel prize) des comparaisons avec Derrida ou Heidegger ou encore à un héros de la philosophie analytique.

Conclusion

[16] C'est au cours des ses années passées au sein de l'université de Californie qui une fois encore lui offre nombre de possibilités de voyager qu'il se marie avec Marcia Cavell, une étudiante qu'il avait connue à Stanford devenue maitre de conférences depuis. Durant quelques mois passés à Oxford à nouveau, il discute avec étudiants et enseignants des causes de l'action. Pour lui, ce sont les désirs, les significations et les croyances mais pour Chris Peacoke et d'autres, il faut rajouter la perception. Les conférences qui suivent avec Marcia Cavell ont lieu en Inde, Argentine, Afrique du sud, Zimbabwe, Irlande, Serbie, Autriche, Slovénie, Israël etc.. Ils y rencontrent de nombreux philosophes menacés, dépossédés.

Les rencontres de Davidson avec les théoriciens de la décision l'incitent au sortir de sa thèse sur le Philèbe de Platon, alors qu'il n'est pas spécialisé dans un domaine précis de la philosophie, à s'intéresser à cette théorie économique. Ses relations avec les philosophes et ses travaux d'importance en philosophie du langage et de l'action ne le détourneront pas de ce premier intérêt comme nous allons le voir tout au long de cette thèse. Au contraire, ils seront mis au service d'un remodelage constant de cette théorie jusque dans les années 1990.

Chapitre 2 :

Les fondements du modèle de Davidson (1957)

Après avoir décrit le contexte intellectuel dans lequel Davidson a évolué au cours des années 1940 et 1950, et avant de mettre en lumière les apports du modèle de Davidson à la théorie de la décision, nous en précisons la construction.

La contribution de Davidson en 1957 intervient alors que les jalons de cette théorie ont déjà été posés notamment au travers du modèle canonique de Von Neumann et Morgenstern (*Theory of Games and Economic Behavior* [1947], deuxième édition où se trouve l'appendice sur l'axiomatisation de l'utilité). En dix ans, les enjeux épistémologiques et analytiques de leur modèle ont été discutés, les concepts auxquels font appel vNM ont été remis en cause si bien que d'autres axiomatiques ont été proposées, en particulier celle de Friedman et Savage en 1952.

La théorie de Davidson, qui s'inscrit dans cette lignée, est une théorie complexe pour plusieurs raisons.

La première tient à son objectif : comme indiqué dans le sous-titre de l'ouvrage écrit avec Siegel et Suppes, *Decision making* (1957), Davidson souhaite adopter « une approche expérimentale ». Pour lui, en effet, « aucune interprétation empirique satisfaisante de la théorie de la décision n'a été donnée, il est ainsi impossible de tester cette théorie » (Davidson, Suppes et Siegel [1957], p. 3). Il ne s'agit évidemment pas pour Davidson de trouver des applications pratiques de la théorie pour un acteur économique. Davidson considère, en effet, que la théorie de la décision est une théorie « normative[17] » (Davidson, Suppes et Siegel [1957], p.2). C'est donc en tant que telle qu'il propose de tester la théorie de la décision. Pour lui, en effet, « même si personne, ou au mieux une personne agit en accord avec la théorie au cours d'une période, cela peut valoir la peine de se demander, pour nous-mêmes ou pour quelqu'un d'autre, si un modèle de décision ou de préférences et d'attentes, est rationnel dans le sens de la théorie. Et si nous voulons poser cette

[17] Même si l'auteur a une position plus complexe puisque selon lui, toute théorie normative est aussi en partie descriptive, nous y reviendrons dans l'introduction du chapitre 3.

question, alors il est clair que nous avons besoin d'une interprétation empirique[18] utilisable de la théorie » (Davidson, Suppes et Siegel [1957], p.3). Il s'agit donc sur la base de cet argument de tester empiriquement la théorie de la décision et dépasser en termes de rigueur et de portée la première expérimentation proposée par Mosteller et Nogee (1951). Cet objectif détermine largement la forme et la nature du modèle de Davidson en 1957.

Pour ce faire, Davidson propose de construire un nouveau « modèle ». Ce dernier terme est toutefois ambigu dans *Decision making* – et c'est la deuxième raison pour laquelle le modèle est complexe. Les auteurs eux-mêmes soulignent la confusion qui règne autour du terme même de modèle : « il nous faut notifier que le terme de modèle est utilisé dans deux sens différents dans notre ouvrage. (…) Il est [d'une part] un ensemble de propositions constituant les axiomes d'une théorie [nous l'appelons modèle au sens 1]. D'autre part, en logique, un modèle est usuellement conçu comme un ensemble théorique non-linguistique qui satisfait un ensemble d'axiomes (…) et dans le chapitre II un ensemble d'hypothèses expérimentales (empirical) » (*ibid*., p.7). Les auteurs restent donc volontairement flous sur la notion de modèle et les limites de celui-ci ; il semble tantôt n'être qu'une somme d'axiomes, tantôt satisfaire des « hypothèses expérimentales » à propos desquelles rien n'est précisé. Davidson suppose que l'utilisation du terme s'éclaircira en fonction du contexte dans lequel il l'emploie.

La distinction entre ces deux usages du même terme est rendue plus obscure encore par celle que font les auteurs entre « théorie » et « modèle » cette fois. En effet, dans le même chapitre introductif, ils proposent de définir une théorie comme « la conjonction d'un modèle formel et d'une interprétation empirique» (*ibid*. p. 5). Le problème concerne donc essentiellement « les hypothèses expérimentales » dont on ne sait pas si elles font partie de « l'interprétation empirique» ou du « modèle ». Si la présentation que nous faisons du « modèle » dans le chapitre 3 éclaire largement son contenu, il peut déjà être noté ici que les auteurs proposent dans *Decision making*, de bâtir une nouvelle axiomatique qu'on peut appeler « modèle » (au sens 1), largement

[18] Davidson, Suppes et Siegel utilisent donc l'adjectif « empirique » dans un sens expérimental.

inspirée de celle de vNM, à laquelle ils ajoutent des hypothèses expérimentales permettant de procéder à des expériences. Nous préférerons ici utiliser le terme de « procédure expérimentale » pour regrouper ces hypothèses.

La complexité de la théorie tient enfin au fait que le modèle canonique de la théorie de la décision, de vNM (1947) à Friedman et Savage (1952), notamment, n'a cessé de se modifier. Pour remplir son objectif, Davidson et *al.* ne testent pourtant par directement la théorie de vNM ou celle de Friedman et Savage. Comme nous l'avons dit, ils construisent leur propre axiomatique susceptible d'être testée. Pour Davidson, Siegel et Suppes, la théorie de vNM ne peut être testée en l'état par exemple car selon eux, elle débouche sur « une liste d'issues infinie, donc impossible à comparer » (*ibid.* p. 8). Les auteurs se montrent donc insatisfaits à l'égard des différents modèles proposés par les théoriciens de la décision et construisent un nouveau modèle qui emprunte à nombre d'entre eux. Le modèle proposé par Davidson fait évidemment appel à l'axiomatique de vNM, et à celle de Friedman et Savage mais aussi aux travaux de Ramsey (1931) par exemple, ou encore de Savage (1954), en choisissant d'intégrer des probabilités subjectives (au contraire de vNM) et en adoptant une démarche complexe qui implique la codétermination des probabilités et des utilités.

Afin de mettre en lumière la construction de la théorie construite par Davidson, Siegel et Suppes en 1957, nous cherchons à identifier la nature de ses emprunts aux différents théoriciens de la décision et les enjeux qu'ils soulèvent, enjeux dont la compréhension s'appuie sur la mise en relation de l'héritage ancien et de débats plus récents de la théorie de la décision.

Nous procéderons ainsi en commençant par présenter le modèle canonique de la théorie de la décision et ses évolutions (2.1).

Nous exposons chronologiquement les types de questionnements auxquels tentent de répondre vNM (2.1.2.) et l'héritage bernouillien (2.1.1.) en présentant à la fois leur démarche analytique (2.1.3) et l'axiomatisation utilisée (2.1.4) avant d'insister sur les remises en cause de ces dernières par Friedman et Savage (2.1.5), deux théories sur lesquelles Davidson, Siegel et Suppes s'appuient.

Puis nous analyserons un second type d'emprunt qui vient enrichir et complexifier la démarche de Davidson : l'emprunt à la théorie de Savage (2.2).

2.1. Le modèle canonique de la théorie de la décision : de l'héritage ancien aux débats modernes

Pour Davidson, Suppes et Siegel, le modèle canonique de la théorie de la décision est celui de vNM formulé en 1947 et remodelé notamment au travers d'une axiomatique précisée et détaillée par Friedman et Savage (1948 et 1952). Ainsi les auteurs de *Decision Making* écrivent-ils : « beaucoup des théories qui ont été développées avec un tant soit peu de précisions sont en fait des modifications mineures de l'axiomatisation de von Neumann et Morgenstern » (Davidson, Suppes et Siegel [1957], p. 3).

Avant d'exposer ces deux modèles retenus par Davidson (celui de vNM en 2.1.4 et sa reprise par Friedman et Savage 2.1.5), il nous faut revenir sur les questionnements analytiques et méthodologiques particuliers au sein desquels ils s'inscrivent.

En effet, selon nous, les apports du modèle de Davidson ne peuvent être compris qu'en resituant son modèle au sein de ces questionnements. Ces derniers, sont à la fois le fruit d'un héritage ancien, notamment associé au nom des Bernoulli, dont il s'agit de mentionner les étapes essentielles (2.1.1) et la résultante de débats au sein de la théorie économique relatifs à la définition, la mesure et la formalisation de l'utilité (2.1.2). La présentation de ces différentes questions nous permettra d'exposer ensuite la manière dont vNM tentent d'y répondre (2.1.3).

2.1.1. L'héritage bernoullien

L'ouvrage de vNM occupe une place essentielle dans la théorie économique à la fois par son caractère fondateur en matière d'axiomatique (de la théorie de la décision) et, d'un point de vue historique, puisqu'il peut être compris comme une

tentative de renouer avec des savoirs anciens, notamment les travaux de Nicolas et de Daniel Bernoulli[19] concernant l'utilité espérée.

Ces deux auteurs prennent place dans une série de débats relatifs à la fois au rôle du calcul des probabilités dans la prise de décision et au choix du critère de décision dans un jeu de hasard. On convient généralement que le critère de choix utilisé[20] avant les travaux de Daniel et de Nicolas Bernoulli était l'espérance mathématique de gain. Comme le souligne Ian Hacking [1975, 2002, p.137], les travaux de Christian Huygens (1629, 1695)[21], auteur du premier manuel de probabilité, faisaient déjà largement usage d'une telle notion. Pour Huygens, l'espérance mathématique correspond au « juste prix »[22], autrement dit au coût qu'il convient de payer lorsque l'on s'engage dans un pari dont les gains varient en fonction du risque, c'est-à-dire en fonction de l'amplitude des gains auxquels sont rattachés des probabilités[23].

Le mémoire de Daniel Bernoulli est une tentative de réponse à la lettre que son cousin Nicolas Bernoulli (1695-1726) adresse à Pierre Rémond de Montmort, lettre dans laquelle est exposé le – communément appelé - paradoxe de Saint-Pétersbourg[24]. Cependant, il semble clair que le critère qui sera mis en avant par Bernoulli ne consiste pas en une théorie alternative à celle de Pascal et Huygens (Jallais, Pradier [1997], p.29)[25].

L'idée est d'essayer de comprendre pourquoi une personne n'est prête à payer qu'une faible somme d'argent pour participer à un jeu dont l'espérance

---

[19] Daniel Bernoulli est cité par von Neumann et Morgenstern lorsqu'ils évoquent la mesure numérique de l'utilité et plus particulièrement l'espérance mathématique comme mesure légitime de celle-ci (von Neumann et Morgenstern [1947, p.28]).

[20] Ce critère était d'ailleurs plus une règle de justice qu'un critère de choix au sens de la théorie de la décision moderne (Jallais, Pradier [1997]).

[21] C'est cependant Blaise Pascal (1623-1662) qui est considéré comme l'inventeur de la théorie de la décision dans la mesure où il est le premier à avoir appliqué, comme le souligne Ian Hacking, « le raisonnement probabiliste à d'autres problèmes que les jeux de hasard » (Hacking [1975, 2002], p. 37]. Il montra avec Pierre de Fermat que l'attrait d'un pari offrant des revenus avec certaines probabilités pouvait être représenté par son espérance mathématique.

[22] Cette idée est déjà présente chez Pascal.

[23] Reprenant la distinction de Frank Knight entre risque et incertitude – « Notre principal intérêt concerne la différence entre le risque comme hasard connaissable et la vraie incertitude » (Knight [1921], p. 21) – nous définissons ici le risque comme un type d'incertitude « probabilisable ».

[24] On peut émettre des doutes quant au statut même de paradoxe posé par Nicolas Bernoulli (Jallais, Pradier [1997].

[25] C'est d'ailleurs Nicolas lui-même qui fait remarquer à Daniel que sa théorie n'est pas concurrente de celle de Pascal.

mathématique de gain est pourtant infinie. Le paradoxe peut plus précisément se formuler ainsi : soit un jeu à partir d'une pièce de monnaie et deux joueurs, si elle tombe sur pile, le joueur A gagne 1 ducat versé par le joueur B et le jeu s'arrête ; si ce n'est pas le cas, on relance la pièce et si elle tombe sur pile, le joueur A gagne 2 ducats ; au troisième jet, B est tenu de remettre 4 ducats à A si la pièce tombe sur pile. Dans le cas contraire, on recommence en attendant que cela arrive. Si le joueur A perd les n-1 parties et gagne la énième, alors le joueur B lui reverse 2^{n-1} ducats. La question est de savoir combien est prêt à payer le joueur A pour avoir le droit de participer à ce jeu ?

On voit ici que l'espérance mathématique pour le joueur A est celle d'un gain infini. En effet, si on gagne au énième coup cela signifie que l'on a obtenu n-1 faces, donc on a une probabilité $(½)^n$ pour un gain de 2^{n-1}, ce qui fait une espérance mathématique de gain de ½. Et en additionnant ces produits (pour obtenir l'espérance de gain du jeu), on trouve une valeur qui tend vers l'infini. Cependant, « un homme de bon sens » (Bernoulli [1725, 1985], p.71) n'acceptera pas de payer une grosse somme pour pouvoir jouer, alors même que l'espérance de gain est infinie et ce, à cause de l'éventualité de gagner trop tôt et donc de gagner peu. Pourtant, l'espérance mathématique nous indiquait que ce jeu était à l'avantage du joueur A puisque celui-ci ne peut payer qu'une somme finie, donc inférieure à l'espérance mathématique du jeu, pour y participer.

Cherchant à résoudre ce problème, Daniel Bernoulli (1700–1782), cousin de Nicolas, propose un autre critère que celui de l'espérance mathématique des gains comme critère de choix de l'action, et intègre les spécificités individuelles du décideur à la détermination du choix.

En effet, comme il le souligne, si l'on accepte plus généralement la proposition suivant laquelle le critère de choix est l'espérance mathématique, deux personnes faisant face au même pari risqué attribueraient la même valeur au risque. Rien ne permettrait donc, avec ce critère, de distinguer les caractéristiques individuelles des décideurs.

C'est pourquoi Bernoulli propose une solution nouvelle incluant *l'utilité* de l'individu lorsque celui-ci est confronté à un pari risqué, ce qu'on appellera

usuellement *l'espérance d'utilité*. Il imagine un homme pauvre obtenant un ticket de loterie qui lui permet d'obtenir 20000 ducats avec une probabilité de ½ et rien avec la même probabilité. Avec les notations modernes, on lui propose donc la loterie L = (20000, 0 ; ½, ½). On peut se demander si cet individu va estimer qu'il va gagner 10000 ducats ou encore s'il serait prêt à vendre son ticket 9000 ducats. Bernoulli répond par la négative à la première question et positivement à la seconde car il considère qu'un homme riche aurait tort de refuser d'acheter ce ticket 9000 ducats (Bernoulli [1725, 1985], p.62). L'auteur en déduit que les gens n'utilisent pas la même règle pour évaluer un pari. Il convient donc de déterminer la valeur de ce dernier à partir non de son *prix*[26], mais de l'utilité qu'il procure et ce faisant, Bernoulli propose de substituer à la valeur monétaire, la « valeur morale » (Bernoulli [1725, 1985], p.72). L'utilité est donc introduite pour prendre en compte les particularités de la personne. Par ce biais, Bernoulli distingue « espérance mathématique » et « espérance morale », cette dernière correspondant à une utilité moyenne[27] (Bernoulli [1738, 1971], p.6)[28].

En introduisant les notions d'utilité et d'espérance d'utilité dans son analyse de la prise de décision en univers risqué, Daniel Bernoulli propose une formalisation de l'utilité sous forme de fonction, celle-ci passe par la valeur subjective de la richesse rattachée à une quantité de monnaie. Elle le conduit à affirmer que l'utilité de la richesse croît moins que proportionnellement que celle-ci et qu'une augmentation de l'utilité, suite à une augmentation de richesse, est inversement proportionnelle au montant de la richesse déjà acquis. Pour la première fois, l'hypothèse d'utilité marginale décroissante est partie intégrante d'une théorie formelle.

Pour représenter l'idée d'utilité marginale décroissante, Daniel Bernoulli propose une fonction d'utilité de la richesse de forme logarithmique, permettant de donner plus d'importance aux petites valeurs de gain et d'exprimer la décroissance de l'utilité marginale de la monnaie. Cette fonction est de la forme $U(x) = k \log \frac{x}{c}$ où k

[26] Si l'on se réfère aux idées de Huygens.

[27] Bernoulli utilise le terme « emolumentum medium » dans le texte original en latin. Ce terme signifie **« rétr**ibution moyenne » ou encore « rémunération moyenne ». Nous conservons le sens économique du terme comme le font Jallais et Pradier [1997].

[28] C'est Gabriel Cramer (1704, 1752) qui, le premier, fit cette distinction.

est une constante, *x* la richesse préalablement possédée et *c* le montant de la richesse nécessaire pour la survie. Cette fonction est concave et permet de décrire la situation d'un individu qui, comme ceux « qui n'ont d'autre fortune que leur force industrieuse » (Bernoulli [1725, 1985], p. 64), considère que le gain d'un ducat sera associée à une valeur d'autant plus grande qu'il est démuni, mais pour qui, plus le gain augmente, plus la valeur morale qui lui est associée croît de moins et moins vite (ce qu'on appelle habituellement l'utilité marginale décroissante).

Cependant, comme nous l'avons mentionné, cette proposition de Bernoulli ne constitue pas en elle-même une nouvelle mesure du risque dans la mesure où elle n'est pas concurrente de la théorie développée par Pascal par exemple. Il semble, en effet, mal venu de comparer une règle de justice et une observation empirique (Jallais, Pradier [1997], p.34). La théorie de Daniel Bernoulli constitue, toutefois, l'une des premières formulations de l'utilité espérée.

Mais l'axiomatique de l'utilité espérée ne figurera que dans la seconde édition de *Theory of games and economic behavior* [1947] de von Neumann et Morgenstern, soit près de deux siècles après l'intuition de Bernoulli. Dans cet intervalle, nombre de débats relatifs à la théorie de l'utilité et à la mesure de celle-ci – essentiellement en univers certain - occupèrent les économistes. Nous allons présenter ces débats afin de comprendre dans quel contexte théorique viennent se placer les travaux de von Neumann et Morgenstern.

2.1.2. La théorie de l'utilité avant von Neumann et Morgenstern : une esquisse des débats et des enjeux en économie.

Lorsque paraît la deuxième édition de *Theory of Games and Economic Behavior* en 1947, la plupart des économistes considèrent que l'étude des préférences ne permet de construire qu'une fonction d'utilité ordinale et se satisfont donc d'une spécification de la fonction d'utilité unique à une transformation monotone

croissante près. Ce positionnement théorique est en particulier celui de Vilfredo Pareto (1848-1923)[29].

Dans son *Manuel* (1909), Pareto rompt, en effet, avec une tradition qui remonte au moins à Stanley Jevons et qui perdure jusqu'aux travaux de Francis Ysidro Edgeworth (1845-1926). Nous revenons dans un premier temps (2.1.2.1) sur la rupture réalisée par Pareto puisque c'est à lui que se réfèrent von Neumann et Morgenstern lorsqu'ils présentent la théorie dominante au moment où ils publient leur ouvrage. La rupture parétienne est véritablement achevée avec l'article de Hicks et Allen [1934]. Cet article tire véritablement les conséquences des travaux de Pareto sur la théorie du consommateur. Nous analyserons les impacts de la rupture parétienne présentés dans cet article (Hicks et Allen, 1934) (2.1.2.2) afin de mettre en relief le cadre dans lequel prend place l'ouvrage de von Neumann et Morgenstern.

2.1.2.1 L'échelle de préférences de Pareto, une rupture par rapport aux théories de Walras, Jevons, et Edgeworth.

De Walras (1834-1910) à Marshall (1842-1924) en passant par Edgeworth (1845-1926), l'utilité était considérée comme une quantité mesurable[30]. Selon ces auteurs, tout individu pouvait en effet, par introspection, associer à chaque panier de bien un nombre réel positif reflétant l'intensité de sa satisfaction à le consommer, c'est à dire son utilité. Ce nombre ayant un sens en soi, l'utilité est dite « cardinale »[31]. A l'instar de Jevons (1835-1882)[32], ces auteurs utilisaient en outre une fonction d'utilité

---

[29] Pareto est mentionné à deux reprises par von Neumann et Morgenstern, voir von Neumann et Morgenstern [1947], pp. 18 et 23.

[30] La théorie de l'utilité a véritablement pris une place centrale en économie à partir des travaux des marginalistes que sont Jevons, Menger et Walras. Pour les trois fondateurs du marginalisme, l'utilité est un fait de l'expérience, un fait conforme à l'introspection la plus commune. Ces trois auteurs raisonnent à la marge, ce qui explique pourquoi ils ouvrent une réflexion approfondie sur l'utilité marginale décroissante.

[31] Par opposition à l'approche « ordinale » où le réel associé à chaque panier de biens ne mesure pas son exacte utilité pour un individu, mais indique simplement si cet individu le juge plus ou moins utile que chacun des autres paniers ; ce réel a donc un sens, non pas en soi, mais relativement aux réels associés aux autres paniers.

[32] Jevons utilise une fonction d'utilité additive et séparable de la forme : $U(x_1, \ldots, x_i, \ldots, x_n) = u_1(x_1) + \ldots + u_i(x_i) + \ldots + u_n(x_n)$ où la satisfaction qu'un agent tire de la consommation d'un ensemble de biens $\{x_1, \ldots, x_i, \ldots, x_n\}$ est la somme de la satisfaction qu'il tire de la consommation de chacun des biens. L'utilité d'un bien ne dépend que de la quantité de ce bien, et non des quantités des autres biens consommées comme dans la théorie d'Edgeworth.

additive. L'hypothèse d'additivité étant d'ailleurs constitutive d'une conception cardinale de l'utilité.

Formellement, une fonction d'utilité additive est de la forme $U(x, y) = V(x) + W(y)$ où x et y sont des quantités de biens. La conséquence principale de cette hypothèse est que les utilités marginales ne dépendent que de x et y[33].

La première critique centrale de cette conception de l'utilité est sans doute celle d'Irving Fisher (1867-1947) qui dans ses *Mathematical Investigations* remettait en cause la conception « psychologique » de l'utilité d'Edgeworth à savoir l'utilité conçue comme un accroissement tout juste perceptible de plaisir. Fisher substitue à cette conception de l'utilité une analyse en termes de choix. Si un individu choisit x par rapport à y c'est que x a une utilité plus grande que y pour l'individu. Plus précisément, les déterminants psychologiques du choix n'ont pas à intervenir dans l'analyse économique de celui-ci (Moscati [2004]).

Cette analyse de Fisher est proche de celle de Pareto lorsque celui-ci examinera les choix des consommateurs, tout en préservant l'analyse des courbes d'indifférence proposée par Edgeworth.

Plus précisément, deux approches relatives à l'analyse de la demande avaient été proposées par Pareto dans le *Manuel*. La première était une approche comportementale selon laquelle la théorie du consommateur devait être fondée sur le comportement de choix observable. La seconde approche, purement ordinale, revenait à l'idée d'un classement des préférences grâce à un index. C'est cette première approche qui sera conservée par des auteurs comme Hicks et Allen qui comme on le verra, initient un nouveau paradigme appelé à être dominant, fondé sur les travaux de Pareto.

Pareto considérait l'utilité à travers une notion d'ordre[34]et abandonnait ainsi l'idée d'une utilité cardinale – et donc les opérations mathématiques comme l'addition et la multiplication sur les utilités – et la remplaçait par un concept d'échelle de

[33] A cela, on peut ajouter que la matrice des dérivées secondes est diagonale et que Si les utilités marginales sont décroissantes, ses éléments diagonaux (dérivées secondes de V et W) sont négatives et on a donc les conditions du second ordre pour un maximum.

[34] Comme le souligne Granger [1960], « la grandeur n'était ici qu'un vêtement assez arbitraire, et […] seule subsistait, comme schématisation raisonnable et motivée de l'expérience, la structure d'ordre » (Granger [1960], p.134).

préférences. En effet, selon lui, la science économique est empirique de nature. Cependant, elle n'est pas empirique dans le même sens que le sont les sciences telles que la physique et la chimie, car celles-ci peuvent avoir accès à l'expérience alors que des sciences comme l'économie ou l'astronomie doivent se contenter de l'observation (Pareto [1909, 1966, p. 16). La seule possibilité qui reste accessible à l'économiste est une expérience de pensée dans laquelle on pourrait déterminer les courbes d'indifférence (Moscati [2004], p.4).
Certains auteurs comme Louis Léon Thurstone en 1930 tenteront de déterminer expérimentalement ces courbes d'indifférence avec un succès relatif, nous y reviendrons dans la section qui suit.

Dès lors, l'idée d'une fonction d'utilité additive ne constituait qu'une approximation qui, même si elle peut être commode, ne recouvre pas véritablement la réalité du phénomène et ne constitue pas un instrument utile au projet de Pareto[35]. Pour ce dernier, tout comme chez Irving Fisher (1867-1947) d'ailleurs, la méthode des courbes d'indifférence ne permet donc pas de déduire l'existence d'une fonction d'utilité, au sens cardinal du terme[36].

Cette conception de Pareto, qui rompt avec celle des marginalistes comme Jevons, Walras ou Edgeworth, ne sera pas sans implications sur leur représentation de l'utilité. C'est précisément ce qu'essaient de montrer Hicks et Allen dans un article écrit en 1934. Ils examinent plus précisément les ajustements dans la structure de la théorie marginale de la valeur-utilité que la proposition de Pareto rendrait nécessaires. Ils montrent en particulier que cette proposition permet de passer d'une théorie subjective de l'utilité propre aux marginalistes à une logique générale du

[35] Cependant, comme le soulignent Roberto Marchionatti et Enrico Gambino [1997, p.334-335], la position de Pareto n'a pas toujours été claire. En effet, selon eux, dans le *Cours d'économie politique* [1896], Pareto évoque la possibilité de mesurer l'utilité – ou plutôt l'ophélimité. Celle-ci est conçue comme une quantité. Tout comme Fisher, Pareto tente de trouver des mesures indirectes de l'utilité avec une mesure similaire à celle utilisée pour déterminer la longueur d'ondulations lumineuses à partir de phénomène optiques. Cependant cette hypothèse semble contradictoire avec la méthodologie expérimentale de Pareto. Une telle mesure n'est en effet pas accessible aux individus et n'est pas empiriquement observable. Cette position est ainsi abandonnée dans les *Ecrits de politique économique pure* [1900, 1982] où cette fois, Pareto évoque une conception purement ordinale. Cette fois le simple fait d'affirmer que l'on a un plaisir plus grand qu'un autre suffit à construire une théorie tout à fait satisfaisante.
[36] Pareto [1909], p.159.

choix (Hicks & Allen [1934, p. 54]) et ce, grâce au concept d'échelle de préférences évoqué précédemment.

2.1.2.2 Pareto et le passage à l'ordinalisme : l'article de Hicks et Allen [1934]

Pareto a non seulement mis en lumière l'importance du concept d'échelle de préférences, mais il est aussi parvenu à élaborer une théorie de l'utilité qui constitue, selon Hicks et Allen, un témoignage des méthodes utilisées par les économistes, et le paradigme dominant avant vNM.

Nous nous appuierons sur cet article de Hicks et Allen pour présenter ce paradigme et, ce, toujours dans l'objectif de mettre en lumière les bouleversements suscités par vNM.

Dans la lignée de Edgeworth, Fisher et Pareto, Hicks et Allen cherchent à établir une définition claire des concepts de la théorie de l'utilité de l'époque. Comme nous l'avons dit, la vraie révolution opérée par Pareto était pour eux de réussir à transformer une théorie subjective de l'utilité en une logique générale du choix, permettant ainsi d'étendre son applicabilité à de larges domaines (Hicks & Allen [1934a, p.54]). Hicks et Allen estimaient que la théorie « pure de la valeur d'échange » (Hicks & Allen [1934a], p.52), après une période de recherches intensives par les économistes de la génération de Jevons et Marshall, avait reçu relativement moins d'attention depuis le début du XXème siècle. Mises à part quelques enquêtes sur la dynamique du sujet, dues au Cercle de Vienne, une seule réalisation majeure – selon Hicks et Allen – avait été faite sur ce sujet depuis 1900. Il s'agissait du *Manuel* de Pareto (1909), ouvrage qui constituait une théorie statique de la valeur, et dont l'une des positions importantes était celle de l'incommensurabilité de l'utilité (Hicks & Allen [1934a, p.52]).

Pareto a en effet été l'un des plus importants porte-paroles des doutes relatifs à l'existence de fonctions d'utilité uniques et à la pertinence de telles fonctions pour comprendre le comportement économique. Indépendamment de Fisher, il a relevé le problème de l'existence d'une fonction d'utilité dès 1892. Peu après, la plus grande partie de sa théorie mathématique était développée. Dans le *Cours* (1896), Pareto

continuait ainsi à accepter les comparaisons interpersonnelles d'utilité pour des problèmes de bien-être. Dans le *Manuel* (1909), au contraire, l'utilité mesurable a sombré vers l'arrière plan (Stigler [1950], p.380).
Avec Pareto, la conception dominante de l'utilité est ordinale. vNM vont rompre avec cette tradition parétienne.

2.1.3. La construction d'une théorie de l'utilité espérée par von Neumann et Morgenstern : parcours intellectuel et enjeux théoriques.

L'ouvrage de von Neumann et Morgenstern (*Theory of games and economic behavior* [1947]**)** est une tentative pour renouer avec le principe de maximisation de l'utilité espérée. Notons que cette tentative de vNM se démarque de la perspective de Bernoulli sous plusieurs aspects
VNM proposent une axiomatique de décision, ce que ne fait pas Bernoulli.
La fonction d'utilité de Bernoulli concernait le cas certain alors que celle de vNM est relative aux situations de risque au sens défini plus haut.

En reprenant l'idée de l'utilité espérée de Bernoulli, vNM font un choix que les économistes qui les précèdent avaient rejeté ou négligé. Friedman et Savage expliquent d'ailleurs ce rejet dans un article écrit en 1948 qui constitue un état des lieux des avancées théoriques impliquées par l'ouvrage de vNM (tout comme celui de Hicks et Allen (1934) constitue un témoignage du revirement théorique proposé par Edgeworth). Selon Friedman et Savage, l'idée que les choix risqués peuvent être expliqués par la maximisation de l'utilité espérée fut rejetée du fait d'une « croyance » - enracinée chez les économistes – selon laquelle l'utilité marginale décroissante empêche d'expliquer les phénomènes de paris.
La fonction d'utilité correspondant à la proposition de Bernoulli est en effet concave pour respecter la loi de l'utilité marginale décroissante.
Pour Alfred Marshall (1842-1924), le principe de maximisation de fonction d'utilité concave ne permet dès lors pas d'expliquer uniquement les comportements d'individus ne prenant pas de risque car pour lui, « le jeu aboutit toujours à une perte

économique » (Marshall, *Principles of Economics*, vol. I, p. 112). Donc si un individu décidait de participer à un jeu ou pari, il perdrait.

Friedman et Savage expliquent que c'est l'aveuglément confiant dans la loi de l'utilité marginale décroissante qui a amené les économistes comme Marshall à rejeter le principe de maximisation de l'utilité espérée.

Pour Marshall, par exemple, l'analyse de l'utilité devait ainsi s'en tenir la à la technique des courbes d'indifférence, tout comme celle d'Edgeworth, Fisher et Pareto. Pour rationaliser les choix certains (choix pour des individus averses au risque), selon eux, il était suffisant de supposer que les individus pouvaient classer des paniers de biens selon leur utilité. Il n'était pas nécessaire de supposer qu'ils pouvaient comparer des différences d'utilité (comme nous l'avons vu dans la partie 2.1.2).

Pour von Neumann et Morgenstern, à partir de l'analyse en termes de courbe d'indifférence un tout petit effort seulement est nécessaire pour parvenir à une utilité numérique. Ils espèrent en effet montrer que l'on peut déduire l'utilité cardinalement. Afin de préciser les apports de VNM, nous commençons par revenir sur le parcours intellectuel des deux auteurs (2.1.3.1) et en particulier sur leur démarche spécifique, largement inspirée de la théorie physique, pour construire leur axiomatique (2.1.3.2).

2.1.3.1 Le parcours intellectuel de von Neumann et Morgenstern

- Parcours de von Neumann

John von Neumann (1903-1957) est né à Budapest. Très tôt, il manifeste des aptitudes exceptionnelles. Comme le souligne Pierre Richard Halmos [1937], mathématicien et disciple de von Neumann, il aurait compris à 12 ans le *Traité des Fonctions* d'Emile Borel[37]. Dissuadé par son père de faire des études de mathématiques pour des raisons financières[38], et sur les conseils de Theodore von

[37] Emile Borel, (1871-1956), est un mathématicien constructiviste et le fondateur de la théorie de la mesure et de l'étude moderne des fonctions.
[38] Pierre Richard Halmos, *ibid.*, page 383.

Kármán[39], il entreprend des études de chimie à Berlin puis à Zurich de 1921 à 1925. Il obtient en 1926 un diplôme d'ingénieur en chimie et une thèse en mathématiques sur l'axiomatisation des ensembles théoriques à Budapest. Après sa thèse, par ailleurs passionné de l'histoire de Byzance, de Jeanne d'Arc et de la guerre civile américaine, von Neumann devient Privatdozent à Berlin de 1926 à 1929, puis à Hambourg de 1929 à 1930. Il travaille alors sur la physique quantique et la théorie opératoire.

Ses travaux en physique sont unanimement salués dès les années 1930. Leur influence sur sa manière d'aborder les jeux stratégiques en économie est d'ailleurs maintes fois soulignée dans l'ouvrage qu'il a écrit avec Morgenstern. Entre 1922 et 1927, von Neumann publie près de vingt articles majeurs en mathématiques, travaux marqués d'une double approche : une étude des fondations des mathématiques et une tentative d'axiomatisation de la physique mathématique.

En 1926, il établit son théorème du minimax qui sera publié sous forme d'article en 1928. Pour Robert Leonard [1995], ce travail est davantage le résultat d'une réflexion commune avec ses pairs qu'un moment isolé de son inspiration (Leonard [1995], p.732). Le théorème du minimax cherche une réponse au problème de Pierre Rémond de Montmort (1678-1719) concernant un jeu de duel. Ce dernier soulignait dans son *Essai d'analyse sur les jeux de hasard* (1713) que ces questions étaient très simples mais insolubles, ce qui était selon lui très dommageable étant donné le nombre de situations de jeux de duel dans la vie quotidienne. Le théorème du minimax se propose de trouver au moins un équilibre de stratégies mixtes (c'est-à-dire une distribution de probabilités affectée par un joueur à l'ensemble des stratégies pures, ces dernières correspondant à des variables certaines) pour tout jeu à somme nulle et à deux joueurs qui ont fait leurs choix dans des ensembles finis de stratégies pures.

Von Neumann n'est pas le premier à envisager ces problèmes. Emile Borel, décrit par Christian Schmidt [2001] comme l'un des initiateurs de la théorie des jeux mais aussi l'un de ses représentants les plus sceptiques, est parmi les premiers à avoir appliqué le calcul des probabilités pour déterminer la solution de certains jeux de

[39] Theodore von Kármán (1881-1963), prodige hongrois des mathématiques, est l'un des pionniers modernes de la mécanique des fluides et de l'aéronautique.

société (Schmidt [2001], p.96). Mais là où Borel cherchait une solution « accessible aux joueurs », von Neumann espére trouver une solution purement théorique (Schmidt [2001], p.98).

Leonard souligne que le contexte historique dans lequel von Neumann a évolué est déterminant pour comprendre l'enjeu de ses travaux et leur origine. Au début du XXème siècle, un dialogue s'était noué entre les mathématiciens d'origine allemande et hongroise à propos du programme d'Hilbert[40], tentative d'établir les mathématiques sur des bases axiomatiques sûres[41], et au sujet d'une seconde tentative qui consistait à montrer comment la formalisation mathématique pouvait constituer un outil d'explication dans plusieurs domaines ? Von Neumann était au fait des travaux de l'époque concernant l'axiomatisation et plus généralement l'ère du formalisme mathématique qui voyait le jour. Il prendra très vite la suite de Zermelo[42], fondateur, selon la typologie de Schmidt, des jeux de positions axés sur la théorie des ensembles[43]. Selon Leonard : « Ce sont les relations entre un ensemble théorique et des jeux de société qui ont formé la base intellectuelle du travail de Von Neumann sur le théorème du minimax » (Leonard [1995], p.734). La plupart des mathématiciens de cette époque pensaient que les mathématiques permettaient de pénétrer la « psychologie du jeu »[44].

Mais un contact véritable avec la science économique n'est pas encore établi.

---

[40] David Hilbert (1862-1943) est une figure emblématique des mathématiques. Il inaugure notamment la méthode axiomatique à l'Université de Göttingen en donnant une formulation rigoureuse de la géométrie euclidienne.

[41] Programme dont on peut retracer la trame dans la découverte de paradoxes et d'antinomies par Georg Cantor (1845-1918) et Bertrand Russell (1872-1970). Von Neumann fait parti de ces mathématiciens voulant « nettoyer » l'ensemble théorique de Cantor pour établir une base axiomatisée solide, fondée sur un nombre limitée de postulats.

[42] Ernest Zermelo (1871-1953), philosophe et mathématicien allemand, a notamment pris part au programme de recherche lancé par Hilbert en appliquant aux ensembles ordonnés son axiome de choix. Il mit en évidence une solution au jeu d'échecs raisonnant en termes de position de joueurs, « il ne cherche pas davantage un équilibre mais s'intéresse aux conditions logiques permettant à un joueur de gagner s'il se trouve en position gagnante, et de reporter se défaite s'il se trouve en position perdante » (Christian Schmidt, *ibid*., page 85).

[43] Christian Schmidt explique en détails les différents courants de la théorie des jeux avant la parution de *Theory of games and economic behaviour* (*ibid.*, page 244).

[44] Robert J. Leonard, *ibid.*, page 733.

- Von Neumann et l'analyse économique

Si, muni de ce bagage mathématique, vN investit la théorie économique, il ne choisit pas de contribuer à n'importe laquelle. Comme Davidson, par la suite, il s'intéresse à une théorie économique particulière, la théorie de la décision individuelle.

Selon Nicholas Kaldor, von Neumann a exprimé un intérêt pour la science économique dès 1927. Kaldor aurait d'ailleurs conseillé à Von Neumann la lecture de Knut Wicksell, plus précisément de l'ouvrage *Value, Capital and Rent*, ce qui conduisit Von Neumann à connaître Walras mais aussi la théorie du capital de Böhm-Bawerk. Dès 1932, von Neumann présente son modèle de croissance économique linéaire, modèle publié en 1937.

Un autre événement a compté dans la sensibilisation de von Neumann à la science économique : il s'agit de la lecture de l'ouvrage du théoricien français Georges Guillaume, *L'économie rationnelle*, à la suite du conseil d'Abraham Flexner, directeur de l' « Institute for Advanced study » de Princeton, que von Neumann avait rejoint en 1930, comme beaucoup d'autres intellectuels européens fuyant le nazisme. Von Neumann explique à Flexner : « Je pense qu'en dépit de bonnes et retentissantes idées sur la méthodologie, la technique mathématique des auteurs n'est pas assez bonne pour prendre en compte toutes les structures théoriques et statistiques ».[45]

Puis, en pleine Seconde Guerre mondiale, von Neumann, alors membre de l' « Atomic Energy Commission » voulue par le président Eisenhower, est chargé de réfléchir, au sein des Nations Unies, aux effets de la bombe atomique[46]. A cette même période, il trouve les ressources nécessaires pour s'intéresser à la théorie des jeux et ses applications en économie. Von Neumann innove encore et propose de remplacer les anciens outils mathématiques utilisés à l'époque en économie mathématique (essentiellement le calcul des variations) par d'autres, plus nouveaux et inédits, les combinatoires et la convexité[47]. Mais le passage d'une réflexion sur les

[45] Robert J. Leonard, *ibid.*, page 737.
[46] Pierre Richard Halmos, *ibid.*, page 391.
[47] Pierre Richard Halmos, *ibid.*, page 392.

mathématiques des jeux de société à une démarche globale concernant la théorie des jeux n'a pu se faire qu'avec la rencontre avec Morgenstern.

- Morgenstern : construire une théorie prédictive ?

La rencontre avec Morgenstern sera déterminante. Oskar Morgenstern (1902-1976) est né en Allemagne, en Silésie. Après une thèse sur la productivité marginale soutenue à Vienne, Morgenstern devient dans cette même université Privatdozent, et succède à Hayek comme directeur du « Vienna's institute for Business Cycle Research » et ce jusqu'en 1938, date de l'Anschluss. Par la suite, Morgenstern est l'un des membres permanents de la faculté de Princeton aux Etats-Unis, jusqu'en 1970, date à laquelle il s'installera jusqu'à sa mort à New-York. Pendant ses années en Autriche, il travaille essentiellement sur le cycle des affaires, et prône l'introduction du temps et de l'incertitude dans le modèle d'équilibre général. Il connaît alors de multiples influences. Leonard en souligne au moins deux qu'il considère comme déterminantes : celle d'Othmar Spann (1878-1950) et celle de Hans Mayer. Spann avait été l'étudiant de Carl Menger auquel il s'opposa sur le plan de la « politique économique libérale » (Leonard [1995], p. 740). Sous l'influence d'Adam Müller (1779-1829), Spann développa une doctrine de l' « universalisme » en opposition à l'individualisme comme catégorie centrale de la pensée moderne. Sa perspective populiste est notamment imprégnée de nationalisme allemand. Si, comme le dit Léonard, Morgenstern est attiré par ce versant antilibéral et antisémite, cet aspect sombre de sa vie ne semble pas entacher sa relation avec von Neumann.
Morgenstern s'intéresse aussi à l'idéalisme philosophique de Fichte, Hegel et Schelling. Suite à certains désaccords avec Spann, Morgenstern se tourne vers Hans Mayer, un assistant de Friedrich von Wieser travaillant notamment sur la possibilité d'incorporer le temps au sein de la théorie de l'équilibre. Morgenstern est très impressionné par l'œuvre de Wieser, et lui consacrera dès 1927, un article commémorant sa mort (Morgenstern [1927]). A cette période, Morgenstern voyage beaucoup. Il rencontre Edgeworth à Oxford, et assiste au séminaire d'Alfred Whitehead. L'influence de la logique mathématique sur Morgenstern est indéniable comme l'auteur le souligne lui-même : « J'ai été énormément influencé par les

travaux de Hermann Weyl, Bertrand Russell et d'autres en mathématiques et physiques. J'ai aussi lutté avec le *Tractatus Logico-Philosophicus* de Ludwig Wittgenstein. »[48] Morgenstern publie son premier ouvrage en 1928, *Wirtschaftprognose* (littéralement la « prédiction économique »). Ce qui marque les esprits à propos de cet ouvrage c'est notamment un certain scepticisme voire un certain pessimisme (Leonard [1995], p. 741). Selon Morgenstern, les sciences sociales ont la possibilité d'affecter leur objet d'étude. Contrairement aux sciences de la nature, elles peuvent influencer le cours des événements. Toute prédiction est donc un leurre. Morgenstern prend part au « Privatseminar » inauguré par Ludwig von Mises où il côtoie entre autres Friedrich von Hayek, Fritz Machlup, et Paul Rosenstein-Rodan. Moins fréquemment, mais avec autant d'intérêt, il participe au « Karl's Menger Mathematical Colloquium » où Gödel et Wald sont notamment présents.

- Le rôle de la formalisation mathématique chez Morgenstern

Morgenstern est également très influencé par Karl Menger, fils de Carl. Karl Menger avait discuté très tôt le paradoxe de Saint Pétersbourg dans son article de 1934, « Das Unsicherheitsmoment in der Wertlehre. Betrachtungen in Anschluss an das sogenannte Petersburger Spiel ». L'une des critiques qu'adresse Menger à la résolution de ce paradoxe est le manque de précision dans les outils formels utilisés (Leonard [1995], p. 744). Il insiste plus généralement sur les limites de la formalisation concernant le comportement humain, et plus particulièrement le comportement de pari. Morgenstern, d'après Leonard, a pris très au sérieux les remarques de Menger (*ibid.*) et l'influence de ce dernier est évidente notamment lors des discussions de Morgenstern avec von Neumann à Princeton dans les années 1940. On y sent en particulier l'influence sur Morgenstern de l'ouvrage de Menger, *Morality, Decision and Social Organization* de 1934. Dans celui-ci, Menger s'interroge sur les implications logiques d'une réunion de groupes sociaux étant

[48] Oskar Morgenstern, The collaboration between Oskar Morgenstern and John von Neumann on the Theory of Games, *Journal of Economic Literature*, vol. 14, n°3, septembre 1976, p. 805.

données les différentes attitudes des individus et un ensemble de règles hypothétiques de comportement ? (*ibid.*).

Il existe aussi une correspondance entre Morgenstern et Frank Knigt peu après la parution de l'article « The time moment in Economic Theory » (Knight, 1934). En 1939, alors que Morgenstern est à Princeton, il adresse à son correspondant, une vive critique au *Traité de la monnaie* de John Maynard Keynes notamment sur le plan des anticipations et des prévisions.

Mais tous ces éléments ne suffisent pas, semble-t-il, à produire l'étincelle de la collaboration avec von Neumann. Leonard conçoit celle-ci comme un processus, plus qu'une conséquence inévitable (Leonard [1995], p. 747). Jusqu'en 1941, d'après les nombreuses correspondances de Morgenstern, ce dernier n'était pas sûr que ses entrevues avec des mathématiciens donneraient des résultats. L'impact de von Neumann sur lui s'est, cependant, vite fait sentir, notamment dans la « Review of Hicks 1939 *Value and Capital* » au cours de cette même année. Il faudra attendre la rupture idéologique de Morgenstern (celui-ci axant ses recherches sur le problème de l'interdépendance des individus au niveau de leurs choix) avec Menger pour voir se dessiner les contours d'une collaboration féconde.

Toutefois le rôle de chacun des auteurs dans l'ouvrage de 1944 peut être questionné. Selon Léonard, « tous les aspects techniques de la théorie doivent être mis au crédit de von Neumann », mais alors quelle fut la contribution de Morgenstern ? Morgenstern souligne de lui-même les multiples questions qu'il posait à von Neumann, toutes plus « intéressantes et provocantes » [49]. Mais son apport est plus profond. C'est la personnalité de Morgenstern, et tout particulièrement son insatisfaction perpétuelle, selon les termes de Leonard, qui fut véritablement décisive dans les débats qu'il eut avec von Neumann. Cet aspect de la personnalité de Morgenstern apparait vivement dans la critique sous-jacente à l'ouvrage du modèle dominant de l'époque, le modèle Hicks-Samuelson. Christian Schmidt partage cette opinion mais de manière plus équivoque. Il montre que Morgenstern n'a jamais saisi les possibilités de rapprochement entre la théorie des jeux et l'analyse économique

[49] Pour les deux citations, (Leonard [1995], p. 753).

par le biais de la théorie de l'équilibre[50]. Malgré cela, Schmidt rappelle le poids de l'économiste allemand : c'est lui qui insiste pour incorporer une « préhistoire de la discipline »[51] et qui, par son impatience, et sa « disposition critique » à l'égard des autres traditions économiques, a rendu possible l'achèvement d'une rencontre entre deux continents théoriques : la problématique mathématique des jeux, et la réflexion économique sur les interactions des choix de décideurs rationnels (Schmidt [2001], p.377).

2.1.3.2 De la physique à la théorie de la décision : le parallélisme entre centre de gravité et utilité

Comme nous l'avons mentionné dans le paragraphe (2.1.3.1), la référence à la physique est l'un des éléments centraux de l'ouvrage de vNM. Pendant longtemps, selon eux, les physiciens ont émis des réserves sur la possibilité d'attribuer une mesure quantitative à la chaleur, réserves qui paraissent rétrospectivement caduques tant la théorie moderne fait usage de telles mesures. Selon vNM, cette trajectoire sera probablement celle de la théorie de l'utilité.
Nous identifions chez eux au moins trois manières d'utiliser les sciences physiques pour construire leur théorie de l'utilité.

- La notion d'utilité, notion physique ?

L'analogie entre les théories physiques, qui nécessitent une mesure de quantités, et le principe de mesurabilité en théorie économique est le point de départ des auteurs de la *Theory of Games and Economic Behavior* (vNM [1947], p. 3).
Ainsi vNM considèrent-ils que l'utilité s'apparente à une notion physique (vNM [1947], p.15), qui peut être traitée comme une quantité mesurable. Pour justifier cette position, ils avancent deux arguments : le premier est qu'historiquement, l'utilité a été conçue comme quantitativement mesurable, c'est-à-dire comme un nombre ; le second est que toute tentative de mesure doit être basée ultimement sur une sensation

[50] Christian Schmidt, *ibid.*, page 208.
[51] Christian Schmidt, *ibid.*, page 376.

immédiate - tout comme en physique on évoque la sensation de chaleur, de lumière – et dans le cas de l'utilité, c'est la sensation immédiate de préférence (d'un objet ou d'une agrégation d'objets par rapport à un autre). Cependant, cela ne permet pas de procéder à des comparaisons numériques d'utilités pour une personne ni à une comparaison entre des personnes selon les auteurs (vNM [1947], p.17).

Cette position rapproche la méthodologie des sciences de la nature de celle des sciences de l'homme[52] et pose, de ce fait, toute une série de questions, relatives à l'interprétation d'un tel rapprochement, que l'on analysera notamment dans la partie 2, lorsque Davidson évoque les difficultés d'interprétation qu'il a rencontrées lorsqu'il était théoricien de la décision. En effet, comme le souligne Granger [1988], nous sommes en droit de nous demander à quelles conditions une structure mathématique s'applique à un domaine phénoménal (Granger [1988], p.254). Cette question a des répercussions sur le niveau expérimental et c'est précisément ce que mettra en évidence Davidson dès 1957 puis en 1980.

- Des opérations sur les grandeurs aux opérations sur l'utilité

Outre la mesure de l'utilité, les auteurs utilisent l'analogie avec la physique pour insister sur les opérations que l'on peut réaliser sur l'utilité.

Selon les auteurs, on rencontre en effet dans les sciences des grandeurs qui ne sont pas *a priori* « mathématiques » mais qui sont, sous certains aspects, rattachées au monde physique. Occasionnellement, ces grandeurs peuvent être groupées dans des domaines où des opérations physiques qu'ils qualifient de « naturelles » sont possibles. La quantité physique de poids, par exemple, permet l'opération d'addition, tout comme la distance. Cela ne veut pas dire, selon les auteurs, que les deux opérations, qui ont le même nom, sont identiques (vNM [1947], p.21). Cela signifie seulement qu'elles ont des traits similaires (il s'agit de mesure d'entités physiques) et que l'on espère faire des correspondances entre elles (c'est-à-dire trouver un ensemble de coordonnées qui rendent les deux échelles compatibles). Pour cela, il

[52] Selon Granger [1988], « depuis la réduction galiléo-cartésienne [dans les sciences de l'homme], les problèmes de mesure sont toujours ramenés plus ou moins indirectement aux trois catégories de la longueur, du temps, de la masse » (p.255).

faut trouver des modèles mathématiques à l'intérieur desquels ces quantités sont définies par des nombres. A partir de ce moment, l'addition devient une addition ordinaire. Il n'est cependant pas sûr que la description du modèle mathématique fournisse une manière unique de relier des quantités physiques à des nombres. Il pourrait exister une famille entière de ces applications. Le passage d'une application à une autre s'appelle une transformation. On dit alors que dans cette théorie, les quantités physiques sont décrites pas des nombres relativement à un système de transformations. Il serait même concevable de dire qu'une quantité physique est un nombre à une transformation monotone près. C'est le cas des quantités pour lesquelles seule la relation « naturelle » « est plus grand que » existe et rien d'autre. C'est le cas pour la température par exemple mais aussi pour *l'utilité* quand elle est basée sur l'idée conventionnelle de préférence, selon les auteurs.
Ainsi peut-on peut considérer que la seule donnée naturelle dans le domaine de l'utilité est la relation « est plus grand que », c'est-à-dire le concept de préférence. Dans ce cas, les utilités sont numériques à une transformation monotone croissante près. Et c'est ce qui est généralement accepté par la littérature et exprimé par la technique des courbes d'indifférence.

- Du concept physique de centre de gravité à la mesure de l'utilité

vNM poursuivent l'analogie en transposant la méthodologie utilisée en physique pour déterminer le centre de gravité à la théorie de la décision *dans le risque.*
La quantité physico-géométrique de « position » ne permet pas l'opération d'addition mais elle permet l'opération consistant à former un « centre de gravité » de deux positions dans l'espace. Ils choisissent l'exemple de « positions » dans un espace à trois dimensions où celles-ci sont, comme les autres quantités vectorielles, considérées en fonction de triplets de nombres appelés des coordonnées. L'opération « naturelle » de « centre de gravité » de deux positions $\{x_1, x_2, x_3\}$ et $\{x_1', x_2', x_3'\}$ avec les « masses » α, 1-α devient $\{\alpha\, x_1 + (1-\alpha)\ x_1', \alpha\, x_2 + (1-\alpha)\ x_2', \alpha\, x_3 + (1-\alpha)\ x_3'\}$.
VNM transposent cette opération pour les utilités. Soient deux utilités, *u*, *v* les auteurs utilisent la relation « naturelle » $u > v$ qui signifie *u* est préféré à *v* et

l'opération « naturelle » $\alpha u + (1- \alpha) v$ (avec $0 < \alpha < 1$) qui signifie : le centre de gravité de *u*, *v* avec les poids respectifs α, 1-α ; ou la combinaison de *u*, *v* avec les probabilités α, 1-α. Si l'on conçoit l'existence de ces concepts, il s'agit de trouver une correspondance entre les utilités et les nombres qui porte la relation $u > v$ et l'opération $\alpha u + (1- \alpha) v$ pour les utilités.
C'est donc à partir de cette méthodologie empruntée à la physique que vNM décrivent comment mesurer les utilités.

En partant de l'analogie avec la physique, vNM sont en mesure de construire une théorie axiomatisée de l'utilité dans le risque. S'agissant de la mesure de l'utilité – même si l'ouvrage n'a pas trait à l'analyse de celle-ci – une procédure simple permet d'obtenir une estimation des différences d'utilité qu'accorde un individu à trois événements A, B et C ; chacun de ces événements étant combinés à des probabilités.
Les auteurs considèrent un individu qui, muni d'un préordre complet sur tous les objets de son ensemble de choix, est amené à comparer non seulement des événements mais aussi des combinaisons d'événements associés à des probabilités.
Si l'on note deux événements B et C ayant tous la probabilité de 50%, la combinaison consiste alors en la perspective de voir B advenir avec la probabilité de 50% et C (si B ne se produit pas) avec la même probabilité. Les deux issues étant mutuellement exclusives. On suppose alors que l'individu a l'intuition de sa préférence entre l'événement A ou la combinaison d'événements B ou C, ou l'inverse. Dès lors, s'il préfère A à B et A à C, il préférera A à toute combinaison de B et C. Mais s'il préfère A à B et en même temps C à A, la comparaison entre A et la combinaison de B et C apporte une nouvelle information. En effet, si maintenant il préfère A à la combinaison de B et de C ayant chacun une chance sur deux de se réaliser, alors cela permet d'estimer numériquement que sa préférence de A par rapport à B excède celle de C par rapport à A[53]. Si ce point est acquis, il existe un critère avec lequel comparer la préférence de C par rapport à A et la préférence de A par rapport à B. Les utilités, ou les différences d'utilités deviennent numériquement

[53] Il n'y ici pas de pétition de principe car comme le soulignent les auteurs, on ne postule pas de prime abord l'existence d'une échelle de mesure numérique de l'utilité (Von Neumann et Morgenstern [1947], p.20).

mesurables. Cet exemple leur suggère, en effet, une méthode permettant de mesurer directement les utilités.

Pour accéder à cette mesure, on peut procéder comme suit : si l'on considère trois événements C, A et B avec les mêmes préférences évoquées ci-dessus et que l'on considère un nombre réel α compris entre 0 et 1 tel que A est désiré tout autant que la combinaison d'événements consistant en l'événement B avec la probabilité 1-α et l'événement C avec la probabilité α. Les auteurs suggèrent d'utiliser α comme une estimation numérique du rapport de la préférence entre A et B et entre C et B.

Cette procédure a l'avantage de pouvoir être exprimée dans les termes de l'analyse par les courbes d'indifférence. Ainsi, si l'on note q le rapport des utilités entre le fait de posséder une utilité du bien et le fait d'en posséder deux, on peut proposer à l'individu le choix entre obtenir une unité du bien, ou tenter d'obtenir deux unités avec la probabilité α et rien avec la probabilité 1-α. S'il préfère la première possibilité, alors $\alpha < q$; s'il préfère la seconde, alors $\alpha > q$; et s'il ne peut pas établir de préférence entre les deux, alors $\alpha = q$. C'est précisément cette méthode qui va être utilisée pour construire l'axiomatique de l'utilité espérée.

2.1.4. L'axiomatique de von Neumann et Morgenstern.

La particularité de l'axiomatique de vNM est qu'elle est largement imprégnée de l'analogie avec la physique mentionnée plus haut comme cela se voit dans la formulation première de leur axiomatique par vNM (voir encadré). Toutefois les notations mathématiques à l'époque choisies par vNM pour formuler leurs axiomes sont pour le lecteur d'aujourd'hui peu claires. Nous présenterons donc dans un premier temps la formulation moderne des axiomes de manière à disposer d'une présentation suffisamment conventionnelle pour nous permettre de comparer la théorie de vNM à celle de Savage par exemple.

Le point de départ de l'axiomatisation de vNM est la théorie ordinale des préférences.

Comme on l'a signalé plus haut, leur présentation des postulats et axiomes de la théorie peut sembler archaïque au théoricien de la décision moderne tant les objets et

les opérations de la théorie manquent de raffinement. Là où les partisans d'une conception ordinale des préférences utilisaient comme éléments de base des préférences sur des objets, vNM – du fait de l'analogie mentionnée plus haut, entre théorie physique et théorie économique – considèrent que la relation de préférence portent sur des utilités et c'est précisément ce que va remettre en cause la théorie moderne, lorsqu'elle reformulera la relation de préférence de vNM en la faisant porter non pas sur des utilités mais sur des loteries.

Pour exposer cette présentation moderne, nous avons choisi de reprendre celle de Luce et Raiffa [1957, 1985][54] particulièrement claire et précise[55].

Selon Luce et Raiffa, on peut partir de l'idée qu'un individu dispose d'un certain classement de préférences entre des issues (que ce soient des gains monétaires, ou des récompenses d'un jeu) A, B, et C tel que A est préféré à B, B à C et A à C. Un pari risqué consisterait à proposer à l'individu de choisir entre l'option 1 ayant pour résultat d'obtenir B pour sûr et l'option 2 revenant à un pari dont les deux issues sont A et C pondérée respectivement par les probabilités *p* et 1-*p*. On demande donc au sujet de comparer et de décrire ses préférences entre une option certaine et une loterie (c'est-à-dire une option revenant à deux issues mutuellement incompatibles pondérées par des probabilités relatives à leur obtention). Comme le soulignent Luce et Raiffa [1957, p. 21], il semble clair que plus *p* se rapprochera de la valeur 1, c'est-à-dire de la chance maximale, plus l'individu sera tenté de choisir l'option 2, et symétriquement, plus *p* se rapprochera de 0, plus l'individu sera tenté de choisir l'option 1.

Le pari est généralement considéré comme le point de départ des problèmes de prise de décision en situation de risque (Luce et Raiffa [1957, 1985], p. 19). Le pari permet, en effet, non seulement d'avoir accès à l'évaluation des issues par l'individu mais aussi à la manière dont il les évalue dans une situation particulière (Luce et Raiffa [1957, 1985], p.21)[56].

---

[54] Nous nous appuierons sur la seconde édition de 1985.

[55] Bien qu'il existe évidemment d'autres reformulations modernes comme celles de Herstein et Milnor [1953], celle de Jensen [1967] ou encore celle de Tversky [1975].

[56] Comme le souligne Emmanuel Picavet [1996], cette référence au pari remonte au moins à Kant qui dans la *Critique de la raison pure* avançait l'idée que la « pierre de touche communément employée pour déterminer si quelque chose que quelqu'un affirme est une simple persuasion, ou du moins une conviction subjective, c'est-à-dire une croyance solide, est le *pari* » (Kant [1781, 2001], pp.669-670).

Cette représentation des préférences sur des issues dont certaines sont des loteries constitue le point de départ et la condition nécessaire pour attribuer, dans un second temps, des utilités, et donc des nombres à ces issues.

L'objectif de la théorie de la prise de décision (*decision making*) en situation risquée est de présenter un certain nombre d'hypothèses (sous formes d'axiomes)[57] permettant non seulement de représenter un modèle « idéalisé » des préférences – et donc de présenter les conditions de cohérence et de rationalité - mais aussi de préciser les modalités de la représentation des préférences par des utilités numériques.

La première hypothèse (axiome) H1 est relative aux choix d'un individu entre des paires de tickets de loteries[58] notés $L = (p_1A_1, p_2A_2,\ldots, p_rA_r)$ et $L' = (p_1'A_1, p_2'A_2,\ldots, p_r'A_r)$ où $\{p_1, \ldots, p_r\}$ représentent l'ensemble des probabilités et $\{A_1, ..., A_r\}$ l'ensemble des lots.

Comme le soulignent Luce et Raiffa, si L est préférée à L', alors l'individu préfère « l'expérience associée » à la loterie L à celle de la loterie L'.

Le symbole $\gtrsim$ est utilisé pour représenter la préférence faible (c'est-à-dire la situation ou soit l'individu préfère A_i à A_j soit la situation où l'individu est indifférent entre les deux) d'un individu.

H1 (hypothèse d'ordre). Pour tout A_i à A_j dans l'ensemble des lots, l'une de ces relations est valable : soit $A_i \gtrsim A_j$, soit $A_j \gtrsim A_i$. Cette relation de préférence est transitive : si soit $A_i \gtrsim A_j$ et soit $A_j \gtrsim A_k$ alors soit $A_i \gtrsim A_k$.

Cette hypothèse d'ordre permet de relier les débats relatifs à la décision en situation risquée à la conception ordinale dominante avant la publication de l'ouvrage de vNM en 1944. Toutefois, vNM n'évoquent ni un ensemble de résultats ni la question de l'indifférence comme on le verra dans l'encadré 1.

---

[57] Luce et Raiffa [1957, 1985], p. 24.

[58] « Un ticket de loterie est un mécanisme de probabilité (*chance mechanism*) produisant des lots représentés par des résultats ayant une certaine probabilité » (Luce et Raiffa [1957, 1985], p. 24). Autrement dit, un ticket de loterie offre la possibilité de participer à un pari offrant des résultats conditionnés à des probabilités connus à l'avance (dans le cas présent) – ou perspectives aléatoires ; ce pourquoi on qualifie ces résultats de risqués dans la mesure où nul ne peut avoir l'assurance que tel résultat va se produire.

Si l'on considère les loteries $L^{(1)}$, $L^{(2)}$,…, $L^{(s)}$ composées les lots A_1, A_2, ..., A_r ainsi que des nombres réels non négatifs q_1, q_2, …, q_s dont la somme égale 1, alors (q_1 $L^{(1)}$, q_2 $L^{(2)}$,…, q_s $L^{(s)}$) représentent une loterie composée dont les lots sont eux-mêmes représentées par des loteries.

H2. Axiome de loteries composées

Toute loterie composée est indifférente à une loterie simple dont les lots sont A_1, A_2, ..., A_r dont les probabilités sont calculées selon le calcul ordinaire des probabilités. En particulier, si

$L^{(i)} = (p_1^{(i)} A_1, p_2^{(i)} A_2, \ldots, p_r^{(i)} A_r)$ pour i = 1, 2, …., s,

Alors (q_1 $L^{(1)}$, q_2 $L^{(2)}$,…, q_s $L^{(s)}$) ~ ($p_1A_1, p_2A_2, \ldots, p_rA_r$), avec $p_i = q_1p_i^{(1)} + q_2p_i^{(2)} + \ldots + q_sp_i^{(s)}$.

Autrement dit, cette hypothèse rend possible la réduction de loteries composées à des loteries simples. Cet axiome est explicitement présent chez vNM (axiome (Cb) voir encadré). Même si certains auteurs comme Samuelson [1952] considèrent que c'est un axiome purement technique qui relèverait uniquement de l'algèbre et non du comportement humain (Samuelson [1952], p. 671), les implications de cet axiome pour la théorie sont pourtant significatives.

En effet, comme le soulignent Luce et Raiffa, cet axiome permet de faire abstraction de tout « plaisir dans le jeu » (*joy in gambling*), d' « atmosphère de jeu » (*atmosphere of the game*) et de « plaisir du suspense » (*pleasure in suspense*) puisque cette réduction des loteries composées à des loteries simples signifie que l'individu est indifférent au fait de jouer une fois ou plusieurs fois (Luce et Raiffa [1957, 1985], p. 26).

Cette hypothèse est donc essentielle si l'on se place dans une perspective expérimentale de la théorie puisque les choix proposés aux sujets sont généralement des choix répétés comme dans les expériences de Davidson et Suppes [1957].

H3. Axiome de continuité

Chaque lot A_i est indifférent à un ticket de loterie impliquant A_1 et A_r c'est-à-dire qu'il existe un nombre u_i tel que $A_i \sim [u_iA_1, (1-u_i) A_r]$.

H4. Axiome d'indépendance

Pour toute loterie L, si $A_1 \sim A_r$ alors, $[\alpha A_1, (1-\alpha)\ A_2] \sim [\alpha A_r, (1-\alpha)\ A_2]$.

Autrement dit, si une loterie est jugée indifférente à une autre, le fait de combiner chacune d'elle par une même tierce loterie ne changera pas la relation initiale d'indifférence.

Cet axiome n'est pas explicitement présent dans l'axiomatique de vNM (voir encadré). La présentation moderne de cet axiome semble être initialement apparue dans l'article de Marschak [1950]. L'axiome existe aussi dans l'article de Friedman et Savage [1952] sous le nom de troisième postulat (p.358) bien que les auteurs ne lui donnent pas le nom d'axiome d'indépendance. La présentation de cet axiome sous l'expression « axiome d'indépendance » se trouve toutefois dans l'article de 1952 de Samuelson qui évoque lorsqu'il présente son deuxième axiome, l' « indépendance forte ». Malinvaud [1952] montrera le caractère implicite de cet axiome dans l'axiomatique de vNM.

A partir de ces quatre hypothèses, il est possible de montrer qu'il existe une fonction $u(\cdot)$ à valeurs réelles sur l'ensemble des loteries.

En effet, si un individu dispose d'une relation de préférences $\gtrsim$ transitive sur l'ensemble des loteries et si pour chaque loterie on peut assigner un nombre $u(L)$ tel que :

(a) $u(L) \geq u(L')$ si et seulement si $L \gtrsim L'$

Alors, nous pouvons dire qu'il existe une fonction d'utilité $u(.)$ sur l'ensemble des loteries et que cette fonction d'utilité a la propriété suivante :

(b) $u\ (L\ ,\ L'\ ;\ p\ ,\ 1-p) = pu(L) + (1-p)\ u(L')$

Autrement dit, elle est linéaire en probabilités.

La classe des fonctions d'utilité qui respecte ces conditions est définie à une transformation affine positive près telle que $v(x) = au(x) + b$ avec $a > 0$.

La présentation de l'utilité espérée de vNM avec les notations originales des auteurs(1947)

Les objets de la théorie, c'est-à-dire ceux sur quoi vont porter les différentes opérations issues de la théorie physique, sont donc les utilités elles-mêmes c'est-à-dire des objets physiques tout comme les poids.
vNM considèrent que la relation d'ordre[59], définie sur l'ensemble des utilités, est une relation binaire « naturelle », notée « > » au même titre que l'opération naturelle mentionnée plus haut – est complète et totale. La complétude est assurée par les deux opérations strictes « > » et « < »[60]. Cette relation est complète si l'on peut écrire :
(A) $u > v$ ou $v < u$.
vNM n'intègrent pas véritablement l'égalité qu'ils considèrent comme une vraie identité (vNM [1947], p. 617).
vNM considèrent que pour tous u, v une seule de ces trois relations est valable :
(A_1) $u = v$, $u > v$, $u < v$.
Les auteurs insistent aussi sur l'hypothèse de transitivité :
(A_2) $u > v$, $v > w$ implique $u >$ w.

La deuxième étape consiste à décrire les axiomes d'ordre et de combinaison (B) (vNM [1947], p.26) :

(Ba) $u < v$ implique que $u < \alpha u + (1-\alpha)v$[61] ;
(Bb) $u > v$ implique que $u > \alpha u + (1-\alpha)v$;
(Bc) $u < w < v$ implique l'existence d'un α tel que $\alpha u + (1-\alpha)\ v < w$;
(Bd) $u > w > v$ implique l'existence d'un α tel que $\alpha u + (1-\alpha)\ v > w$.

L'interprétation de ces axiomes est la suivante :
(Ba) : si v est préférée à u, alors la combinaison de u et de v pondérées respectivement par α et 1 – α est préférée à u ;
(Bb) cela correspond à l'axiome (Ba) avec la relation « moins préférée à » à la place de « préférée à » ;
(Bc) si w est préférée à u et que v est préférée aux deux autres, alors la combinaison de u avec la probabilité α et v avec la probabilité 1-α ne va pas affecter la relation de préférence par rapport à w à condition que la probabilité soit suffisamment faible ;
(Bd) même interprétation que (Bc) en remplaçant « moins préférée » par « préférée à ».
La littérature de la théorie de la décision qualifie (Bc) et (Bd) de « principe de la chose sûre ».

[59] vNM utilisent le terme « ordering » et non « preordering ». En effet, vNM postulent une relation de préférence stricte – en référence aux sciences physiques – ce qui explique pourquoi ils notent la relation de préférence avec les signes <,> et =. Par ailleurs, l'ordre des préférences portent sur les « utilités » - comprises comme des entités physiques comme la chaleur ou le poids.
[60] vNM [1947], p.26.
[61] On retrouve ici la référence à la physique évoquée plus haut.

Troisième étape, les axiomes de combinaison (C) :

(Ca) $\alpha u + (1-\alpha)\, v = (1-\alpha)\, v + \alpha u$.
(Cb) $\alpha\, (\beta u\ (1-\beta)\, v) + (1-\alpha)\ v = \gamma u + (1-\gamma)\, v$, où $\gamma = \alpha\beta$.

Ces axiomes, et tout particulièrement le dernier, rendent possible la réduction de combinaisons d'utilités pondérées par des probabilités (selon l'expression moderne des « loteries composées ») à de simples combinaisons d'utilités conditionnées à des probabilités (des loteries simples). Cette assimilation préserve le l'ordre complet valable pour les simples combinaisons d'utilités pondérées par α et (1-α).

Grâce à ces axiomes, les auteurs ont défini l'utilité numérique à laquelle on peut appliquer le calcul des espérances mathématiques.
L'objectif ultime est de montrer que les axiomes sur le comportement permettent d'élaborer un théorème de représentation des choix comme résultant de la maximisation de l'espérance mathématique des utilités.
La première étape consiste à trouver une correspondance entre les utilités et les nombres qui vérifie la double relation : on a pour les utilités u et v la relation « naturelle » $u > v$ (c'est-à-dire u est préféré à v) et l'opération « naturelle » $\alpha\, u + (1-\alpha)\, v$, $(0 < \alpha < 1)$ (lire : la combinaison de u, v avec les probabilités α, 1-α). On doit trouver une correspondance entre les utilités et les nombres qui *supporte* la relation $u > v$ et l'opération $\alpha\, u + (1-\alpha)\, v$.
La correspondance peut s'écrire :
$u \rightarrow \rho = \mathrm{v}(u)$; u étant l'utilité et v(u) le nombre qui lui est rattaché.
Les exigences sont les suivantes :

- $u > v$ implique que v (u) > v (v)
- $v\ (\alpha u + (1-\alpha)\ \mathrm{v} = \alpha\ \mathrm{v}\ (\mathrm{u}) + (1-\alpha)\ \mathrm{v}\ (v)$

Ceci, si les deux correspondances suivantes existent :

- $u \rightarrow \rho = \mathrm{v}\ (u)$,
- $u \rightarrow \rho' = \mathrm{v}'(u)$.

Dès lors, on a une correspondance entre les nombres qui peut s'écrire $\rho' = \Phi\ (\rho)$. Si les deux correspondances *fonctionnent*, il s'ensuit que les deux exigences évoquées sont satisfaites. En outre, si $\rho' = \Phi\ (\rho)$, cela conduit à la relation $\rho > \sigma$ qui permet l'opération $\alpha\rho + (1-\alpha)\ \sigma$, c'est-à-dire que les opérations sur les nombres suivent les opérations sur les utilités. D'où :

- $\rho > \sigma$ implique que $\Phi\ (\rho) > \Phi\ (\sigma)$
- $\Phi\ (\alpha\rho + (1-\alpha)\ \sigma = \alpha\Phi\ (\rho) + (1-\alpha)\ \Phi\ (\sigma)$, où $\Phi\ (\rho)$ est une fonction linéaire.

Si une évaluation numérique des utilités existe, alors elle est déterminée à une transformation monotone linéaire près.

(D) la fonction d'utilité

Von Neumann et Morgenstern définissent deux utilités certaines (non pondérées par des probabilités) u_0 et v_0. A partir de la relation $u_0 < v_0$, et pour tout w défini sur l'intervalle $u_0 < w < v_0$, on définit la fonction numérique $f\ (w) = f\ u_0, v_0\ (w)$ comme suit :

- $f\ (u_0) = 0$
- $f\ (v_0) = 1$

- $f(w)$, pour $w \neq u_0, v_0$ c'est-à-dire avec $u_0 < w < v_0$, est un nombre α dans l'intervalle $0 < \alpha < 1$ dans le sens où l'on a les deux opérations mentionnées plus haut : $u > v$ et $\alpha u + (1-\alpha) v$ ainsi que la relation $\alpha \rightarrow w = (1-\alpha) u_0 + \alpha v_0$.

On a, dès lors, la carte[62] $w \rightarrow f(w)$ qui a les propriétés suivantes :

- elle est monotone
- pour $O < \beta < 1$ et $w \neq u_0$, on a $f((1-\beta) u_0 + \beta w = \beta f(w)$
- pour $O < \beta < 1$ et $w \neq v_0$, on a $f((1-\beta) v_0 + \beta w) = 1- \beta + \beta f(w)$.

On trouve ici l'expression de la théorie de l'utilité espérée. Ce dernier met en relief l'existence d'une fonction *f* représentant la satisfaction de l'agent, aussi bien dans le cas certain que dans le cas risqué. Dans ce dernier cas, elle prend la forme d'une espérance mathématique d'utilité sur les gains possibles. Le fait d'utiliser une même fonction d'utilité témoigne d'une volonté de vNM d'intégrer l'analyse par les courbes d'indifférence au sein de leur propre schème. Cette idée sera reprise dans l'article de Friedman et Savage (1948) et fera même l'objet d'une tentative d'expérimentation par ces deux auteurs.

2.1.5 Friedman et Savage, une refonte du modèle de von Neumann et Morgenstern ?

Le modèle proposé par Davidson, Suppes et Siegel (1957) est une tentative de rendre la théorie de la décision empiriquement testable.

Pourtant, pour ce faire, ils n'utilisent pas uniquement le modèle de vNM, ils tiennent compte des évolutions et transformations que celui-ci a subies et s'inspirent en particulier des refontes proposées par Friedman et Savage en 1948 et 1952 dont la nouvelle axiomatique fait apparaître ce que l'on a appelé rétrospectivement l'axiome d'indépendance.

Ces deux articles vont influencer Davidson tant au niveau méthodologique que théorique. C'est pourquoi nous allons les présenter successivement. En 1948, Friedman et Savage donnent une place à la théorie de l'utilité espérée au sein de la discipline en présentant les applications de cette théorie au domaine de l'assurance par exemple (2.1.5.1). En 1952, ils précisent les fondements épistémologiques et

[62] Von Neumann et Morgenstern définissent une « carte » comme une corrélation entre une entité, ici une utilité, et un nombre, p. 22.

méthodologiques de la théorie de l'utilité espérée en produisant une nouvelle axiomatique permettant de faire apparaître de nouveaux enjeux (2.1.5.2).

2.1.5.1 Friedman et Savage (1948)

L'article de Friedman et Savage (1948) constitue une tentative pour valider empiriquement l'hypothèse d'utilité espérée de vNM. Plus précisément, il s'agit de rassembler des observations relatives au comportement de choix d'individus face à des issues risquées puis de vérifier si ces observations sont cohérentes avec ce que les auteurs appellent « l'hypothèse d'utilité espérée » de vNM pour enfin examiner les conséquences qu'impliquent ces observations sur la forme de la courbe qui représente la fonction d'utilité notamment en termes d'aversion ou d'amour pour le risque.

L'intérêt de cet article pour l'analyse du modèle de Davidson (1957) est multiple. Il est d'abord présenté par Friedman et Savage comme une première tentative de test de la théorie de vNM - alors que Davidson cherche précisément à tester le modèle canonique de la théorie de la décision. Le modèle de Friedman et Savage sert ensuite de socle à la première expérimentation réalisée par Mosteller et Nogee. C'est précisément à l'égard de cette dernière que Davidson, Suppes et Siegel sont très critiques, que ceux-ci bâtiront leurs propres expériences.

Suivre la chronologie de l'article de Friedman et Savage (1948) permet de distinguer trois parties, toutes liées analytiquement :

- leur définition du cadre théorique dans lequel la théorie de vNM prend place (i).
- leur tentative de tester empiriquement cette théorie à partir de données statistiques (ii).
- la méthode construite pour ce faire à partir de la théorie de vNM, méthode qui cherche à vérifier la validité empirique de celle-ci ; ce qui amène Friedman et Savage à mettre en évidence des profils d'individus en fonction de

leur aversion au risque et de leurs revenus puis à proposer des restrictions sur la forme de la fonction d'utilité (iii).

(i) Selon Friedman et Savage (1948), l'apport de vNM à la théorie économique est avant tout relatif à leur façon de concilier choix risqués et critère de la maximisation de l'utilité. Ils reviennent pour l'expliquer sur les liens entre la théorie de vNM et la théorie orthodoxe de l'utilité dans le certain. Cette dernière, notamment défendue par Edgeworth, Fisher et Pareto est fondée sur une analyse par les courbes d'indifférences.

Pour eux, les choix en univers certain sont volontiers expliquées par les économistes en termes de maximisation de l'utilité : les individus sont supposés choisir comme s'ils attribuaient une caractéristique commune aux différents biens, appelée utilité, et sélectionnaient la combinaison de biens produisant la plus grand montant de cette caractéristique commune. En revanche, les choix parmi des issues impliquant différents degrés de risque, par exemple, entre différents métiers, sont selon eux expliqués de manière très différente, de manière très différente dans la théorie économiques traditionnelle, par l'ignorance des chances ou par le fait que les « jeunes hommes dans une disposition aventureuse sont plus attirés par les perspectives ayant un grand succès qu'ils ne sont détournés par la peur de l'erreur » (Friedman et Savage [1948], p. 280).

La théorie orthodoxe fait donc une distinction entre choix en univers risqué et choix en univers certain et propose de les analyser selon des démarches divergentes.

Selon Friedman et Savage – reprenant l'argument de vNM - une large gamme de réactions des individus face au risque peut toutefois être analysée par une simple extension de l'analyse orthodoxe de l'utilité dans le certain.

Le rejet de la maximisation de l'utilité comme représentation des choix risqués est, selon les auteurs, la conséquence directe de « la croyance en l'utilité marginale décroissante » (*ibid*). Si l'utilité marginale de la monnaie diminue, un individu cherchant à maximiser son utilité ne participera jamais à un jeu équitable, par exemple un jeu consistant à gagner ou perdre un dollar avec la même probabilité. En effet, imaginons la loterie suivante : L = (1, -1 ; ½, ½). Si l'utilité marginale de la

monnaie est décroissante, la perte d'utilité associée au fait de à perdre 1$ est nécessairement supérieure au gain d'utilité liée au fait de gagner 1$. Dès lors, l'utilité espérée de cette loterie est négative. L'utilité marginale décroissante et la maximisation de l'utilité espérée impliquent que les individus doivent être payés pour les inciter à supporter le risque (dans cet exemple).

Selon Friedman et Savage, « cette assertion est nettement contredite par le comportement réel. Les individus ne s'engagent pas seulement dans des jeux équitables, ils s'engagent librement et souvent impatiemment dans des jeux non équitables comme des loteries.

La contribution de vNM est une tentative de réhabilitation de la maximisation comme explication des choix parmi des issues risquées. Puisque le rejet de la maximisation est liée à la décroissance de l'utilité marginale, l'idée sous jacente à la théorie de vNM est de réintégrer le principe de maximisation en laissant de côté cette fois la décroissance de l'utilité marginale. Mais le fond du problème consiste plus précisément pour vNM à produire une théorie qui puisse à la fois donner une explication des choix en univers certain et en univers risqué.

Fisher et de Pareto notamment considèrent qu'il n'est pas nécessaire de calculer des différences d'utilité (procéder de manière cardinale) pour déterminer l'utilité maximale. Il suffit en effet de classer les biens selon un critère ordinal. Pourtant leur hypothèse de décroissance d'utilité marginale suppose bien de comparer des différences d'utilité en univers certain.

Friedman et Savage suivent la perspective cardinaliste introduite par vNM en utilisant ces différences d'utilité qui ne sont pas difficiles à obtenir.

L'idée de Friedman et Savage est de supposer que si un individu montre, par son comportement, qu'il préfère A à B, et B à C, l'analyse économique traditionnelle suppose qu'il attache plus d'utilité à A qu'à B, et plus d'utilité à B qu'à C. Toute fonction d'utilité qui donne le même classement de ces issues constitue une aussi bonne représentation de ces préférences.

Si un individu pouvait montrer par son comportement qu'il préfère une combinaison de probabilité de A et C avec la même probabilité ½ par rapport à la certitude de B, les auteurs considèrent pouvoir « analyser son comportement en supposant que la différence entre les utilités qu'il associe à A et B est plus grande que la différence

entre les utilités qu'il attache à B et C, de telle sorte que l'utilité espérée de la combinaison est supérieure à l'utilité de B » (Friedman et Savage [1948], p. 282). En introduisant des calculs à partir des différences d'utilité, les auteurs adoptent une perspective cardinale de l'utilité.

Les auteurs remarquent que la classe des fonctions d'utilité, si elles existent, qui représentent toutes le même classement d'issues A, B, C risquées est plus petite que celle des fonctions d'utilité qui représentent toutes le même classement des issues A, B, C non risquées.
Ces fonctions diffèrent seulement au niveau de leur origine et de leur unité de mesure, c'est-à-dire les fonctions d'utilité qui sont dans la même classe sont des fonctions affines des autres.
Pour les auteurs, dès lors, « les propriétés ordinales des fonctions d'utilité peuvent être utilisées pour rationaliser les choix non-risqués, et les propriétés numériques – cardinales - de celles-ci pour rationaliser les choix risqués » (Friedman et Savage, 1948, p. 242).

(ii) Friedman et Savage affirment avoir pour objectif de « fournir un test empirique rudimentaire (*crude*) [de la théorie de vNM] en en rassemblant des observations relativement larges sur le comportement d'individus choisissant entre des issues risquées » (*ibid.*, p. 282). Friedman et Savage recueillent effectivement toute une série d'observations du comportement des individus face à des issues risquées. Toutefois cette procédure de recueil n'est pas détaillée et aucune procédure expérimentale ne semble être toutefois mise en place. Il n'est donc en fait guère question de test de la théorie de VNM par le biais d'une expérimentation comme Mosteller et Nogee le feront.
Les auteurs se contentent de proposer des données, observations et exemples.
Les phénomènes économiques pour lesquels l'hypothèse de vNM est pertinente sont les phénomènes de pari et d'assurance. Ici, l'influence du risque est plus marquée et plus significative que dans d'autres domaines.
Sans avoir recours à une investigation empirique de grande échelle, les auteurs choisissent un exemple illustratif. Ils expliquent les décisions économiques majeures

d'un individu pour lesquelles le risque joue un rôle important sont les décisions relatives à l'emploi de ses ressources par l'individu, par exemple le métier qu'il choisit - dans leur exemple les auteurs n'attribuent pas de probabilités précises à ces situations. Leur exemple s'appuie sur l'identification de trois types de risque (d'investir de perdre les ressources investies) : les situations impliquant « peu ou pas de risque comme enseigner par exemple »(idem, p. 284*)*, les situations où « le risque est modéré mais pouvant conduire rarement à des gains ou des pertes extrêmes comme par exemple l'activité de dentiste » (idem) et enfin les situations les plus risquées où « la possibilité de gains ou de pertes extrêmes est importante comme les emplois de pilote automobile par exemple » (Friedman et Savage [1948], p. 284).

Une fois ces degrés de risque identifiés, les auteurs expliquent que les économistes s'intéressent essentiellement au premier et troisième cas, très rarement au second. Pour Friedman et Savage, bien qu'aucunes données n'apparaissent dans l'article, nombreuses sont les personnes achetant de l'assurance tout en prenant part dans le même temps à des paris. Il est assez peu probable, selon eux, que ces deux types de comportement soient dissociés et qu'ils caractérisent des individus foncièrement différents. Pour le montrer, les auteurs s'appuient sur une première série de statistiques qui mettent en évidence le fait que les individus aux revenus les plus bas s'orientent généralement vers des activités peu ou pas risquées – s'approchant de l'assurance - et les revenus élevés vers des activités risquées – correspond à de la participation à des loteries. Ces données sont ensuite nuancées par d'autres montrant que ce sont généralement les individus aux revenus les plus bas qui achètent des produits financiers extrêmement risqués. Ainsi semble t-il plus pertinent d'avancer l'idée que les individus achètent à la fois de l'assurance et des tickets de loterie.

(iii) La méthodologie de l'article de 1948 est calquée sur celle de vNM. L'intérêt de la présentation axiomatique de ces derniers réside toujours pour Friedman et Savage dans sa capacité à faire apparaitre le peu de différence qu'il existe entre « l'hypothèse de maximisation de l'utilité de vNM » (d'ordre cardinal) et

l'explication usuelle par les courbes d'indifférence en univers certain (d'ordre ordinal).

Selon Friedman et Savage « l'hypothèse de vNM » peut-être ainsi synthétisée : « en choisissant parmi des issues risquées ou non qui sont ouvertes à lui, un individu se comporte comme si :

(a) il avait un ensemble cohérent de préférences ;

(b) ses préférences peuvent être représentées par une fonction d'utilité qui attache une valeur numérique à chaque issue considérée comme certaine ;

(c) l'objectif de l'individu est de rendre son utilité espérée la plus grande possible » (Friedman et Savage [1948], p. 287).

Cette structure en trois parties est une simple variante des postulats de vNM présentés dans la partie précédente.

En reformulant ces propositions dans les termes de la théorie de vNM, celles-ci deviennent le système suivant :

- Le système de préférences de l'individu est complet et cohérent. Ainsi, un individu peut dire lequel de deux objets il préfère ou s'il est indifférent entre les deux, et – considérant que A, B et C sont des objets et des combinaisons d'objet - s'il ne préfère pas C à B ni B à A, il ne préfère pas C à A. Cette dernière assertion est l'expression de ce qu'on appellera par la suite l'axiome de transitivité.
- Tout objet qui est une combinaison d'autres objets pondérés avec des probabilités n'est jamais préféré à chacun de ces autres objets et aucun d'eux n'est préféré à la combinaison.
- Enfin, si A est préféré à B et B à C, il existe alors une combinaison de probabilité de A et B telle que l'individu est indifférent entre cette combinaison et C.

Friedman et Savage tentent ensuite de déterminer les implications d'une telle hypothèse en particulier en termes d'aversion pour le risque. Pour ce faire, ils se placent d'abord dans un cas certain et supposent que l'utilité totale U (.) est fonction du revenu monétaire I de l'individu et cherche à représenter cette fonction dans un

plan. Partant du principe qu'un individu préfère toujours posséder plus, ils en déduisent que l'utilité sera une fonction croissante du revenu.
Soit un cas risqué où la loterie A permet de gagner un revenu I_1 avec la probabilité *a* ($0 < a < 1$) ou un revenu I_2, inférieur à I_1, avec la probabilité $1 – a$ et la loterie B offre la certitude de recevoir un revenu I_0.
Selon *l'hypothèse* de vNM, l'utilité espérée de A est $\bar{U}(A) = aU(I_1) + (1 – a)\, U(I_2)$.
Selon cette *même hypothèse*, un consommateur choisira A si $\bar{U}(A) > U(I_0)$, et B si $\bar{U}(A) < U(I_0)$ et sera indifférent $\bar{U}(A) = U(I_0)$.

Friedman et Savage expliquent ensuite ce qu'ils appellent la préférence pour une situation risquée. L'espérance mathématique de gain de la loterie A est $\bar{I}(A) = aI_1 + (1 – a)I_2$. Supposons que $I_0 = \bar{I}(A)$. Dès lors, si le consommateur choisit A, il manifeste une préférence pour une situation risquée, autrement dit, que $\bar{U}(A) > U(\bar{I})$ et $\bar{U}(A) – U(\bar{I})$ est « la mesure de l'utilité attachée à ce risque spécifique »(idem, p. 289) ou participer au jeu**).** Si, au contraire, il choisit B, il manifeste une préférence pour la certitude, autrement dit on a : $\bar{U}(A) < U(\bar{I})$.

Friedman et Savage introduisent ensuite l'équivalent certain de la loterie A, autrement dit le revenu I* qui procure à l'individu la même utilité que A : $U(I^*) = \bar{U}(A)$

Comme l'utilité est pour eux une fonction croissante du revenu alors
$\bar{U}(A) > U(\bar{I})$ implique que $I^* > \bar{I}$. Et $\bar{U}(A) < U(\bar{I})$ implique que $I^* < \bar{I}$.

Si $I^* > \bar{I}$, pour Friedman et Savage, on se situe dans un cas où le consommateur préfère la situation risquée le risque à la valeur espérée et sera prêt à payer la somme de $I^* – \bar{I}$ pour participer à un pari.
Friedman et Savage détaillent ensuite les choix du consommateur dans cette situation en comparant les loteries A et B avec les revenus $\bar{I}$ et I*. Si I_0 est supérieur à I*, il choisira la loterie B, dans le cas inverse, il choisira A comme le montre le graphique suivant.

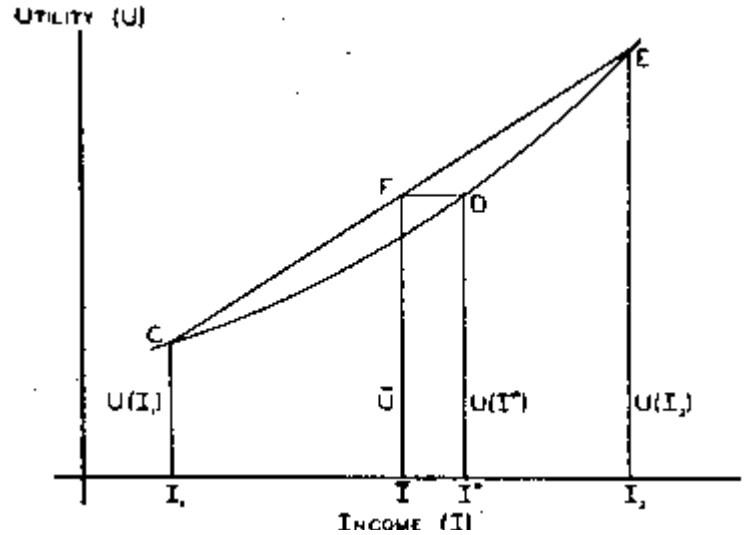


Figure 1 (graphique b de Friedman et Savage, 1948 ; p. 290)

Si I*< $\bar{I}$, pour Friedman et Savage, on se situe dans un cas où le consommateur préfère la certitude et sera prêt à payer $\bar{I}$ - I* pour s'assurer contre le risque. Dans cette situation, si I_0 < I*, le consommateur choisira A, dans le cas inverse, il choisira B comme le montre le graphique suivant.

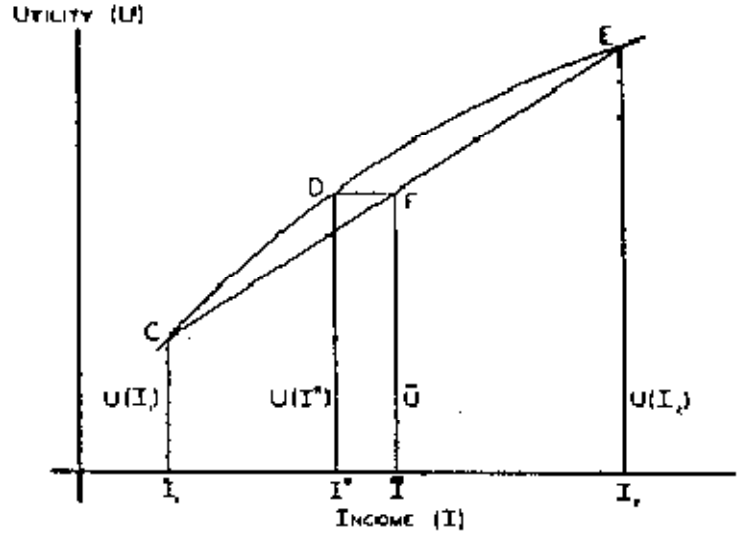


Figure 2 (graphique a de Friedman et Savage, 1948 ; p. 290)

Cette analyse du comportement des individus à partir de l'axiomatique de vNM permet donc de représenter l'aversion ou la préférence de l'individu pour le risque mais aussi de définir ce que les auteurs appellent des faits stylisés qui sont les suivants :

- concernant les revenus certains, les agents préfèrent les revenus supérieurs aux revenus inférieurs ;
- les agents ayant un revenu inférieur achètent ou veulent acheter de l'assurance
- les agents ayant un revenu supérieur achètent ou veulent acheter des tickets de loterie
- plusieurs agents disposant de revenus inférieurs achètent ou veulent acheter de l'assurance et des tickets de loterie
- les loteries possèdent souvent plus d'un prix

Les auteurs avaient déjà évoqué la croissance de l'utilité avec le revenu qui est une première restriction sur la fonction d'utilité. L'existence de ces faits stylisés impliquent de nouvelles conditions ou plutôt restrictions sur la fonction d'utilité. Les auteurs construisent en effet une courbe d'utilité de la monnaie permettant d'expliquer ces cinq faits à la fois.

Représentée dans un plan qui porte les revenus en abscisse et les utilités en ordonnées, la courbe d'utilité qui remplit toutes ces conditions serait d'abord concave, puis convexe pour redevenir concave[63].

[63] Remarquons que cette courbe « hypothétique » ressemble aux courbes de Davidson, Suppes et Siegel [1957] comme nous le verrons dans le chapitre 3.

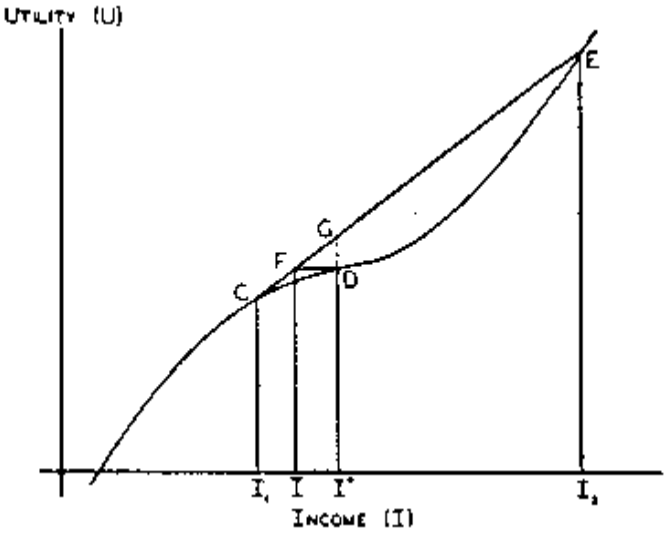


Figure 3 (graphique a de Friedman et Savage, 1948 ; p. 290)

Les auteurs font apparaître les différentes catégories de revenus et ainsi les classes socioéconomiques des agents distinctement.

.

La courbe décrivant l'utilité du revenu monétaire est convexe jusqu'au dessous d'un certain revenu (les personnes à revenus faibles ont une préférence pour une situation risquée), concave entre ce dernier et un revenu plus important (les personnes à revenus modérés préférence pour la certitude), convexe à nouveau pour les revenus les plus hauts, ce qui signifie que l'utilité marginale du revenu monétaire est respectivement décroissante, croissante puis de nouveau décroissante. Notons que les auteurs ont une interprétation particulière de la phase convexe comme une phase de transition de revenus, *c*orrespondant à une prise de risque pour entrer dans une nouvelle classe).

Friedman et Savage aboutissent par cette procédure à un certain nombre de résultats. Tout d'abord la démarche théorique qu'ils ont choisie permet à Friedman et Savage de mêler l'orthodoxie marshallienne, la théorie de vNM, et - en ce qui concerne le dernier point – l'analyse du comportement des individus parmi des issues risquées (Friedman et Savage [1948], p. 303). Ils proposent plus précisément un critère de maximisation de l'utilité qui comme on l'a dit ne suppose plus systématiquement une utilité marginale décroissante.

Rappelons que l'article de 1948 propose un traitement particulier de la théorie de vNM : l'objectif est ici de fournir un « test empirique rudimentaire » en liant des observations concernant le comportement des individus face à des issues risquées et en vérifiant si les observations sont cohérentes avec les hypothèses de von Neumann et Morgenstern. Selon Friedman et Savage, les observations empiriques sont totalement cohérentes avec les hypothèses si une forme particulière est donnée à la courbe de l'utilité totale de la monnaie.

Ensuite, l'article de 1948 permet à Friedman et Savage, à partir d'observations empiriques, d'élaborer des profils d'individus qui, en fonction de la forme de leur courbe d'utilité, vont préférer jouer ou s'assurer. Si nous ne cherchons pas ici à rendre compte en détail de chacun des cas envisagés, nous nous attardons cependant sur ce que les auteurs appellent une « expérience conceptuelle pour déterminer la fonction d'utilité » (*idem*, p. 292), car la procédure proposée par Friedman et Savage pour déterminer la fonction d'utilité sera reprise par Davidson (1957).

Friedman et Savage proposent de considérer deux revenus, l'un de 500$ et l'autre de 1000$. On peut assigner à ces deux revenus des utilités arbitraires, par exemple U(500$) = 0 et U(1000$) = 1. Si l'on choisit par ailleurs, un autre revenu, par exemple 600$, on peut proposer aux individus le choix entre les deux loteries suivantes :

(A) une probabilité a d'obtenir 500 $ et une probabilité $1 - a$ d'obtenir 1000 $

ou

(B) obtenir 600$.

Selon les auteurs, si l'on fait varier la valeur de a de telle sorte à obtenir une relation d'indifférence entre (A) et (B), on a la possibilité de déterminer U(600$). Supposons en effet que l'indifférence soit obtenue pour $a = a^*$, on peut en déduire que $U(600\ \$) = a^*U(500\$) + (1 - a^*)U(1000\$) = 1 - a^*$. En réitérant l'opération pour tous les gains possibles compris entre 500 et 1000 $, on peut ainsi déterminer les niveaux d'utilité que l'individu leur accorde (Friedman et Savage [1948], p. 293). Selon les auteurs, cette expérience renouvelée doit conduire à la même fonction d'utilité à chaque fois, à une unité de mesure près. Autrement dit, pour vérifier l'hypothèse de

vNM, il doit exister une classe de fonctions d'utilité qui vérifie toutes le même ordre de préférence.

Cette procédure permet donc, selon les auteurs, de fournir un test de l'hypothèse. Mieux, la connaissance complète des préférences des individus parmi les issues comme (A) et (B) permet de prédire les réactions des individus face à d'autres choix risqués. Même si les auteurs reconnaissent que l'on pourra trouver des exemples susceptibles de contredire l'hypothèse. Toutefois, et on retrouve ici l'épistémologie de Friedman, il ne s'agit pas pour les auteurs de faire des hypothèses réalistes, ni de décrire la réalité mais de « fournir des prédictions suffisamment précises sur les décisions concernées par l'hypothèse de vNM » (*ibid.*, p. 298)

C'est cette méthode qui sera utilisée par Mosteller et Nogee en 1951

2.1.5.2 Friedman et Savage (1952)

Dans leur second article écrit en 1952, Friedman et Savage questionnent à nouveau la validité de la théorie de l'utilité espérée pour expliquer les choix risqués et non risqués.

Ils reviennent en particulier sur les débats autour de l'utilité, et, en particulier sur deux problèmes que les économistes semblent voir dans la présentation de vNM : l'existence d'une classe de fonctions d'utilité affines représentant les mêmes préférences semblent en effet poser problème de même que l'utilisation d'une « utilité mesurable » pour rationaliser les choix certains et risqués (Friedman et Savage, [1952]p. 464).

En effet, Pareto et Slutsky ont pu montrer qu'il n'est pas nécessaire d'avoir une utilité numérique pour analyser les choix certains.

Friedman et Savage essaient dans cet article de montrer que l'hypothèse de vNM est pertinente en montrant son intérêt pour l'analyse des choix certains et incertains et en revenant sur la classe des fonctions d'utilité représentant les préférences. Ils répondent ainsi aux critiques que les économistes adressent à la théorie de vNM. Tout en privilégiant le problème de la mesure de l'utilité sur lequel ils estiment n'avoir pas eu le temps d'insister dans leur premier article (1948).

Pour ce faire, ils utilisent l'article de Baumol publié en 1951 qui regroupe un ensemble de reproches faits au modèle de vNM (i) et (ii). C'est à partir de ces critiques que Friedman et Savage tentent précisément de rétablir la pertinence et la signification de l'hypothèse d'utilité espérée (iii).

Dans son article publié en 1951, Baumol, outre proposer de s'en tenir à « un point de vue ordinaliste » (1951, p. 61), adresse les deux critiques suivantes à la théorie de vNM reprise par Friedman et Savage en 1948 :

> (i) la première est que l'index d'utilité de vNM est selon lui arbitraire, choisi de manière injustifiée.
> (ii) la seconde est que la construction théorique de von Neumann et Morgenstern peut être incompatible avec l'échelle de préférences d'un individu. Cette critique a trait aux fondements sur lesquels la théorie doit être acceptée ou rejetée.

(i) La première critique de Baumol

Selon Baumol, vNM suggèrent qu'il n'existe qu'une seule échelle d'utilité qui peut être établie à partir des préférences d'un agent selon leur hypothèse : « dans la présentation réalisée dans leur ouvrage, comme dans les autres discussions, l'impression est donnée qu'il existe une échelle numérique unique d'utilité (la vraie mesure) qui peut être déduite d'un nombre suffisant d'informations obtenues à partir du comportement observé des individus » (Baumol [1951], p. 61). Or selon lui, cette échelle n'est qu'arbitraire, il s'agit d'une échelle d'utilité parmi d'autres compatibles avec les préférences d'un individu.
Plus précisément, selon Baumol, pour vNM, si U(a) décrit les préférences d'un individu, alors V (*a*) = s U (*a*) + t les décrit aussi. Il n'existe alors pour lui qu'une marge réduite de liberté, s et t en fait, pour faire varier l'utilité. Selon lui, vNM font ce faisant explicitement ou implicitement une hypothèse de psychologie humaine afin de prouver l'unicité de V (*a*) = s U (*a*) + t, c'est-à-dire la cardinalité : « il est nécessaire de postuler […] une description simplifiée de la réaction psychique

aux issues risquées qui revient à dire que les utilités sont calculées avec des espérances » (Baumol [1951], p. 62).

Baumol explique « adopter un point de vue ordinaliste » (Baumol [1951], p. 65) contraire à celui de vNM. Pour lui, il n'est guère nécessaire de faire une hypothèse aussi restrictive sur la classe des fonctions d'utilité V(.) compatibles avec un même ensemble de préférences. Cette classe est selon lui « trop étroite » (Baumol [1951], p. 65) alors que toute fonction d'utilité qui respecte l'ordre des préférences est valable : « deux échelles sont substituables si elles conduisent toutes deux dans tous les cas à prendre exactement les mêmes décisions (comme acheter ou ne pas acheter) » (*ibid*).

(ii) La seconde critique de Baumol

Outre mettre en relief des hypothèses implicites fortes faites sur les réactions des agents, Baumol cherche à montrer que la fonction d'utilité proposée par vNM ne décrit pas systématiquement les préférences des agents. Le second argument de Baumol ([1951], p.62) consiste à dire que l'utilité de vNM peut être incompatible avec l'échelle de préférences d'un individu et donc fausse.

Ce que tente de montrer Baumol, c'est qu'il n'est pas difficile de construire des index de vNM qui mènent à des résultats contradictoires concernant les préférences. Par exemple, si l'on considère trois issues auxquelles on assigne les valeurs 600, 420, et 60, en termes d'utilité. Si l'on considère deux loteries offrant les paris suivants :

(A) : une probabilité de 5/6 de gagner 600 avec 60 comme lot de consolation en cas de perte ; (B) : une probabilité de 1/6 de gagner 600 avec cette fois-ci 420 comme lot de consolation.

La théorie de vNM indique que l'individu devrait préférer *a* à *b* (car 5/6 *600 + 60*1/6) > (1/6*600 + 420*5/6), le gain de A est donc 510, et celle de B est 450. Mais Baumol se demande pourquoi le choix inverse devrait être nié, il refuse d'assimiler le choix de b comme la manifestation d'une pathologie. Le choix de b se justifierait, en effet, par la préférence d'une issue assurée à une issue risquée.

Dès lors, l'index d'utilité de vNM ne correspond pas forcément aux préférences effectives des individus, il n'est donc pas toujours validé.

La fonction d'utilité proposée par Friedman et Savage en 1948 qui pourtant tient compte de l'aversion ou non pour le risque est également critiquée par Baumol. L'hypothèse de Friedman et Savage relative au fait d'investir à la fois en assurance et en loterie nécessite une fonction d'utilité particulière comme on l'a vu où l'utilité marginale est tantôt décroissance tantôt croissante.

Si Baumol ne voit pas d'incohérences au sein de l'axiomatique de Friedman et Savage, s'il juge que dans certains cas les individus aux revenus plus faibles et ceux aux revenus se comporteront effectivement comme le disent Friedman et Savage, Baumol considère d'abord que le passage d'un revenu à un autre est trop imprécis, et que ces cas ne peuvent être érigés en cas généraux : « il existe de sérieux doutes sur la validité universelle de l'hypothèse de vNM » (Baumol [1951], p. 65).

En conséquence, Baumol propose de reformuler le modèle de décision sans l'hypothèse jugée « cardinale » de classe de fonctions d'utilité du type V (*a*) = s U (*a*) + t ni l'introduction d'une utilité espérée (Baumol [1951], p. 66).

(ii) Les réponses des Friedman et Savage, 1952

L'article de 1952 vise à répondre aux critiques de Baumol.

Leur première réponse est d'ordre épistémologique : alors que Baumol critique l'arbitraire, la particularité voire l'irréalisme des hypothèses de vNM et de Friedman et Savage, ces derniers rappellent les fondements sur lesquels on peut rejeter ou accepter une hypothèse. Suivant l'épistémologie friedmanienne, la fonction d'une théorie est « de permettre de prédire des phénomènes non encore observés » (Friedman et Savage, 1952, p. 465). Une théorie doit en outre pouvoir être contredite sans l'être effectivement.

Plus la classe des phénomènes observables capables de contredire la théorie est large, plus son potentiel est important. De même plus le nombre d'occasions où une théorie a passé le test de la contradiction est grand, plus la confiance en cette théorie est importante.

L'ensemble des arguments de Baumol consistant à mettre en lumière des situations contredisant la théorie d'utilité espérée, ne constituent pas, pour Friedman et Savage,

une critique de cette théorie. Eux-mêmes avaient d'ailleurs en 1948 évoqué plusieurs comportements susceptibles de contredire la théorie. Pour ces derniers, les arguments de Baumol sont vains car ce dernier fait appel à l'observation courante et l'introspection pour justifier le fait que la théorie de l'utilité espérée est fausse. Or, Baumol n'évoque aucune observation concrète et il ne peut démontrer qu'une telle observation a été effectuée, même sommairement.

Baumol ne produit pas, en somme, de preuves permettant d'invalider l'hypothèse d'utilité espérée.

La stratégie de Friedman et Savage consiste alors à exclure la supposée invalidité de la théorie au nom d'absence de données empiriques la vérifiant et à la sauvegarder car sa validité peut être prouvée indirectement : l'hypothèse doit être considérée comme une conjecture prometteuse dont la validité provient davantage des preuves indirectes que des preuves directes (1952, p. 466).

Les preuves indirectes de sa validité résident essentiellement dans « sa cohérence avec le reste de la théorie économique » (idem, p. 466). Selon Friedman et Savage, la preuve indirecte est plus précisément fournie par la validité d'un ensemble de postulats suffisants pour en déduire l'hypothèse et sont eux-mêmes déduits d'elle tels qu'ils puissent être considérés comme une alternative à cette hypothèse » (*ibid*). Il suffit alors de montrer que ces postulats n'ont jamais été eux-mêmes contredits.

Friedman et Savage cherchent ensuite à répondre à Baumol sur un autre point. Selon eux, Baumol remettrait en question un axiome spécifique de la théorie (telle que reformulée par Savage, en insistant davantage sur les probabilités). Il ne s'agit ni de l'axiome d'ordre complet ni de l'axiome de continuité mais de l'axiome d'indépendance.

Friedman et Savage cherchent ensuite à répondre à ce que l'on a appelé ci-dessus la première objection de Baumol, autrement dit la nature « arbitraire » de l'échelle d'utilité de vNM. Selon Friedman et Savage, cette critique a en fait trait à la mesurabilité de la fonction d'utilité. Selon eux, en effet, Baumol propose de s'en tenir à une conception ordinale de la fonction d'utilité qui correspond à leur axiome P1, une fonction d'utilité qui d'une part respecte le classement ordinal des préférences sans introduire un critère d'utilité espérée et d'autre part est valable à une fonction monotone près.

Selon eux, toutefois le type de fonction d'utilité monotone ainsi définie « n'a alors pas plus de droits qu'une autre d'être appelée « la fonction d'utilité. En ce sens, l'utilité n'est pas mesurable » (idem, p. 470). Ils vont jusqu'à qualifier une telle conception de « théorie générale du choix presque vide » (*ibid*) et ce, parce qu'elle serait « relativement inutile pour prédire le comportement » (*ibid*). Il suffit pour donner un sens à cette théorie et une valeur à l'utilité de restreindre la classe des fonctions d'utilité. Il suffit pour cela de supposer qu'il existe une fonction du revenu dont la valeur espérée qui respecte le classement des préférences de l'individu considéré. Cette fonction permet d'une part de prédire le comportement (le choix se porte sur l'issue dont l'utilité espérée est la plus grande) et d'autre part est valable à une transformation affine près. L'utilité espérée ainsi définie donne une « mesure comme la température et la longueur sont mesurables » (*ibid*, p. 472). Ce n'est donc pas la seule mesure possible mais celle qui est choisie car elle est la plus pratique (*convenient*).

Les auteurs en concluent que « la mesurabilité est utilisée en référence à l'étroitesse de la classe des fonctions d'utilité » (*ibid*). Elle ne doit pas selon eux être rejetée au nom du « réalisme » (*ibid*).

2.2. L'axiomatique de Savage

Si Davidson et Suppes utilisent les modèles de vNM (1947) et Friedman et Savage (1948 et 1952), ils envisagent leurs expériences dans le cadre de la théorie de la décision où les agents doivent évaluer à la fois les issues ou résultats mais aussi la probabilité d'événements qui les provoquent. En cela, ils s'inscrivent dans la perspective ouverte par Savage. En distinguant états de la nature, actes et conséquences, ce que vNM et Friedman et Savage ne proposent pas de faire, Savage complexifie l'axiomatique initiale.

L'ouvrage de ce dernier, *The Foundations of Statistics* (1954), est une tentative de construire une théorie des préférences rationnelles. L'auteur s'intéresse à un agent imaginaire et idéalisé qui doit planifier l'ensemble des choix qui s'offrent à lui. A partir de sept axiomes, l'auteur prouve, notamment, qu'il existe une représentation,

en termes d'espérance d'utilité (dont l'origine remonte à Daniel Bernoulli), du préordre des préférences de l'agent sur les actions entre lesquelles il peut choisir. En décidant d'une action, l'agent doit prendre en compte les états possibles du monde, c'est à dire l'ensemble des états de la nature, ainsi que l'ensemble des conséquences implicites à chaque acte dans chaque état du monde possible (Savage [1954], p.13). Savage[64] définit une conséquence comme tout ce qui peut arriver à la personne qui choisit. Mieux, les conséquences peuvent être appelées « les états de la personne » en opposition aux états du monde. L'idée que l'individu doit tenir compte des états possibles du monde renvoie à une conception particulière des probabilités. Plus précisément, Savage opte pour une interprétation des probabilités personnelles (Savage [1954], p.3) permettant de mesurer la « confiance » (*confidence*) ou croyance qu'un individu a de la vérité d'une proposition particulière[65], par exemple « le fait qu'il pleuvra demain ».

Il s'agit à présent de comprendre comment avoir accès non seulement aux préférences de l'individu mais aussi aux probabilités qu'il attribue aux événements qui vont conditionner ces actions, contrairement au modèle de vNM dans lequel les probabilités sont objectives.

Pour illustrer l'interconnexion entre les actes, les conséquences et les décisions, Savage donne l'exemple suivant (Savage [1954], pp.13-14): imaginons un homme qui se trouve face à cinq bons œufs cassés dans un bol et qu'il se propose de terminer la préparation de l'omelette. Il s'agit pour l'individu de savoir s'il doit ajouter un œuf à la préparation ou la laisser en l'état sachant qu'il ne sait pas si le sixième œuf est frais ou périmé. Autrement dit, il doit choisir entre trois actes : casser le sixième œuf et l'incorporer aux cinq autres, casser le sixième œuf dans un récipient à part afin de l'inspecter ou le jeter sans l'inspecter.

En fonction de l'état du sixième œuf, chacun de ces trois actes aura les conséquences suivantes :

---

[64] On peut noter que deux influences majeures pour Savage sont les travaux de Ramsey [1931], et de Finetti [1937].

[65] Cette position est celle de Jeffrey [1983]. Sur les conseils de Savage, ce dernier fera même porter la relation de préférence usuelle sur les propositions comme nous le verrons.

Acte	Etat	
	Œuf frais	Œuf périmé
Casser l'œuf dans le bol	Une omelette de six œufs	Pas d'omelette et cinq œufs gâchés
Casser l'œuf dans un récipient à part	Une omelette de six œufs et le récipient à laver	Une omelette de cinq œufs et le récipient à laver
Jeter l'œuf	Une omelette de cinq œufs et un bon œuf détruit	Une omelette de cinq œufs

Cet exemple permet de préciser la méthodologie de Savage[66]. Il adopte très clairement une vision behavioriste de l'individu car, plutôt que de demander aux gens la probabilité qu'ils attribuent aux événements, il préfère la déduire de leurs choix dans diverses circonstances : « il serait préférable, au moins en principe, d'interroger la personne, non pas littéralement à travers sa réponse verbale aux questions mais plutôt dans un sens figuratif qui rappelle quelque peu celui dans lequel une expérience scientifique est parfois considérée comme une interrogation de la nature » (Savage [1954], p.28). D 'où l'idée de demander à une personne de choisir entre deux issues, en lui proposant une récompense s'il vise juste, tout en sachant qu'il choisira l'évènement qui lui semble le plus probable.

Savage définit ce qu'il appelle un ordre simple (*simple ordering*) entre événements, notée $\lesssim$, qu'il appelle une « probabilité qualitative » (Savage [1954], p.30), qui a les propriétés habituelles d'un classement (préordre complet, réflexif et transitif), et qui

[66] Nous allons nous attacher à présenter la théorie de Savage afin de mettre en relief l'emprunt de Davidson, Suppes et Siegel sans insister sur les différents débats relatifs aux axiomes proposés par Savage. Pour une présentation détaillée de Savage, voir Savage [1954], Anscombe et Aumann [1963].

a les deux propriétés suivantes : si *A, B* et *C* sont des événements quelconques, alors (Savage [1954], p. 32) :

. $A \lesssim\cdot B$ si et seulement si $A \cup C \lesssim\cdot B \cup C$ à condition que $A \cap C = B \cap C = \emptyset$.

. $\emptyset \lesssim\cdot A$, $\emptyset \lesssim\cdot E$, où *E* est l' « événement universel » (union de tous les évènements possibles).

Cette dernière condition signifie que n'importe quel événement est au moins aussi probable (préféré à) que l'évènement vide (ne se réalise jamais), et cela est vrai aussi pour l'événement universel, qui se réalise toujours.

La première condition traduit l'idée que le classement effectué entre deux évènements n'est pas modifié si on tient compte d'un autre événement (*C*) qui se réalise alors que ni *A* ni *B* ne le sont. Savage donne pour exemple le fait que l'importance de la récompense donnée dans le cas où l'événement choisi se réalise n'a pas d'influence sur ce choix (le classement effectué ne dépend pas du fait que la récompense soit plus ou moins grande).

Cette condition est à l'origine de l'égalité qui sert, entre autres, à caractériser une loi de probabilité $P(\cdot)$:

$$P(A \cup B) = P(A) + P(B) \text{ à condition que } A \cap B = \emptyset.$$

Savage montre qu'il existe une fonction $P(\cdot)$ à valeurs positives définie sur l'ensemble des événements, avec $P(E) = 1$ et $P(\emptyset) = 0$, associée à la relation de préférence $\lesssim\cdot$ de façon à donner le même classement qu'elle (Savage dit que $P(\cdot)$ est en accord - « *agree* » - avec $\lesssim\cdot$) :

$$A \lesssim\cdot B \text{ si et seulement si } P(A) \leq P(B) \text{ (Savage [1954], p.34).}$$

Pour que la fonction $P(\cdot)$ existe, il faut que la probabilité qualitative $\lesssim\cdot$ soit complète - elle classe tous les éléments (événements) de *E* –, mais cela ne suffit pas. Savage ajoute une proposition - P6 - facile à interpréter, selon lui, et qui est relative à la partition de E en parties qui n'ont aucun point commun. Il conclut que la fonction $P(\cdot)$ est en accord (*agree*) avec la probabilité qualitative $\lesssim\cdot$, « probabilité quantitative personnelle » (Savage [1954], p.33), expression qu'il préfère à « probabilité subjective » ou à « degré de conviction ». Il signale que, une fois admis les axiomes permettant l'existence de la probabilité quantitative personnelle, il n'y a plus besoin de recourir aux probabilités qualitatives, cette remarque étant relative au

mathématicien. Mais, pour l'expérimentateur, qui ne connaît pas ces probabilités, l'approche ensembliste, par classement des événements, demeure importante, puisqu'elle permet de déduire ces probabilités, ou du moins de les encadrer, comme le font Davidson et Suppes.

Après avoir ainsi donné un fondement rationnel – c'est-à-dire basé sur un classement complet et transitif de l'ensemble des événements – à la loi de probabilité personnelle, Savage reprend le théorème d'existence de vNM qui suppose seulement l'existence d'une loi de probabilité sur les événements (où les états de la nature) envisagés. Dans l'approche de Savage, il y a toutefois trois « relations de préférence » (ou relations d'ordre) qui interviennent :

i) celle qui concerne ce que Savage appelle « actions », qu'il note $\lesssim$. Une action est une fonction $f(\cdot)$ qui associe à l'état de la nature s (événement élémentaire), quel qu'il soit, la conséquence $f(s)$ (qui peut être monétaire ou autre). Ainsi, un acte peut être identifié à toutes ses conséquences possibles (Savage [1954], p.14). Dès lors, une relation de préférence sur des actes revient à des préférences sur des loteries comme par exemple : si l'acte f à deux conséquences c_1 et c_2 et que les probabilités associées à c_1 et c_2 sont respectivement p_1 et p_2, alors l'individu qui choisit f sera face à la loterie suivante L : $(c_1, c_2 ; p_1, p_2)$.

ii) la relation sur les événements eux-mêmes, qu'il note $\lesssim_{\cdot}$, qui est à l'origine de la loi de probabilité associée aux événements ;

iii) la relation de préférences sur les conséquences qui est une relation du type 1$ R 2$ (2 $ est préféré à 1$), relation qui est construite à partir de la relation de préférences sur les actes grâce à l'axiome P3 qui est relatif au transport des préférences sur les actes aux préférences sur les conséquences : $\forall s \in E$, $a_1(s) = c_1$, $a_2(s) = c_2$; si E se réalise $a_2 \gtrsim a_1 \Rightarrow c_2 \gtrsim c_1$.

Savage remarque que, dans cette perspective, toutes les actions et leurs conséquences étant donnés, les préférences sur les actes dépendent seulement de la distribution de probabilité des conséquences des actes, du moins si l'ensemble K des

conséquences est fini[67] – ce qui est le cas des expériences envisagées par Davidson et Suppes. Il considère donc des jeux de hasard ou paris de la forme $\{(f_i, p_i), i \in K\}$, où $f_i = f(s)$, avec $s \in B_i$ (où B_i est une partition de B, B étant un événement), auxquels s'appliquent le théorème de vNM, pourvu que les axiomes « de rationalité » (essentiellement, complétude et transitivité) concernant les classements $\lesssim$ et $\lesssim\cdot$ soient vérifiés.

Davidson et Suppes se situent dans la perspective « expérimentale » de Savage ; comme lui, ils cherchent un moyen d'évaluer l'importance relative attribuée aux divers événements à travers le classement fait à leur propos par les sujets des expériences – supposés être rationnels au sens donné plus haut.

Conclusion

Le modèle que Davidson, Suppes et Siegel construisent en 1957 emprunte donc largement aux théoriciens de la décision antérieurs. Ils reprendront comme nous allons le voir

- l'axiomatique de vNM dans la formulation moderne qu'en proposent Friedman et Savage
- la notion de probabilité subjective de Savage (et Ramsey) qui n'existe pas chez vNM et n'est pas introduite dans l'expérience de Mosteller et Nogee
- l'intuition de Friedman et Savage mise en œuvre par Mosteller et Nogee qui consiste à proposer un test de la théorie de vNM.

---

[67] Un exemple particulièrement intéressant de la conception de Savage est celle où celui-ci évoque l'axiome d'indépendance par un exemple.
Savage propose d'imaginer un homme d'affaires ayant l'idée de faire un certain investissement, immobilier par exemple. Il peut penser que la valeur de son futur bien est influencée par le résultat de la prochaine élection présidentielle. Dès lors, Il peut être amené à demander s'il ferait l'acquisition du bien en question si c'est le candidat républicain qui est élu. Il estime effectivement que si le candidat républicain est élu, cette opération sera profitable mais aussi dans le cas où il est battu. Dès lors, on peut dire qu'il ne préfère pas la situation ou il n'investit pas à celle où il investit, que le candidat républicain soit élu ou non.

Mais la construction ne s’arrête pas là, en s’inspirant de la méthode opérationnelle de Ramsey (qui consiste à déterminer une proposition éthiquement neutre), ils proposent d’expérimenter une théorie de l’utilité espérée avec des probabilités subjectives.

Chapitre 3.

Le modèle de Davidson (1957) : théorie et « hypothèses expérimentales »

L'ouvrage de Davidson, Suppes et Siegel (1957) est l'aboutissement d'un programme de recherches entamé en 1954 à l'université de Stanford par Davidson et Suppes. L'objectif central est de présenter toute une série de recherches et de travaux expérimentaux visant à évaluer la validité empirique de la théorie de l'utilité espérée ; mieux, d'apporter une « interprétation empirique de la théorie qui soit testable ». *Decision Making* publié en 1957 a été précédé par deux articles fondamentaux (Davidson, McKinsey et Suppes, 1955) et (Davidson et Suppes, 1956) dont la présentation ici permet de préciser le projet de Davidson et Suppes mais aussi d'expliquer les raisons des expériences à venir.

L'article de 1955 constitue plus précisément le premier travail de Davidson en théorie de la décision. Il s'agit pour les trois auteurs de proposer différentes mesures de ce qu'ils appellent « un ensemble rationnel de préférences » et de justifier leur intérêt pour une mesure « forte » qui utilise des échelles d'intervalles (3.1). En 1956, Davidson et Suppes mettent cette fois moins l'accent sur le cadre théorique dans lequel ils situent leur travail que sur la nature des outils mathématiques nécessaires et, en particulier, sur la nécessité de formuler une axiomatique finitiste où la mesure de l'utilité tient une place centrale, ouvrant la voie aux expériences de 1957 (3.2). C'est en fait dans ce second article que des fonctions d'utilité et de probabilité subjectives sont pour la première fois définie et que le cœur de ce qui constituera le modèle de 1957 est présenté.

Après avoir posé les jalons de cette axiomatique dont nous montrerons qu'elle est aussi fortement influencée par les travaux de Ramsey, nous présenterons la théorie de 1957. Les expériences et procédures de tests y apparaissent pour la première fois. Nous expliciterons ces dernières en les mettant en relation avec les expériences (3.4) et procédures (3.5) antérieures et notamment celle de Mosteller et Nogee.

Nous pourrons enfin déterminer les apports et limites du modèle de Davidson et ce, en identifiant les bénéfices et insuffisances internes de sa démarche au regard des objectifs affichés mais aussi par rapport aux modèles proposés depuis 1947 (3.6 et 3.7).

3.1. Différentes mesures de l'utilité : Davidson, Suppes et McKinsey (1955)

Dans leur premier article de 1955 « Outlines of a formal theory of value », Davidson, McKinsey et Suppes présentent leur programme de recherche comme une tentative de construction d' « une théorie du choix rationnel » en s'appuyant sur ce qu'ils appellent « une théorie formelle de la valeur ». Dès les premiers paragraphes, le lecteur doit toutefois se rendre à l'évidence, les auteurs cherchent avant tout à montrer que l'on peut mesurer les valeurs relatives qu'un individu associerait à des biens ou loteries sans toutefois fournir de définition précise de cette valeur. L'article de 1955 coécrit avec Suppes et McKinsey constitue en ce sens le premier article de Davidson relatif à la théorie de la décision et plus précisément à la mesure de l' « utilité ».

Comme nous l'avons mentionné, Davidson arrive à Stanford en janvier 1951, peu de temps après Patrick Suppes. McKinsey, quant à lui, est un professeur de logique dont les recherches sont orientées à l'époque vers la théorie des jeux comme en témoigne son ouvrage *Introduction to the Theory of Games* en 1952.

Le point de départ des trois auteurs consiste à s'interroger sur la théorie de la valeur au sens large, telle que cette question est abordée par les philosophes comme Emmanuel Kant et plus tardivement Ralph Barton Perry, John Dewey ou encore Clarence Irving Lewis, autrement dit des figures éminentes de la philosophie américaine de l'époque. Autrement dit, il ne s'agit pas pour eux de questionner la théorie de la valeur des économistes classiques ou contemporains, de remettre en cause la notion même d'utilité ou de chercher à la définir. La « théorie de la valeur » n'est interrogée et utilisée qu'en tant qu'elle permet de construire une théorie du

choix rationnel. Portés par l'intérêt de McKinsey, alors professeur influent, pour la logique et la théorie des jeux, les auteurs suivent la voie ouverte par Frank Ramsey.
L'idée est simple : tout comme Frank Ramsey a utilisé la logique pour définir les conditions formelles de la croyance rationnelle, Davidson, McKinsey et Suppes proposent de définir les conditions formelles du choix rationnel, formulation faisant directement allusion aux travaux de Arrow (1951).
Pour construire une théorie du choix entre plusieurs biens ou plusieurs issues, encore faut-il en mesurer les valeurs relatives. Davidson, McKinsey et Suppes abordent donc la théorie de la valeur à travers la mesure de celle-ci. Les auteurs proposent plusieurs méthodes pour mesurer la valeur et plus spécifiquement l'utilité. Il n'est, selon eux, nul besoin de s'attarder sur des questions d'ordre métaphysique ou sémantique lorsque s'interroge sur la valeur ; le but ultime étant de proposer des méthodes permettant de mesure la valeur et de persuader les philosophes que ces méthodes sont utiles (Davidson, McKinsey, Suppes [1955], p.140).

Ces méthodes sont présentées au travers d'axiomatiques que nous présentons ici. Nous revenons pour commencer sur la première mesure proposée par les auteurs (3.1.1), mesure qu'ils considèrent comme « faible » et limitée (3.1.2), ce qui justifie la construction d'une mesure forte (3.1.3). Cette dernière se modifie toutefois en fonction de l'ensemble sur lequel porte le choix des individus. Les auteurs proposent en effet deux mesures fortes de la valeur (3.1.4), l'une portant sur un ensemble d'issues fini, l'autre infini.

Avant de présenter les différentes mesures exposées par Davidson, McKinsey et Suppes, il nous faut présenter succinctement la typologie des mesures qu'il est d'usage de présenter car c'est à cette typologie[68] que vont faire implicitement référence les auteurs.
Tous ces types de mesures ont au moins deux points communs : attribuer des nombres à des objets et être déterminées expérimentalement. Chaque type de mesure

[68] Pour une présentation détaillée, voir Granger [1988], Reuchlin [1970].

se réduit à un ensemble de règles visant à établir une correspondance entre certaines propriétés des nombres et certaines propriétés des objets.

On peut distinguer quatre types d'échelles de mesure :

i) Les échelles nominales correspondent au premier degré de la mesure, degré le plus faible en raison du type de correspondance exigée : il s'agit de dissocier les objets d'une classe suivant certaines caractéristiques plus ou moins détaillée. Les nombres éventuellement utilisés ne le sont que pour distinguer une classe d'une autre mais nullement pour des comparaisons de quelque sorte. La relation d'équivalence peut être utilisée pour placer les objets dans chaque classe. Par exemple, dans une école, on pourrait dissocier tous les enfants qui font du sport, qui n'ont jamais fait de sport ou qui ont cessé de faire du sport. Chaque membre de chaque classe est considéré comme équivalent à un autre selon le critère de l'activité sportive.

ii) Les échelles ordinales sont des échelles nominales auxquelles s'ajoutent de nouvelles propriétés. L'idée est d'entrer un peu plus dans la comparaison des membres d'une classe. Par exemple, dans le cas de la chaleur, il s'agit de trouver un dispositif permettant de dire laquelle de deux sensations est supérieure à l'autre. On recherche donc ici une application qui conserve la relation d'ordre construite sur les membres d'une classe. La classe des transformations d'une telle échelle est celle des transformations monotones.

iii) Les échelles d'intervalles sont quant à elles définies à une transformation linéaire près. L'idée générale de ce type d'échelles est de donner une définition et donc un sens à la distance ou à la différence de valeurs entre deux objets. C'est cette méthode qui est utilisée par Ramsey [1931] et Davidson, Suppes et Siegel [1957].

Les échelles de rapports peuvent être des versions sophistiquées d'échelles d'intervalles correspondent à un usage encore plus déterminant des nombres et de leurs propriétés car il s'agit de déterminer une estimation des rapports entre des objets d'une classe. Chez Ramsey par exemple, on a vu que les échelles d'intervalles étaient suffisamment sophistiquées pour aboutir à des rapports de distances.

iv) Enfin, dernier type d'échelles, les échelles absolues, définies de manière unique et qui constitue la mesure par excellence comme le Kelvin par exemple.

Voyons maintenant les différentes méthodes retenues par Davidson, McKinsey et Suppes pour mesurer ce qu'ils appellent la « valeur » au sens défini plus haut.

3.1.1 Une mesure faible fondée sur un quasi-ordre et un classement de préférences rationnelles

Les auteurs commencent par définir un « quasi-ordre ». Il s'agit pour eux d'une relation R réflexive et transitive, dans un ensemble K d'issues rangées par ordre de préférences (préférences asymétriques et transitives).

La relation R se décompose en deux relations : la relation de préférence P et la relation d'équivalence E. L'expression *x* P *y* décrit le fait que *x* est préféré à *y*, et la relation E décrit une relation d'équivalence en termes de préférence. E est transitive et symétrique.

Ils définissent ensuite « un classement de préférences rationnelles » (« Rational Preference Ranking »)[69]:

Définition 1 : Le triplet <K, P, E> est un classement de préférences rationnelles si et seulement si :

P1. La relation P est transitive ;

P2. La relation E est transitive ;

P3. Si *x* et *y* sont dans K, alors on a exactement l'une de ces relations : *x* P *y*, *y* P *x*, *x* E *y*.

Cette définition vise à décrire les conditions nécessaires de rationalité inhérentes à toute attribution de valeur. Elle n'a pas valeur de prescription selon les auteurs (Davidson, McKinsey, Suppes [1955], p.141) mais sert plutôt de point d'ancrage pour caractériser une préférence rationnelle.

Ainsi les auteurs considèrent-ils qu'un classement des préférences rationnelles dans le sens de la Définition 1 correspond à un classement ordinal représentant une

[69] Les auteurs expliquent que cette définition est « neutre » (Davidson, McKinsey, Suppes [1955], p.144) dans le sens où elle n'impose aucune restriction sur le contenu de K.

mesure « faible » de la valeur contrairement aux mesures cardinales qui sont qualifiées de mesures fortes.

Les auteurs reconnaissent que cette première mesure de la valeur fondée sur la définition comporte des limites car P1, P2 et P3 peuvent être tour à tour critiqués comme c'est le cas dans la littérature sur ce thème.

L'une des objections les plus sérieuses qui puisse être faite à la Définition 1 et plus précisément à P1 est l'argument de la pompe à finance, déjà présent chez Ramsey [1931] et qu'ils présentent sous la forme d'un exemple :

« Mr S se voit offrir trois opportunités d'emploi par un directeur de département : il peut être professeur à plein temps pour 5000$ (issue *a*), professeur associé pour 5500$ (issue *b*) ou assistant professeur pour 6000$ (issue *c*). Les raisonnements de Mr S sont les suivants : *a* P *b* puisque l'avantage en gloire l'emporte sur la petite différence de salaire, *b* P *c* pour la même raison, et *c* P *a* puisque la différence de salaire est maintenant suffisamment importante pour l'emporter sur une question de rang » (Davidson, McKinsey, Suppes [1955], p.155).

Pour les auteurs, il s'agit de démontrer que l'ensemble des préférences présenté dans l'exemple est irrationnel car intransitif. La transitivité apparaît alors comme une condition de rationalité. Pour le montrer, ils poursuivent leur exemple en imaginant la scène suivante :

« Le directeur du département, avisé des préférences de Mr S, dit : ' Je vois que vous préférez *b* à *c*, alors je vais vous laisser avoir la chaire de professeur associé pour une petite rémunération. La différence doit valoir quelque chose pour vous'. Mr S accepte de glisser 25$ au directeur de département pour obtenir son issue préférée. A présent le directeur du département dit : ' Puisque vous préférez *a* à *b*, je suis prêt, si vous me payer pour le dérangement, à vous laisser avoir la chaire entière'. Mr S remet 25 autres dollars et commence à partir, satisfait. 'Attendez' dit le directeur du département, 'Je viens de m'apercevoir que vous préférez avoir *c* à *a*, et je peux arranger cela…' »[70]. (*ibid.*)

[70] On retrouve ici la référence au pari hollandais, encore appelé pompe à finance, cette idée est notamment présentée par Ramsey [1926].

Ici, l'irrationalité provient de la cyclicité de la relation de préférence de Mr S, de l'intransitivité de ses préférences, ce qui constitue une violation de P1 : en poussant Mr S à faire plusieurs échanges et en le laissant payer pour chaque échange, il peut être ramené à sa position initiale mais avec moins d'argent dans les poches qu'au départ.

Une deuxième objection importante qui peut être formulée à l'encontre de la Définition 1 est relative à P2. Il est en effet d'usage de considérer comme Amstrong [1950] et plus tard Duncan Luce [1956] que la relation d'équivalence n'est pas transitive du fait d'un problème de discrimination entre les différents objets soumis à la comparaison. Ainsi, comme le suggèrent Davidson, McKinsey et Suppes, imaginons une relation d'ordre sur des éléments ($x_1 \ldots x_n$) dans laquelle x_1 a plus de valeur que x_n. Il se peut qu'il n'y ait pas de manière directe de détecter une différence de valeur entre les membres adjacents de la séquence $x_1, x_2, \ldots, x_n$, de telle sorte que $x_1 \, E \, x_2$, $x_2 \, E \, x_3$. Puisque la transitivité de E entraîne que $x_1 \, E \, x_n$ et donc non $x_1 \, P \, x_n$, on peut dire que l'axiome P2 demande inutilement une discrimination infinie entre les issues.
Mais selon les auteurs, « la Définition 1 n'implique pas que si nous croyons que l'on peut voir une différence entre x_1 et x_n, alors, pour être rationnel, on devrait être capable de voir une différence entre au moins deux membres adjacents de la séquence $x_1, x_2, \ldots, x_n$; elle implique simplement que si x_1 est considéré comme ayant plus de valeur que x_n, il doit être rationnellement considéré qu'il y a une différence de valeur entre au moins deux membres adjacents de la séquence » (Davidson, McKinsey, Suppes [1955], p.146).

Enfin, une troisième objection peut relativiser la portée de P3 dans la Définition 1.
P3 postule que toutes les issues prises deux à deux dans un classement rationnel peuvent être comparables. Or il est tout à fait possible, selon les auteurs, que deux ensembles possibles d'issues K_1 et K_2, chacun ordonné rationnellement, peuvent se chevaucher sans nécessairement impliquer que chaque membre de K_1 soit comparable avec chaque membre de K_2.

3.1.2 Une interprétation empirique difficile du classement des préférences rationnelles : la nécessité d'une mesure forte de l'utilité

Les difficultés liées à la définition théorique d'un classement rationnel sont d'une grande importance dès lors que les auteurs cherchent à « interpréter la préférence et l'équivalence de choix réels » (*ibid.*, p. 147). L'expression « choix réels « ne renvoie toutefois pas à la construction d'une théorie qui rendrait compte des comportements d'individus observés. Il s'agit davantage pour eux d'imaginer les choix possibles et les difficultés lies aux choix au sein d'exemple hypothétiques qui servent à bâtir un modèle puis une expérience (en 1957 seulement).

Selon eux en effet, la mesure faible de la valeur qu'on peut construire à partir de la définition s'appuie sur un classement rationnel des préférences ordinales. L'établissement d'un classement rationnel pose toutefois problème. Comme nous l'avons montré, les auteurs suggèrent qu'il existe des situations dans lesquels les individus comme M. S qui choisissent a à b, b à c, et c par rapport à a.

L'ensemble des préférences ainsi proposées n'est pas rationnel car il est intransitif. Cela ne signifie pas que M. S soit irrationnel. On peut en effet interpréter ses choix comme non simultanés. Autrement dit, on pourrait envisager que ces choix ont été faits de manière décalée. Le choix entre a et b en t, le choix entre b et c en un temps t' et le choix entre c et a en un temps t''.

Dans ce cas, on peut aussi établir selon la définition 1 que chaque choix particulier en un des trois temps t, t' et t'' provient d'un classement momentané de préférences qui est rationnel (Davidson, McKinsey, Suppes [1955], p.147) à ce moment. Le problème est qu'alors « on ne pourra jamais prouver que le classement rationnel s'étend à plus de deux issues » (*ibid.*).

Reprenant la théorie de Ramsey concernant les croyances, le point de vue des auteurs consiste à interpréter la préférence et l'équivalence comme des dispositions qui caractérisent des individus sur une période de temps. Dès lors, on peut considérer les

choix comme les preuves de cette disposition, mais non comme ces dispositions elles-mêmes.

Ainsi la définition 1 établit ce que serait un classement rationnel des préférences. Mais dans toutes les situations concrètes des choix, l'ensemble de préférences étudié ne se conforme pas a priori à cette définition. Tel est le cas des préférences de Mr. S. Il faut donc proposer ce que les auteurs appellent « une interprétation empirique opérationnelle de la préférence, les conditions formelles nécessaires pour construire un modèle de préférence » (*ibid.*). Comme lorsque pour déterminer si on fait un bon ou un mauvais usage du langage, on doit proposer au préalable des règles pour dire ce qui est vrai et ce qui est faux, les auteurs proposent ici des conditions formelles pour déterminer la rationalité d'un classement de préférences rationnelles. Il faut considérer les choix comme révélant les préférences, considérée ces dernières comme des dispositions et enfin proposer une mesure de ces dispositions.

3.1.3 Une mesure forte des préférences

Une théorie de la mesure, issue d'une théorie de la valeur, permet donc de donner une interprétation des préférences.
L'intérêt d'une mesure de la valeur et plus précisément d'une estimation des différences de valeurs entre plusieurs actions accessibles est qu'elle peut servir à la fois de guide à la décision mais elle permet aussi d'avoir une représentation numérique d'états mentaux qui sont parfois difficiles à décrire qualitativement.
Davidson, McKinsey et Suppes proposent d'établir, au travers d'un exemple, ce que serait une mesure forte de l'utilité en mettant l'accent sur l'importance des différences de valeur pour un individu et des probabilités qu'il attachent aux différentes situations possibles :

« Imaginons un homme politique, Wright, s'intéressant aux aides fédérales pour l'éducation. Wright a proposé un projet de loi visant à autoriser de telles aides, mais en incluant à ce projet des dispositions nécessaires indiquant que les sommes

allouées ne devaient aller aux écoles que si elles acceptaient de ne pas pratiquer la discrimination raciale.

Au moment de la présentation de ce projet de loi, il devint clair que ce dernier pourrait passer si la clause d'anti-ségrégation était abandonnée. D'un autre côté, Wright estime que le projet n'a seulement qu'une chance sur trois de passer avec la clause. Wright classe alors les trois résultats possibles par ordre de préférence : a) le projet passe dans son intégralité ; b) le projet passe mais sans la clause ; c) l'échec du projet. Cependant, son classement de préférences, et son estimation des chances de chaque issue ne servent pas, seuls, de base à la décision. Wright se demande s'il doit faire pression pour que le projet passe dans sa totalité avec une chance substantielle de défaite (action 1) ? ou doit-il accepter un projet affaibli avec la quasi certitude qu'il passerait (action 2) ? « (*ibid.*, p. 148).

C'est à partir de cet exemple que les auteurs justifient le besoin d'une mesure des écarts d'évaluation entre a et b et entre b et c.

En effet, selon eux si Wright accorde beaucoup d'importance à la clause, il s'intéresse peu au fait que le projet passe ou non. Dès lors, il choisira l'action 1. Si on suppose à présent que Wright estime que b est aussi bon que c, tout comme a est aussi bon que b, si on suppose que b serait à mi-chemin entre a et c en valeur.

Les auteurs proposent alors d'assigner des chiffres aux résultats possible : « si on assigne le nombre 2 à a, le nombre 0 à c et le nombre 1 à b (puisque ces nombres mesurent seulement les valeurs relatives de a, b et c, leur choix est arbitraire, excepté pour les magnitudes relatives de leurs différences) ».

Grâce à ces nombres, Wright peut selon les auteurs calculer les mérites relatifs de ces deux actions.

- Ainsi, puisque l'action (2) assure d'obtenir b, et que b a la valeur relative de 1, l'action (2) vaut 1.
- l'action (1) entraîne une chance sur trois d'obtenir a avec une valeur relative de 2, et deux chances sur trois d'obtenir c avec une valeur relative de 0. L'action (1) vaut donc 2/3 [(1/3).2 + (2/3).0 = 2/3].

Les « scores » relatifs des deux actions vont déterminés le choix de Wright. Les auteurs montrent en effet que tant que le ratio des différences de valeurs entre a et b,

et b et c est inférieur à 2, Wright choisira l'action 2, alors que si le ratio est supérieur à 2, il choisira l'action 1. C'est-à-dire [n (a) – n (b) / n (b) – n (c)] < 2.
Cette manière de calculer indique simplement qu'ayant décidé des mérites relatifs des issues, Wright a tempéré le poids qu'il voulait assigner à chacune d'entre elles par son opinion sur les probabilités.

La mesure proposée ici à partir de différences de valeur et de probabilités correspond à ce que les auteurs considèrent comme une mesure « forte » de la valeur.

Il ne s'agit toutefois pas d'une mesure forte comme la taille ou le poids que les auteurs qualifient d' « échelle de proportion, qui nécessite une unité arbitraire « (ibid., p. 150) mais d'un autre type de mesure qui peut être soit l'échelle absolue (*absolute scale*) c'est-à-dire le dernier degré de quantification ayant pour caractéristique d'être absolument unique comme nous l'avons mentionné plus haut; et enfin l'échelle d'intervalles, ayant pour caractéristique d'avoir une unité et un point 0 fixés arbitrairement, c'est l'exemple du temps ou de la longitude. C'est cette dernière forme qui est privilégiée par les auteurs.

Comme on l'a mentionné plus haut, cette investigation dans le domaine de la théorie de la valeur a pour but de convaincre les philosophes que la méthode proposée par Davidson, McKinsey et Suppes est utile, et peut conduire à des résultats intéressants.
Cette tentative de persuader les philosophes peut être comprise en présentant brièvement le positionnement théorique dans lequel la philosophie américaine se trouvait dans les années 1950.

L'exemple de Mr S et toutes les autres remarques de Davidson, McKinsey et Suppes avaient pour but de mettre en évidence le fait qu'une mesure de la valeur par une échelle d'intervalles était plus « forte » qu'une mesure par un classement des préférences rationnelles.

3.1.4 Deux axiomatiques conduisant aux mesures fortes d'un ensemble de préférences rationnelles

Une théorie cohérente de la mesure est déterminée selon eux par des conditions spécifiées via une axiomatique ; ces conditions sont imposées aux opérations et relations qui peuvent être expérimentées concrètement. Les auteurs proposent, à partir de cette idée, de présenter deux axiomatiques pour lesquelles la mesure par intervalles peut représenter un modèle de préférence ; l'idée étant simplement de mettre en relief les multiples variations possibles d'un même modèle, basé sur une mesure par intervalles.

i) Une axiomatique portant sur un ensemble K infini

En partant du modèle initial (décrit par la Définition 1) il est possible d'introduire une nouvelle relation *h*, fonction à trois arguments telle que si les issues *x* et *y* sont dans K, et si α est la probabilité qui n'est pas égale à 0 et à 1 (c'est-à-dire que α est nombre réel tel que $0 < \alpha < 1$), alors h(*x*, *y*, α) est l'issue consistant en *x* avec la probabilité α, et *y* avec la probabilité 1- α. Autrement dit, on construit une loterie : *x* avec la probabilité α ou *y* avec la probabilité (1- α). Si l'on reprend les termes traditionnels de vNM, elle peut s'écrire (*x*, *y* ; α, (1- α)).

Définition 2 : Le quadruplet ordonné <K, P, E, h> est un modèle de préférences rationnelles au sens 1 si et seulement si, pour tout *x*, *y* et *z* dans K, et pour tout α et β dans (0,1) ;

H1. Le triplet <K, P, E> est un classement de préférences rationnelles (au sens de la Définition 1) ;

H2. h (*x*, *y*, α) est dans K ; l'ensemble K doit contenir non seulement les issues données mais aussi toute les combinaisons de probabilité construites sur ces issues

H3. Si *x* E *y*, alors h (*x*, *z*, α) E h (*y*, *z*, α) ; H3 signifie que si *x* est équivalent à *y* en préférence, alors la combinaison de *x* avec n'importe quelle issue *z* avec la probabilité α est équivalente à la combinaison de *y* et *z* avec la probabilité α. L'axiome H3 peut dont être considéré comme une variante de l'axiome d'indépendance ou du principe de la chose sûre.

H4. Si *x* P *y*, alors, *x* P h (*x*, *y*, α) et h (*x*, y, α) P *y* ; H4 indique que si *x* est préféré à l'issue *y*, alors *x* est préféré à toute combinaison de probabilité de *x* et *y*, et que toute combinaison de probabilité de *x* et *y* est préférée à *y*.

H5. Si *x* P *y* et *y* P *z*, alors il *y* a un nombre γ sur (0,1) tel que y P h (*x*, *z*, γ) ;

H6. Si *x* P *y* et *y* P *z*, alors il y a un nombre γ sur (0,1) tel que h (*x*, *z*, γ) P *y* ; ces deux hypothèses signifient que si *y* est située entre *x* et *z* sur l'échelle des préférences, alors il existe une combinaison de probabilité de *x* et *z* qui est préférée à *y*, et une autre à laquelle *y* est préférée.

H7. h (*x*, *y*, α) = h (*y*, *x*, 1- α) ;

H8. h (h (*x*, *y*, α), *y*, β) E h (*x*, *y* αβ) est une règle de combinaison des probabilités.

Pour permettre à cette axiomatisation d'être en adéquation avec une mesure sur la base d'une échelle d'intervalles, les auteurs proposent le théorème suivant :

Théorème 1. Si <K, P, E, h> est un modèle de préférences rationnelles dans le sens de la Définition 2, alors :

(A) il existe une fonction Φ qui est définie sur l'ensemble K et dont les valeurs sont des nombres réels, telle que pour chaque x et y dans K et α dans (0,1)

(i) *x* P *y* si et seulement si *u*(*x*) > *u* (*y*),

(ii) *x* E *y* si et seulement si *u* (*x*)= *u* (*y*),

(iii) *u* (h(*x*,*y*, α)) = α *u* (*x*)+(1- α) *u* (*y*) ;

(B) si u_1 et u_2 sont deux fonctions satisfaisant (A), alors il existe des nombres réels a et b avec a >0 tel que pour tout x dans K,

$u_1(x) = a\, u_2(x) + b$.

(A) signifie qu'il est toujours possible d'assigner des nombres à des issues de manière à préserver la structure d'un modèle de préférences rationnelles.
(B) signifie que cette attribution de nombres a la propriété unique d'être la caractéristique d'une échelle d'intervalles, c'est-à-dire qu'une fois qu'une origine et une unité ont été choisies, cette attribution est unique.

ii) Une axiomatique portant sur un ensemble K fini

Le problème majeur de la définition 2 est, comme le remarque les auteurs, qu'elle considère que l'ensemble K doit contenir non seulement les issues de base mais aussi les combinaisons finies de probabilité construites à partir de ces issues, autant dire, un nombre infini d'éléments. La difficulté est de rendre cette hypothèse cohérente avec une interprétation empirique viable de la théorie. Or, selon les auteurs, il n'est pas possible de considérer que les individus soient en mesure de comparer un nombre infini d'éléments (Davidson, McKinsey, Suppes [1955], p. 156).
C'est pourquoi Davidson, McKinsey, Suppes proposent une autre axiomatique permettant de séparer les hypothèses relatives au nombre d'issues (c'est-à-dire au contenu de K) et celles au nombre de combinaisons d'issues possibles.
Pour cela, ils proposent, en plus du modèle de base, (c'est-à-dire en plus de définition 2) d'introduire une nouvelle relation T quaternaire décrite de la façon suivante:

« si *x*, *y* et *z* sont dans K, et α est une probabilité comprise dans l'intervalle [0, 1], c'est-à-dire $0 \leq \alpha \leq 1$, alors T (*x*, *y*, *z*, α) si et seulement si *y* n'est pas préférée à *x*, *z* n'est pas préférée à *y* et l'issue consistant en *x* avec la probabilité α et *z* avec la probabilité 1- α est équivalente à l'issue *y*. On pourrait penser que T (x, y, z, α) tient quand h (x, z, α) E y » (*ibid.*)

La relation T fonctionne sur le même mode que *h* (sauf qu'au lieu de porter sur deux éléments, elle porte sur trois). Elle n'est cependant pas utilisée dans le même objectif. En reprenant les notations contemporaines, on pourrait écrire : si *x* E *y* E *z* alors on peut construire une relation T telle que l'individu est indifférent entre (*x, z* ; α, (1- α)) et *y* ce qui revient au cas spécial où h (x, z, α) E y. La définition de T ressemble donc à celle d'un axiome de combinaison qui respecte la relation d'équivalence que l'on pourrait comprendre ainsi : si trois issues sont estimées équivalentes deux à deux alors il est possible de construire une loterie à partir de deux issues, loteries qui est elle-même indifférente à la troisième issue obtenue pour sûre.

Le problème est si *x* P *y* ou *z* P *x* la relation ne tient plus puisqu'elle est fondée sur l'équivalence dès lors si *x* P *y* ou *z* P *x* alors non T (y, x, z, α) c'est-à-dire que l'on a pas (*y, x* ; α, (1- α)) E *z* puisque par exemple *z* P *x* donc *z* sera toujours préféré à *x* même s'il est combiné à *y* et ce du fait de l'axiome d'indépendance (cf. H3).

Cette définition de la relation T leur permet d'établir la « Définition 2' : <K, P, E, T> est un modèle de préférences rationnelles au sens deux, si et seulement si pour chaque x, y, z, et w dans K et pour chaque α et β dans [0, 1] :

T1. <K, P, E> est un classement de préférences rationnel (dans le sens de la définition 1) ;

T2. Si x E y et y E z, alors T (x, y, z, α) ;

T3. Si x P y ou z P x, alors non T (y, x, z, α) ;

T4. Si T (x, y, z, α) et x E w, alors T (w, y, z, α) ;

T5. Si T (x, y, z, α) et y E w, alors T (x, w, z, α) ;

T6. Si T (x, y, z, α) et z E w, alors T (x, y, w, α) ;

T7. Si x P z, alors y E z, si et seulement si T (x, y, z, 0) ;

T8. Si x P z, alors x E y si et seulement si T (x, y, z, 1) ;

T9. Si x P z, non y P z et non z P y, alors il existe un unique γ dans [0, 1] tel que T (x, y, z, γ) ;

T10. Si x P y, y P z, et z P w, et pour les deux assertions suivantes, alors on a les suivantes :

T (x, y, w, α)

T (x, z, w, β)

T (y, z, w, β/α)

T (x, y, z, α – β/ 1 – β) » (*ibid*., p. 56-57).

La définition 1 permet à Davidson McKinsey, Suppes de proposer une mesure faible des préférences qui est en fait un classement ordinal. Les deux axiomatiques respectivement fondées sur les définitions 2 et 2' ont pour objectif de montrer l'existence d'une mesure plus forte d'un ensemble rationnel de préférences et ce à partir d'une échelle d'intervalles. La définition 2' prend pour point de départ un K fini, ce qui constitue selon eux une amélioration de la définition 2 où K est infini.

3.2. Une axiomatique finitiste de la probabilité subjective et de l'utilité : Davidson et Suppes (1956)

L'article écrit par Davidson et Suppes en 1956 s'inscrit dans un cadre particulier sur lequel nous reviendrons plus loin, puisqu'il fait partie du « Stanford Value Theory Project » qui constituait dans les années 1950 l'un des pôles les plus importants de la recherche expérimentale aux Etats-Unis avec le pôle de Pennsylania State University représenté par Sidney Siegel et celui du Michigan symbolisé par Ward Edwards. Pourtant il n'est guère fait état d'approche expérimentale dans cet article.

Les auteurs y présentent dans une écriture complexe l'axiomatique qui sera aussi celle du modèle de 1957, établissant une série de définitions, axiomes et théorèmes sans prétendre toutefois réaliser d'expériences.

S'inscrivant dans la lignée de celui de 1955, les auteurs choisissent la dernière des mesures fortes proposées dans l'article de 1955 puis justifient et détaillent son utilisation. L'objectif est en effet de fournir une axiomatique « finitiste » de la probabilité subjective et de l'utilité. Plus précisément, l'idée est de construire une théorie du choix rationnel finitiste.

Le terme « finitiste » peut être compris – même si les auteurs ne le mentionnent pas – comme une allusion aux travaux de David Hilbert (1862-1943) en mathématiques.

La démarche finitiste d'Hilbert nous enjoint de ne considérer qu'un nombre fini d'objets dans les raisonnements mathématiques. Appliquée à la théorie de l'utilité, cette démarche - utilisée dans l'ouvrage de Davidson, Suppes et Siegel – implique de ne considérer que des ensembles finis pour déterminer les fonctions d'utilité et de probabilité subjective.

Comme le soulignent les auteurs dans les premières pages de *Decision Making*, « la théorie ne peut pas être complètement vérifiée, personne ne peut comparer une liste infinie d'issues […] tant qu'il n'y a pas de preuves convaincantes qu'une telle mesure est possible ; l'idéalisation représentée par l'introduction d'ensembles infinis n'a pas de fondement ferme » (Davidson, Suppes, Siegel [1957], p.8). On comprend dès lors, que l'intervalle sur lequel les auteurs vont tenter de mesure l'utilité est un intervalle fini[71].

Sans aller jusqu'à procéder à une expérimentation comme ils le feront en 1957, Davidson et Suppes (1956) approfondissent la proposition faite en 1955 à partir de la définition 2'. Davidson, McKinsey et Suppes (1955) s'étaient alors contentés d'introduire l'idée d'une mesure forte par intervalles de ce qu'ils appelaient un ensemble de préférences rationnelles. Ils ne se préoccupaient guère de définir une fonction d'utilité ou de probabilité subjective.

En 1956, Davidson et Suppes introduisent enfin le terme d'utilité et s'intéressent à la définition et l'existence d'une fonction d'utilité, ce qui passe par l'établissement d'un certain nombre de définitions (3.2.1) et l'utilisation de la méthode opérationnelle de Ramsey (3.2.2). Celles-ci et les axiomes posés, les auteurs en tirent une mesure de l'utilité par intervalles également espacés (3.2.3). L'axiomatique ainsi proposée sera reprise à l'identique en 1957 et fondera le cœur du modèle.

3.2.1 Définitions et axiomes

Rappelons que les définitions et axiomes présentées ici sont ceux qui servent également de socle au modèle de 1957. Davidson et Suppes conservent, en effet, les notations et la structure de l'axiomatique de 1956 dans l'ouvrage qu'ils publient

[71] L'article de Suppes et Winet [1955] tente de dépasser cette contrainte.

l'année suivante. Afin d'éviter toute forme de redondance tout en respectant la chronologie du travail de Davidson, nous ne présenterons qu'une fois cette axiomatique.

Les auteurs commencent par définir les données de base de leur axiomatique :

(1) un ensemble fini K d'issues accessibles à un individu donné, à un temps donné ;

(2) un ensemble X d'éléments qui sont conçus comme survenant avec une certaine probabilité ;

3) $\mathcal{F}$ une famille de sous-ensembles de X (en suivant la terminologie usuelle, les éléments de $\mathcal{F}$ sont appelés des événements, par exemple l'événement consistant à obtenir un nombre pair lors d'un lancer de dé) ;

(4) une relation binaire *P* de préférences sur le domaine K ;

(5) une relation quaternaire *M* qui est utilisée pour formaliser la situation dans laquelle le sujet est indifférent entre l'option présentée dans (1.1) (c'est-à-dire *x*, *y* M(E)[72] *u*, *v* si et seulement si l'individu en question est indifférent entre recevoir *x* si E survient et *y* si $\overline{E}$ survient et recevoir *u* si E survient et *v* si $\overline{E}$) ;

(6) *E**, un élément de $\mathcal{F}$ (comme mentionné, l'interprétation de *E**est que c'est un événement avec une probabilité subjective de ½).

Les deux ensembles K et X sont des allusions directes à la théorie de Savage (Davidson, Suppes [1956], p. 270). En effet, l'ensemble X correspond, dans l'approche de Davidson et Suppes à l'ensemble des états de la nature, et l'ensemble K à l'ensemble des conséquences. Comme chez Savage, les auteurs considèrent qu'une décision ou une action est une application de X dans K, application conditionnée à la partition de X. Seulement, la partition de X est particulière chez Davidson et Suppes. Ces derniers ne s'intéressent qu'au cas où X se décompose en deux parties, E et son complémentaire Ē.

[72] A noter qu'ici E ne désigne plus l'équivalence mais un événement.

Deux définitions sont par ailleurs nécessaires pour présenter les axiomes de la théorie.

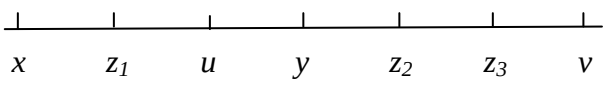
Définition 1 : *x* J *y* si et seulement si *x* P *y* et pour chaque *z* dans K si *x* P *z* alors *y* = *z* ou *y* P *z*.

L'interprétation intuitive de la relation J est que *x* J *y* si et seulement si *y* est l'unique successeur immédiat de *x* relativement à la relation P.
Récursivement, la relation J est définie comme suit :
x J^1 *y* si et seulement si *x* J *y*
x J^n *y* si et seulement il existe un *z* tel que *x* J^{n-1} *z* et *z* J *y*.
L'interprétation intuitive de l'assertion que *x* J^n *y* est que *y* est le n-ième successeur de *x* dans la relation de préférence P.

Définition 2 : r (*x*, *y* ; *u*, *v*) = α si et seulement s'il existe des entiers non négatifs *m* et *n* tels que :

- (i) $\alpha = \frac{m}{n}$;
- (ii) $x \neq y$ ou $u \neq v$;
- (iii) soit *x* J^m *y* et *u* J^n *v* ; ou *y* J^m *x* et *v* J^n *u* ; ou *x* = *y*, m = 0 et (*u* J^n *v* ou *v* J^n *u*) ; ou *u* = *v*, n = 0 et (*x* J^m *y* ou *y* J^m *x*).

On appelle alors r la fonction-ratio (*ratio function*) et son interprétation est simple : le rapport du nombre d'intervalles entre *x* et *y* sur le nombre d'intervalles entre *u* et *v* est α.
Par exemple, si plusieurs issues également espacées en utilité sont présentées comme ici :



Alors, r (*x*, *y* ; *u*, *v*) = ¾.

Puisque K est fini, le nombre d'éléments entre deux éléments est fini et la fonction r est définie pour tous les quadruplets d'éléments de K qui satisfont (ii). L'objectif de (iii) est d'exclure la possibilité qu'on ait à la fois x P y et v P u. Cette assertion n'est pas liée à cette de la fonction r mais plutôt à l'axiome A10 qui sera présenté plus loin.

La fonction-ratio r n'est pas nécessaire pour construire une fonction d'utilité numérique mais elle l'est (ou son équivalent l'est) pour bâtir la fonction de probabilité subjective[73].

3.2.2 La méthode opérationnelle de Ramsey

Cette fonction ratio peut être considérée comme un emprunt direct aux travaux de Ramsey réunis dans un ouvrage paru en 1931 sous le titre *The Foundations of mathematics and other logical essays* (traduit en français en 2003).

En effet, Ramsey, en particulier dans son essai « Vérité et Probabilité » (1926) publié au sein de l'ouvrage de 1931, propose une définition du degré de croyance au travers d'un exemple de la vie quotidienne :

« Imaginons un homme à la croisée des chemins, ne sachant pas où aller mais qui préfère légèrement un chemin à un autre » (Ramsey [1931, 2003], p. 168).

L'individu qui emprunte ce chemin peut être amené à questionner son choix s'il aperçoit au loin une personne pouvant peut-être l'informer. Notre homme peut donc soit faire fi de cette nouvelle information et poursuivre sur sa route tout en sachant qu'il est possible que ce soit le mauvais chemin, ou se détourner de sa route, traverser le chemin et demander conseil. Cette délibération est fonction, selon Ramsey, de la confiance que l'individu a vis-à-vis de son propre jugement mais aussi de l'inconvénient relatif à être sur le mauvais chemin.

Pour répondre à ce problème, Ramsey propose d'utiliser la distance que l'individu serait disposer à franchir comme mesure de la confiance en son opinion (Ramsey [1931, 2003], p. 168). Le théorème de représentation que propose Ramsey indique un

[73] C'est notamment cette méthode qui sera utilisée dans le modèle de Bolker-Jeffrey (voir la seconde partie de la thèse).

isomorphisme entre la structure des distances et la structure des utilités. Ainsi, la distance entre *a* et *b* sera égale à la distance entre *c* et *d* si et seulement si la différence d'utilité entre *a* et *b* est égale à la différence d'utilité entre *c* et *d* (Sahlin [1990], p.30).

Comme nous l'avons mentionné la ligne directrice de Davidson, McKinsey et Suppes en 1955 est de proposer une analyse des conditions formelles du choix rationnel tout comme Ramsey avait utilisé la logique pour définir les conditions formelles de la croyance rationnelle. Voyons à présent en détails la méthode proposée par Ramsey afin de saisir la méthode utilisée par Davidson et Suppes en 1956

L'objectif de Ramsey dans « Truth and Probability » (1926) est de mettre en évidence la connexion entre le degré subjectif de croyance que nous avons dans une proposition *p* et la probabilité que nous lui attribuons[74].

Plus précisément, Ramsey montre comment mesurer le degré de croyance qu'un agent a dans une proposition donnée. Si cet agent suit un certain nombre de normes de rationalité, alors le degré de croyance peut être représenté par une mesure qui satisfait, selon Ramsey, les lois mathématiques de la probabilité.
Pour cela, Ramsey propose une méthode « opérationnelle », c'est-à-dire une règle de mesure permettant de représenter numériquement (à l'aide de différences de valeurs) l'estimation qu'a une personne de la possibilité que tel événement survienne. Suivant cette méthode, il est possible de proposer une théorie qui consiste en un ensemble d'axiomes qui requièrent une conduite rationnelle de la part de l'agent. Cette théorie permet, à partir des axiomes, de calculer à la fois les utilités cardinales (c'est-à-dire les valeurs subjectives de l'individu sur des issues) ainsi que le degré de croyance en une proposition qui conditionne les issues et qui a une incidence sur le comportement de pari de l'individu. Et cette mesure du degré de croyance satisfait les lois de la théorie de la probabilité, en d'autres termes c'est une mesure de la probabilité (subjective). La clé de la méthode de Ramsey permettant de mesurer le degré de croyance d'un agent vis-à-

[74] Nous nous appuierons pour décrire la méthode de Ramsey sur Sahlin [1990].

vis d'une proposition arbitraire *p* - c'est-à-dire une méthode qui permette d'obtenir l'évaluation subjective du sujet - consiste à postuler l'existence de « propositions éthiquement neutres » (Ramsey [1931, 2003], p. 170).

Une proposition atomique *p* est dite éthiquement neutre si deux « mondes possibles »[75] qui diffèrent seulement en fonction de la vérité de *p* sont toujours de valeur égale[76].

La découverte d'une telle proposition éthiquement neutre résout le problème, selon Ramsey, consistant à séparer les probabilités subjectives et les utilités subjectives.

A partir de ces axiomes, il est possible d'établir un théorème de représentation permettant de représenter les préférences de l'individu par une fonction d'utilité.

(1) Il existe une proposition éthiquement neutre *p* crue au degré ½

(2) Si p et q sont de telles propositions et si l'option α si p, δ si non p est équivalente à l'option β si p, γ si non p, alors l'option α si q, δ si non q, est équivalente à l'option β si q, γ si non q, c'est-à-dire $\alpha\beta = \gamma\delta$.

(3) Si $\alpha\beta = \gamma\delta$ alors $\alpha > \beta$ est équivalent à $\gamma > \delta$ et $\alpha = \beta$ est équivalent à $\gamma = \delta$.

(4) Si $\alpha\beta = \gamma\delta$ et $\gamma\delta = \eta\zeta$ alors $\alpha\beta = \eta\zeta$

(5) Pour tout triplet (α, β, γ), il existe un unique *x* tel que $\alpha x = \beta y$

(6) Pour toute paire (α, β), il existe un unique *x* tel que $\alpha x = x\beta$.

(7) Axiome de continuité : toute progression a une limite

(8) Axiome d'Archimède : aussi faible que soit la distance de valeur entre *a* et *b*, et aussi grande la distance entre *c* et *d*, il existe un entier *n* tel que *n* fois d(a, b) est plus grand ou égale à d(c,d).

A partir de ces huit axiomes, Ramsey affirme qu'il est possible d'attribuer à chaque résultat *a* une utilité définie *u(a)* qui est un nombre réel tel que :

d(a,b) = d(c,d) si et seulement si u(a) – u(b) = u(c) – u(d)

Les préférences de l'agent peuvent être représentées par une fonction d'utilité u(.) définie à une transformation affine positive près[77].

---

[75] Selon les termes de Ramsey.

[76] Voyons un exemple de propositions éthiquement neutres que nous empruntons à Sahlin [1990] : On note *a* le résultat « Il va faire beau demain » et *b* « Il va pleuvoir demain ». Soit *p* l'événement suivant lequel le dé va tomber sur 1, 2 ou 3 au prochain lancer.
Si *a* est préféré à *b* et (*a* si *p*) est préféré à (*b* si *p*) alors *p* est une proposition éthiquement neutre.
Par ailleurs, une proposition éthiquement neutre est crue au degré ½ si l'agent est indifférent entre les deux options suivantes :
Option 1 : *a* si *p* est vraie, *b* si ¬*p* est vraie
Option 2 : *b* si *p* est vraie, *a* si ¬*p* est vraie

3.2.3 L'utilité espérée

Ce qui nous intéresse ici, c'est de montrer comment Davidson et Suppes vont exploiter cette méthode opérationnelle de Ramsey pour tester l'existence d'une mesure de l'utilité par intervalles également espacés en 1957 en présentant, avant cela, une axiomatique du choix rationnel analogue à l'axiomatique de la croyance rationnelle de Ramsey.

En effet, à partir de leurs deux définitions fondamentales (déterminées au 3.2.1), les auteurs sont en mesure de présenter une série d'axiomes permettant de décrire une structure de choix rationnel finitiste. Ce sont ces axiomes qui seront utilisés en 1957.

Définition 3.1. Soit K un ensemble fini, soit X un ensemble et $\mathcal{F}$ une famille d'ensembles de X qui est fermée au complémentaire et dont X est un membre. Soit P un sous ensemble K x K et soit M[78] en sous ensemble de K x K x $\mathcal{F}$ x K x K. Soit E* un membre de $\mathcal{F}$. Alors < K, X, $\mathcal{F}$, P, M, E*> est une Structure de Choix Rationnel Finitiste si et seulement si pour chaque *x, y, u, v, x', y', u',* et *v'* dans K et pour tout *E* et *E'* dans $\mathcal{F}$:

A1 : P est un ordre simple dans K ;
A2 : Si $x \neq$ y, alors *x, y* M(E*) *y, x* ;
A3 : Si *x, y* M(X) *u, v,* alors $x = y$;
A4 : Si *x, y* M(E) *u, v,* alors $x \neq$ y et $u \neq v$;
A5 : Si *x, y* M(E) *u, v,* alors *u, v* M(E) *x, y ;*
A6 : Si *x, y* M(E) *u, v,* alors *y, x* M(Ē) *v, u* ;
A7 : Si *x, y* M(E) *u, v* et *u, v* M(E) *u', v',* alors *x, y* M(E) *u', v'* ;
A8 : Si *x, y* M(E) *u, v* et $y \neq v$ et *x* P *u* alors *v* P *y* ;
A9 : Si *x* J *y, u* J *v,* $x \neq v$ et $u \neq y$ alors *x, v* M(E*) *u, y* ;

---

[77] Mais, comme le souligne Sahlin, cette définition n'est pas suffisante pour connecter le degré de croyance dans une proposition avec la probabilité subjective dans cette proposition. Pour faire cela nous avons besoin des probabilités conditionnelles (Sahlin [1990], p.36).
[78] Notons que la relation M est en fait la relation ≈ de *Decision Making*.

A10 : Si r (x, y ; u, v) = r (x', y' ; u', v') et x, v M(E) y, u et $x \neq v'$ et $y' \neq u'$, alors x', v' M(E) y', u' ;
A11 : Si x, y M(E) u, v et x', y' M(E) u', v' et E ⊆ E', alors r (x, y ; u, v) ≤ r (x', y' ; u', v') à condition que $x \neq u$ ou $y \neq v$ et $x' \neq u'$ ou $y' \neq v'$.
A12 : Il existe des éléments x, y, u et v dans K tel que x, y M(E) u, v et $x \neq y$ ou $y \neq v$.

L'axiome A1 correspond à une relation d'ordre sur l'ensemble K. L'axiome A2 assure que la probabilité subjective de E* soit ½, autrement dit, les auteurs cherchent, comme Ramsey, à définir une proposition éthiquement neutre crue au degré ½ comme nous l'avons expliqué plus haut. L'axiome A3 indique que les deux issues x et u suffisent à déterminer l'équivalence de (x, y) et (u, v). L'axiome A4 exclut la possibilité de paris certains[79]. Les axiomes A5, A6, A7 renvoient aux hypothèses usuelles de symétrie et de transitivité. L'axiome A8 renvoie à l'idée que si l'option (x, y) est équivalente à l'option (u, v) et que x est préféré à u alors v est préféré à y[80]. A9 renvoie directement à l'idée de « successeur immédiat » évoqué plus haut dans l'article de 1955. L'axiome A10 exprime la condition d'équivalence de deux ratios[81]. Enfin l'axiome A11 est un axiome de cohérence et l'axiome A12 garantit simplement que les éléments de K nous permettent de déterminer la probabilité subjective pour chaque élément dans $\mathcal{F}$.

A partir de ces axiomes, Davidson et Suppes présentent un certain nombre de théorèmes dont le plus important, le théorème de l'utilité espérée [82].

Théorème 5.1 : Soit < K, X, $\mathcal{F}$, P, M, E*> une structure faible de probabilités subjectives alors :
(A) Il existe une paire ordonnée < u, p > de fonctions à valeurs réelles où $u(.)$ est définie sur K et $p(.)$ sur $\mathcal{F}$ telle que pour tout x, y, u et v et pour tout E et F dans $\mathcal{F}$:
(i) x P y si et seulement si $u(x) > u(y)$

[79] On verra plus loin que cet axiome est révélateur d'une critique adressée à Mosteller et Nogee.
[80] Comme le mentionnent les auteurs, la condition que y et u soit distincts est insérée pour se prémunir du cas limite où la probabilité subjective de E est 0.
[81] Tels qu'ils ont été définis dans l'article de 1955 et dans le paragraphe précédent.
[82] Pour une présentation de ces théorèmes, voir Davidson et Suppes [1956], p.270.

(ii) p (E) ≥ 0

(iii) p (X) = 1

(iv) p (E) + p ($\bar{E}$) = 1

(v) si E $\subseteq$ F ; alors p(E) $\leq p(F)$

(vi) x, y M(E) u, v si et seulement si $u(x) \neq u(y)$et $u(u) \neq u(v)$ *et*

$p\,(E)\,u(x) + p\,(\bar{E})\,u\,(y) = p\,(E)\,u(u) + p\,(\bar{E})\,u\,(v)$;

(B) Si $n^* > 5$ et si $< u_1, p_1 >$ et $< u_2 , p_2 >$ sont deux paires ordonnées de fonctions qui satisfont (A) alors :

(i) $p_1 = p_2$

(ii) Il existe des nombres réels a et b avec $a > 0$ tel que pour tout x *dans* K

$u_1(x) = a\,u_2(x) + b.$

(A) est le théorème de l'utilité espérée et (B) la classe des fonctions d'utilité définies à une transformation affine croissante près.

3.3. Tests empiriques de la théorie de la mesure de l'utilité et de l'axiomatique

Les articles de 1955 et 1956 justifient et donnent à lire pour la première fois l'axiomatique de l'utilité espérée telle qu'elle sera reprise en 1957 et connue comme celle du modèle de Davidson. Toutefois l'ouvrage *Decision making* écrit en 1957 par Davidson, Siegel et Suppes est, à plusieurs égards, spécifique.

D'une part, les auteurs y amorcent pour la première fois toute une série de critiques à l'encontre des autres théories du choix rationnel comme celle de Savage (1954) (3.3.1).

D'autre part, *Decision Making* (1957) peut-être considéré comme un ouvrage pionnier de l'économie expérimentale (3.3.2). C'est en effet en réaction aux premières expériences des théories des choix individuels (en particulier celle de Mosteller et Nogee, 1951) que la théorie de Davidson, Siegel et Suppes constituée, rappelons-le, d'un modèle et d'une expérience se construit. Et le point d'ancrage de la remise en cause des expériences de Mosteller et Nogee se trouve dans la méthode

opérationnelle de Ramsey que ces derniers avaient mentionnée sans l'utiliser alors que celle-ci apporte une réponse directe au problème de la mesure des utilités et des probabilités.
La présentation détaillée de la procédure expérimentale attachée au modèle de 1957 permettra de le vérifier (3.3.3).

3.3.1 Critiques de Davidson, Suppes et Siegel à la théorie de la décision

Comme Savage, Davidson, Suppes et Siegel considèrent que le choix entre des issues risquées fait appel à deux facteurs au moins : « le degré auquel les résultats possibles sont désirés les uns par rapport aux autres et le degré auquel les résultats sont jugées probables » (Davidson, Suppes, Siegel [1957], p.1).
Toutefois, la critique générale la plus sérieuse que formulent Davidson, Suppes et Siegel aux précédentes théories formalisées de la décision parmi des issues risquées est qu'elles ne sont pas, en l'état, empiriquement testables (*ibid.* p.3). Cette critique prend deux formes :

- La première est exprimée ainsi : les agents ne satisfont pas les conditions de rationalité, la question alors se pose de savoir si les préférences sont véritablement transitives.
- La seconde est qu'aucune « interprétation empirique satisfaisante » (*ibid.*) au sens expérimental de la théorie n'a été donnée et dès lors il est impossible de la tester.

La seconde critique est la plus fondamentale car elle est relative à la testabilité de la théorie.
Ainsi, pour Davidson, Siegel, et Suppes, proposer une théorie, comprenant des axiomes qui la sous-tendent, c'est en même temps permettre de la tester.

Précisons que, bien qu'ils parlent de manière ambigüe d' « interprétation empirique » ou de « théorie empiriquement applicable », il ne s'agit guère de produire une théorie descriptive. Ils s'inscrivent dans un cadre normatif. Toutefois, dans l'introduction de

Decision Making, les auteurs expliquent que la normativité n'exclue pas la description :

> « Si personne ne s'accorde avec la théorie, on peut vouloir se demander si nos décisions sont rationnelles dans le sens de la théorie. (…) Une théorie normative de la décision rationnelle doit disposer d'un intérêt pratique et doit donc pouvoir être appliquée empiriquement. Or, si la théorie est capable d'application empirique, il est possible qu'elle soit vraie en tant que théorie descriptive » (*ibid.* p.4).

Il y a donc, pour les auteurs, une connexion nécessaire entre la dimension descriptive et la dimension normative de la théorie de la décision rationnelle : « Il est tout à fait possible que, dans certaines situations, les gens agissent en accord avec certains canons de la rationalité. De même, des personnes tout à fait ignorantes de la logique formelle raisonnent le plus souvent en accord avec celle-ci » (*ibid.*).

Dans chaque cas, à partir du moment où la théorie peut être testée, la question de savoir si elle est vraie – c'est-à-dire si elle est vérifiée - sous des circonstances données est un fait relatif à l'expérimentation.

Ainsi, en suivant la méthodologie initiée par vNM – puis reprise par Savage, les modèles présentés par Davidson, Suppes et Siegel dans l'ouvrage de 1957 sont caractérisés par une liste d'axiomes qui contiennent en plus de l'apparatus usuel des mathématiques et de la logique, des termes primitifs qui se réfèrent à des ensembles, ensembles relatifs aux issues et aux événements comme chez Savage. Les axiomes de Davidson, Suppes et Siegel doivent, selon eux, permettre :

- d'assigner des nombres à des éléments de n'importe quel ensemble à partir duquel le modèle est construit de telle sorte à ce qu'il préserve la structure imposée à de tels ensembles par les axiomes – autrement dit le morphisme doit être respecté entre l'ensemble des nombres et l'ensemble utilités sur les gains par exemple, et,
- que toute paire d'assignation de nombres en accord avec la première condition soit reliée de manière spécifique par un groupe de transformations.

Les modèles présentés dans l'ouvrage de 1957 divergent cependant de ceux de vNM et Savage non pas en demandant quelque chose de différent au décideur rationnel mais en demandant moins : « Ces modèles sont non conventionnels dans la mesure

où ils sont plus modestes mais la limite à la modestie est que les conditions pour la mesure, au moins dans un domaine limité, doivent exister » (Davidson, Suppes, Siegel [1957], p.6). Cette modestie a trait au caractère expérimental de la théorie. En effet, les hypothèses doivent être formulées de telle sorte à être testables en termes de comportement. C'est sur cette base que Davidson, Suppes et Siegel vont construire leur théorie.

3.3.2 *Decision Making*, un ouvrage pionnier de l'économie expérimentale

Même si des traces de ce que l'on qualifie aujourd'hui d'économie expérimentale peuvent être décelées dans des travaux antérieurs aux années 1930, on peut comme le suggère Alvin Roth [1995], prendre comme point de départ la période allant des années 1930 jusqu'aux années 1950 pour mettre en évidence l'émergence de thèmes et de recherches qui constituent les fondements de l'économie expérimentale moderne. Cette période peut, en effet, être qualifiée de période pionnière. L'ouvrage de vNM constitue, de ce point de vue, l'ouvrage de référence en tant que première axiomatique moderne de l'utilité espérée[83] (Guala [2008]).

- **Une économie expérimentale hétéroclite**

Cette période pionnière a plus précisément vu naître ce qu'il convient d'appeler trois branches de l'économie expérimentale (Roth [1995], p.5) qui constituent aussi trois objets différents d'expérience.

La première regroupe des recherches portant sur les théories du choix individuel. Le premier test qui marqua cette branche de l'économie expérimentale est l'expérience de Thurstone en 1931 lorsque celui-ci tenta de représenter les préférences d'individus à partir de choix hypothétiques parmi des paniers de biens composés de chapeaux, de chaussures et de vestes.

La seconde branche regroupe les tests portant sur la théorie des jeux telle que présentée par vNM (1944). On retrouve ici des recherches ayant trait non seulement

[83] Même si, comme on l'a vu, l'objectif de vNM n'est pas expérimental mais axiomatique.

aux jeux stratégiques mais aussi aux processus d'apprentissage en situation de jeux comme les travaux de Suppes et Atkinson en 1960.
La troisième et dernière branche rassemble les recherches portant sur l'entreprise, le mécanisme des prix et la modélisation de marchés expérimentaux comme dans l'expérience de Chamberlin en 1948.

Ces différentes branches qui constituent ce que l'on appelle « l'économie expérimentale » - dont les premières expériences majeures ont lieu entre le début des années 1930 et la fin des années 1950 – ont des statuts particuliers. En effet, bien qu'on ait aujourd'hui coutume de rassembler ces travaux sous la bannière « économie expérimentale », il ne s'agit pas à l'époque de travaux institutionnellement situés. Ils sont en particulier largement liés à la psychologie expérimentale, et géographiquement dispersés.

Ainsi, comme le mentionne Francesco Guala (2008), la grande majorité[84] des recherches portant sur l'économie expérimentale aux Etats-Unis dans les années 1950 avait lieu au sein de trois grands pôles constitués par autant d'universités américaines.
Le premier grand pôle était celui de l'université de Pennsylvanie avec comme principal représentant Sidney Siegel.
Le deuxième grand pôle était celui de l'université du Michigan dont le pionnier était Ward Edwards et dont le successeur est Amos Tversky. Notons dès à présent que le programme de recherche initié par Edwards dans les années 1950 puis poursuivi par Tversky et Kahneman à la fin des années 1970 constitue ce qu'on appelle habituellement des modèles d'utilité non-espérée (Machina [2008]), nous y reviendrons en détails (voir fin du chapitre).
Enfin, troisième et dernier pôle, celui de l'université de Stanford au travers du « Stanford Value Theory Project » dont les représentants sont Donald Davidson et Patrick Suppes.

[84] Certaines expériences font toutefois exception comme celles de Mosteller et Nogee, tous deux professeurs à Harvard.

Les expériences de Davidson, Suppes et Siegel regroupées dans *Decision Making* (1957) appartiennent à la première branche, qui concerne, comme on l'a dit la théorie du choix individuel et dont Thurstone (1931) est le pionnier.

- **La première expérience de Thurstone**

L'expérience de ce dernier, même si elle se fonde sur une conception strictement ordinale de l'utilité, est révélatrice d'un certain type d'exigences et d'une démarche méthodologique embryonnaire qui laisse des traces puisqu'après celle-ci les auteurs s'efforceront de défendre leurs points de vue expérimentaux et leurs méthodes pour assurer la validité et la pertinence de leurs expériences.

Thurstone proposait de recueillir des données permettant de représenter les préférences des individus par des courbes d'indifférence. Ces données consistaient en des choix parmi des paniers de biens hypothétiques. L'idée était de proposer des combinaisons de ces biens de telle sorte à ce que l'on obtienne les taux d'échange nécessaires pour construire les courbes. Près de dix ans plus tard, dans leur article « The empirical derivation of indifference functions », Wallis et Friedman (1942) fustigent la méthode employée par Thurstone. Pour eux, l'expérience de ce dernier s'appuyait sur des choix non seulement hypothétiques mais surtout mal spécifiés car rien n'assurait que les individus prendraient les mêmes décisions s'ils étaient placés dans une situation de choix véritable (Wallis et Friedman [1942], p.179).

- **Preston et Baratta et la déformation subjective**

Il faudra attendre 1948 et l'expérience de Preston et Baratta pour assister à la première tentative de test expérimental impliquant des probabilités et leurs déformations subjectives.

Ces deux auteurs s'intéressent plus précisément à tous les éléments psychologiques qui peuvent déformer le calcul objectif des espérances mathématiques portant sur un jeu de casino comme la roulette. Par exemple, selon Preston et Baratta, un grand nombre de personnes imaginent que la probabilité de voir survenir la couleur rouge à la roulette est supérieure à ½ si l'on vient d'assister à une longue série de couleur

noire (Preston et Baratta [1948], p.183). Ces considérations ont, selon les auteurs, une influence directe sur le prix que serait prêt à payer toute personne désirant jouer à la roulette. De même, s'il ne manque que 50 points à un individu pour remporter une partie qui l'oppose à d'autres participants, il accordera une valeur particulière à la partie finale qui lui permettra de gagner, et ceci peut le conduire à ne pas se comporter en conformité avec le calcul des espérances mathématiques. Pour mettre en évidence ces assertions, Preston et Baratta organisent un jeu tout spécialement conçu pour l'occasion dans lequel les individus sont amenés à acheter aux enchères des cartes (au nombre de 42) composées de points *x* (6 possibilités : 5, 50, 100, 250, 500 ou 1000 points) et de probabilités (0.01, 0.05, 0.25, 0.50, 0.75, 0.95, et 0.99). L'idée était donc de vendre aux enchères l'opportunité de gagner *x* points avec une probabilité *p*.

Le résultat de l'expérience des deux auteurs de l'université de Pennsylvanie est notamment qu'il existe une échelle de probabilités psychologiques qui peut diverger de l'échelle objective des probabilités et que les deux courbes reliant respectivement les déformations psychologiques des attributions de probabilités (courbe de probabilités psychologiques) et les probabilités objectives avaient au moins un point en commun. En dessous de ce point, c'est-à-dire lorsque la courbe de probabilités psychologiques se situe en dessous de celle des probabilités objectives (représentée par la première bissectrice), les auteurs considèrent que les individus surestiment les probabilités objectives et lorsque la courbe de probabilités psychologiques se place au dessus de celle des probabilités objectives, ils considèrent que les individus sous-estiment les probabilités objectives liées au calcul des espérances (Preston et Baratta [1948], p. 193).

L'intérêt de la démarche de Preston et Baratta, même si elle n'est pas très éclairante du point de vue du test de la théorie de l'utilité espérée (puisque les enchères n'étaient pas linéaires en probabilités alors que les points l'étaient) est qu'elle constitue une première ébauche tout à fait instructive de ce que seront les expériences ultérieures. Ces expériences furent critiquées notamment pour la méthode utilisée et l'absence de test rigoureux des hypothèses prises une à une comme ce sera le cas avec Davidson, Suppes et Siegel [1957].

- **Mosteller et Nogee, le test de la théorie de vNM et de la mesure d'utilité par intervalles**

Lorsque Davidson, Siegel et Suppes écrivent *Decision Making*, un seul rapport concernant une expérience construite pour dériver un intervalle de mesure pour l'utilité à partir de véritables choix a été publié, c'est l'article de Mosteller et Nogee (1951). La théorie et l'expérience de Davidson, Siegel et Suppes (1957) sont originalement inspirées par le désir de voir s'il est possible d'améliorer les résultats de Mosteller et Nogee.

En suivant une suggestion faite par Friedman et Savage (1948), Mosteller et Nogee avaient décider de tester « la validité empirique de l'axiomatisation de l'utilité de vNM appliquée aux issues consistant à gagner et perdre de faibles montants d'argent et des combinaisons de probabilité de ces issues » (Davidson, Siegel et Suppes,1957, p.20).

L'idée de Mosteller et Nogee est de rapporter une expérience de laboratoire dans laquelle ils ont tenté de mesurer de manière restrictive la valeur assignée par les individus à des revenus monétaires (Mosteller et Nogee [1951], p. 371). L'objectif de cette expérience est double :

- premièrement, il s'agit de déterminer si l'utilité peut être mesurée en toutes circonstances ;
- deuxièmement, les auteurs cherchent à montrer comment une mesure de l'utilité dans une situation donnée peut permettre de prédire le comportement d'individus placés dans des situations différentes.

C'est pourquoi les auteurs construiront leur expérience en plusieurs étapes :

a. Ils recrutent leurs sujets pour participer à un jeu offrant la possibilité de prendre ou de refuser des paris, jeu impliquant l'usage de monnaie.
b. Puis, à partir du comportement observé au cours du jeu, les auteurs construisent une courbe d'utilité pour chaque sujet.
c. Cette courbe est ensuite utilisée pour faire des prédictions relativement au comportement futur de l'individu face à des paris plus complexes.

Enfin, ces prédictions sont testées (en examinant le comportement des sujets face ces paris plus complexes).

Pour Mosteller et Nogee, en effet, lorsque les économistes ont cherché à tester la validité de la notion d'utilité, ce fut toujours en la considérant comme « une variable comme une autre » (ibid., p. 371). Il s'agit donc ici pour eux de tester une notion plus complexe d'utilité, dérivée de leur lecture de Friedman et Savage et qu'ils résument ainsi « les individus se comportent comme s'ils avaient une échelle subjective de valeurs à attribuer aux différents montants de biens - cette échelle n'étant, selon eux, pas nécessairement une simple dérivation ou translation de celle « physique utilisée sur le marché » (Mosteller et Nogee, 1951, p. 371).

L'expérience proposée trouve donc sa source dans l'article de Friedman et Savage (1948)[85]. Voici la formulation de Friedman et Savage :

« En choisissant parmi des issues ouverte à lui, qu'elles soient risquées ou non, l'utilité de consommation se comporte comme si (a) elle avait un ensemble cohérent de préférences, (b) ces préférences peuvent être complètement décrites par une fonction attachant une valeur numérique – l'utilité – à chaque issue considérée comme certaine, (c) son objectif étant de rendre son *utilité espérée* la plus grande possible. L'apport de VNM est d'avoir montré que le système de préférences doit avoir les propriétés suivantes : 1) le système est complet et cohérent sachant que les préférences portent sur des objets qui peuvent être combinés avec des probabilités ; 2) si A est préféré à B alors A si p, C si non-p est préféré à B si p, C si non-p ; 3) si A est préféré à B et B est préféré à C, il existe une combinaison de probabilité telle que A si p, C si non-p est indifférent à B ».

C'est cette position théorique qui sert de socle à l'expérience de Mosteller et Nogee. Ainsi, afin de la tester expérimentalement, ils la traduisent dans les termes suivants : les objets A, B, et C représentent des montants de gains ou de pertes d'argent tels que

[85] Mosteller et Nogee ont aussi pu avoir accès à une version remaniée de cet article (ajoutant de la précision dans la détermination de la courbe d'utilité) (Mosteller et Nogee [1951], p.373).

A corresponde à l'issue consistant à obtenir 25¢, B à l'issue consistant à ne rien recevoir et à ne pas perdre d'argent et C à perdre 5 ¢.

Selon Mosteller et Nogee, une majorité d'individus préfère A à B et B à C. Selon la proposition 3) de Friedman et Savage ci-dessus, si les préférences sont cohérentes et si la rationalisation du comportement par l'utilité est correcte, il existe une combinaison de probabilité de A et de C telle que l'individu est indifférent entre la combinaison de A-C et B. En posant U(X) pour l'utilité de X, cela revient à écrire qu'il existe une probabilité p telle que :

$$p.U(A) + (1-p)\, U(C) = U(B)$$

En remplaçant U(A), U(B) et U(C) par les valeurs définies ci-dessus, on trouve :
$p\, U(25¢) + (1-p)\, U(-5¢) = U(0¢)$.
Cette équation suggère, selon Mosteller et Nogee, qu'en disposant de trois valeurs monétaires comme ici, il suffit de faire varier p jusqu'à ce que le point d'indifférence soit trouvé. Pourtant, plutôt que partir à la recherche de ce p, les expérimentateurs choisissent de fixer arbitrairement les utilités de B et de C – c'est-à-dire les utilités, pour l'agent, des sommes 25¢ et -5¢ - ainsi qu'une probabilité p_0 afin de déterminer un A qui fournira l'équilibre.
Ainsi, en posant l'utilité de A comme inconnue, et en assignant arbitrairement des valeurs à U (0¢) et U (-5¢) comme par exemple U (0¢) = 0 utile et U (-5¢) = -1 utile où le terme « utile » représente l'unité de valeur, on doit trouver un A tel que pour l'individu :

$$p_0\, U(A) + (1- p_0)\, U(-5¢) = U(0¢)$$

A partir de cette équation, et en remplaçant par les valeurs, on en déduit

$$U(A) = \frac{1-p_\circ}{p_\circ}.$$

Ainsi, pour toute probabilité p_0 de gagner A, l'utilité de A est connue grâce à cette dernière équation. L'important, dans la démarche de Mosteller et Nogee, est que A soit déterminé expérimentalement. Plus précisément, en participant au jeu, l'individu va exprimer indirectement la valeur qu'il accorde à A.

Pour relier leur expérience à la théorie proposée par Friedman et Savage, Mosteller et Nogee tentent de classer les individus relativement à leur choix de parier ou non dans telle ou telle situation. Ainsi, dans un graphique ou les cents dont portés en abscisse et les utilités en ordonnée, la première bissectrice représente selon les auteurs une « offre mathématiquement juste ». Il s'agit alors de classer les individus en fonction de leur position par rapport à cette bissectrice : les individus se plaçant au dessus de celle-ci seront considérés comme « extravagants » car ils choisissent des paris injustes mathématiquement parlant, alors que ceux qui se situent en dessous de la bissectrice seront qualifiés de « conservateurs » car ne prenant que les paris qui offrent plus que ceux situés sur la bissectrice.
L'idée des auteurs est donc d'une part de construire une courbe d'utilité pour chaque individu et, si cela est possible, de classer ces individus les uns par rapport aux autres en fonction de leur comportement de paris d'autre part.

Ce faisant, bien que ne présentant qu'un modèle très simplifié et laissant dans l'ombre de nombreuses questions concernant le passage à l'expérience, l'article de Mosteller et Nogee ajoutent aux débats autour de la notion d'utilité une dimension expérimentale. Leur expérience servira de point de référence à Davidson.

Davidson, Siegel et Suppes font trois grands reproches à l'approche de ces derniers :

(1) Une procédure expérimentale trop succincte

La procédure de Mosteller et Nogee n'offrirait pas de vérification systématique du fait que les nombres assignés aux mesures d'utilité sont uniques à une transformation linéaire positive près. Pour Davidson, Siegel et Suppes, Mosteller et Nogee ne tiennent pas leur promesse, ils s'avèrent incapable de proposer une mesure de l'utilité au sens d'une échelle d'intervalles.
En effet, Mosteller et Nogee considèrent qu'ils ont mesuré l'utilité d'une issue donnée pour un sujet quand (a) ils ont trouvé un modèle empirique de réponses qui peut être interprété comme montrant que le sujet était indifférent entre les deux options mentionnées plus haut, et (b), en utilisant l'équation (2.1) (Ø(0¢) = s(E) Ø(-

5¢) + [1-s(E)] ∅(*x*)) ils ont assigné un nombre à l'issue. Puisqu'il n'y a aucune garantie à l'avance que les modèles de réponses de la sorte demandés dans (a) soit trouvés, on doit être d'accord sur le fait que le sens dans lequel est utilisé le terme de mesure n'est pas trivial.

D'un autre côté, on doit souligner, selon les auteurs, que l'usage de l'équation 2.1 pour assigner des nombres, et la possibilité de la représenter par un graphique dans lequel plusieurs courbes peuvent être comparées significativement (par exemple la courbe du sujet et la première bissectrice qui symbolise la linéarité de l'utilité en monnaie) est seulement justifié dans un cadre expérimental, pour des situations relatives à la théorie alors les nombres sont uniques à une transformation linéaire croissante près. Mais pour Davidson, Suppes et Siegel, la procédure décrite par (a) et (b) ne donne aucune preuve de la manière dont cela pourrait être le cas.

Ils illustrent leur reproche en donnant un exemple : « Imaginons, sur la base des choix des individus, que l'on assigne aux issues a, b, c et d les utilités 0, 1, 2, 3 en utilisant l'équation (2.1). Si l'attribution est unique à une transformation linéaire croissante près, cela signifie que l'intervalle entre a et b est le même que l'intervalle entre c et d, et l'intervalle entre b et c le même que celui ente c et d. Chacune de ces conséquences peut être traduite en une affirmation sur les préférences des individus révélée par ses choix et qui peut être testée. Si ces prédictions ne sont pas vérifiées, alors il est difficile de dire quelle signification attacher aux nombres originellement attribués aux issues ; à tout niveau ils ne mesurent pas les utilités dans le sens d'un intervalle de mesure demandé par vNM » (Davidson, Siegel et Suppes, 1957, pp.22-23)

La critique majeure n'est pas que les résultats ne sont pas concluants mais que l'expérience n'était pas conçue pour construire un test clair pour savoir si un intervalle de mesure de l'utilité est possible.

(2) Un biais de participation au jeu

Une deuxième critique adressée par Davidson, Siegel et Suppes à la conception expérimentale de Mosteller et Nogee est que la plupart des choix offerts aux sujets étaient des choix entre accepter ou refuser un pari. Dès lors, l'une des options implique

toujours de jouer et de prendre un risque alors que l'autre correspond au fait de ne pas jouer et de ne pas gagner ni perdre d'argent. Davidson, Siegel et Suppes considèrent qu'il existe une utilité positive ou négative à la participation du jeu, dont l'expérience de Mosteller et Nogee ne tient pas compte et qui produit pourtant « une distorsion maximale » (Davidson, Siegel et Suppes, 1957, p. 23). C'est précisément ce que tentent de limiter Davidson, Suppes et Siegel.

(3) La disparition des probabilités subjectives

Selon Davidson, Siegel et Suppes, « Mosteller et Nogee ont supposé que la probabilité subjective d'un événement était égale à sa probabilité objective (1957, pp.23 -24). Davidson, Siegel et Suppes vont alors proposer la première expérimentation de l'hypothèse d'utilité espérée et de mesure par intervalles avec des probabilités subjectives en utilisant la méthode opérationnelle de Ramsey.

3.3.3 Cadre théorique

Davidson, Suppes et Siegel avancent l'idée – assez floue et utilisée notamment par Milton Friedman – des hypothèses expérimentales : « Etant donné un modèle formel de décision et une interprétation opérationnelle de ce modèle, on pourrait formuler des hypothèses expérimentales. Ces hypothèses sont, évidemment, des conséquences logiques du modèle et de l'interprétation qui y est attachée ; si le modèle et l'interprétation sont vrais, alors les hypothèses expérimentales, si elles sont soumises au test, seront vérifiées » (Davidson, Suppes, Siegel [1957], pp.4-5). Sans mentionner directement Friedman, la posture épistémologique des auteurs est étonnement proche de celle de ce dernier dans son article de 1952 écrit en collaboration avec Savage. Ainsi, selon Friedman, une hypothèse scientifique doit nous permettre de prédire des phénomènes qui n'ont pas encore été observés (Friedman et Savage [1952], p.465). Plus précisément, même si nous ne disposons pas de preuves irréfutables de la validité de l'hypothèse d'utilité espérée, par exemple, nous pouvons faire appel, selon Friedman et Savage à des preuves

indirectes. Celles-ci ont trait à la cohérence de cette hypothèse avec le reste de la théorie économique. Mieux, la validité de celle-ci est assurée par le caractère plausible des postulats théoriques sur lesquels cette hypothèse est fondée logiquement. Les hypothèses et les postulats sont équivalents logiquement (Friedman et Savage [1952], p.466).

Ce processus itératif entre le modèle et les hypothèses éclaire le projet des auteurs. Selon eux, « un modèle est caractérisé par une liste d'axiomes qui contiennent des termes primitifs relatifs à des ensembles. Ces ensembles peuvent être interprétés dans certains cas comme des ensembles d'issues ou de résultats monétaires et dans d'autres cas comme des ensembles d'événements probables et de relations qui sont interprétées comme exprimant divers types de préférences parmi des résultats ou des jugements subjectifs de probabilités qualitatives» (Davidson, Suppes, Siegel [1957], p.5).

On retrouve l'approche de Savage, avec ses trois relations de préférence :

- Celle qui concerne ce que Savage appelle « actions », fonction $f(\cdot)$ qui associe à l'état de la nature s (événement élémentaire), quel qu'il soit, la conséquence $f(s)$ (qui peut être monétaire ou autre). Toutefois en adoptant la méthode opérationnelle de Ramsey, Davidson commence, comme dans l'axiomatique de 1956 présentée ci-dessus, par réduire l'ensemble des états de la nature aux évènements élémentaires E et $\bar{E}$.
- La relation de préférence sur les conséquences : il ne reste alors plus que deux conséquences, les sommes monétaires associées à ces deux évènements.
- la relation sur les événements eux-mêmes, qui est à l'origine des « probabilités qualitatives subjectives ».

L'idée est ainsi de pouvoir attribuer un nombre aux diverses issues envisageables, qui peuvent alors être classées grâce à eux : « on attend, en gros, que les axiomes suffisent à construire une forme de mesure sur des ensembles basiques » (*ibid*). Pour cela il faut trouver un moyen pour dénouer les rôles respectifs de la probabilité subjective et de l'utilité à partir de véritables décisions (Davidson, Suppes, Siegel [1957], p.10).

Davidson et Suppes proposent une solution behavioriste au problème, selon eux, de la mesure séparée de la loi de probabilité subjective et de l'utilité qui s'appuie pour l'essentiel sur un « jeu à une personne simple », dans lequel la personne doit choisir entre deux options, l'une rapportant un gain x_1 si un événement E se réalise et y_1 s'il ne se réalise pas (ce qu'on note $\overline{E}$), l'autre rapportant un gain x_2 si un événement E se réalise et y_2 s'il ne se réalise pas. Ce qui est résumé par le tableau suivant:

	option 1	option 2
E	x_1	x_2
$\overline{E}$	y_1	y_2

Tableau 1

Si on note $u(\cdot)$ la fonction d'utilité d'une personne sur les gains et $p(\cdot)$ sa loi de probabilité subjective, alors la théorie de l'espérance d'utilité revient à comparer les nombres $p(E)u(x_1) + p(\overline{E})u(y_1)$ et $p(E)u(x_2) + p(\overline{E})u(y_2)$.
Davidson et Suppes accordent une place importante au cas particulier où l'option 2 est telle que $y_2 = x_1$ et $x_2 = y_1$, situation décrite dans le tableau 2 :

	option 1	option 2
E	x_1	y_1
$\overline{E}$	y_1	x_1

Tableau 2

En effet, s'il existe alors un événement E^* tel que ces deux options sont considérées comme équivalentes par le sujet, alors on a :

$$p(E^*)u(x_1) + p(\overline{E}^*)u(y_1) = p(E^*)u(y_1) + p(\overline{E}^*)u(x_1),$$

et donc

$$(p(E^*) - p(\overline{E}^*))u(x_1) = (p(E^*) - p(\overline{E}^*))u(y_1).$$

Ce qui implique, si $u(x_1) \neq u(y_1)$, que :

$$p(E^*) = p(\overline{E}^*).$$

La probabilité de l'ensemble E, union de E^* et de son complémentaire, $\overline{E}^*$, étant égale à 1, il s'ensuit que $p(E^*) = p(\overline{E}^*) = ½$.

Davidson et Suppes considèrent alors le cas où les options 1 et 2 du tableau 1 sont équivalentes pour la personne considérée, l'événement E^* étant privilégié (la situation est donc celle décrite dans le tableau 3).

	option 1	option 2
E^*	x_1	x_2
$\overline{E}^*$	y_1	y_2

Tableau 3

On doit alors avoir (puisque les options 1 et 2 sont équivalentes), si la fonction $u(\cdot)$ existe :

$$p(E^*)u(x_1) + p(\overline{E}^*)u(y_1) = p(E^*)u(x_2) + p(\overline{E}^*)u(y_2),$$

soit, comme $p(E^*) = p(\overline{E}^*)$:

$$u(x_1) + u(y_1) = u(x_2) + u(y_2),$$

et donc :

$$u(x_1) - u(x_2) = u(y_2) - u(y_1).$$

Ainsi, en répétant ce jeu plusieurs fois et en utilisant l'événement particulier E^* on peut mesurer les intervalles d'utilité séparant un nombre fini d'issues.

D'où l'importance que Davidson et Suppes accordent à ce cas. Ils vont alors définir la relation quaternaire, notée $\approx E *$ (Davidson, Suppes, Siegel [1957], p.31), qui s'interprète comme une relation d'équivalence, la notation :

$$x_1\, y_1 \approx E *\ x_2\, y_2$$

signifiant que le sujet de l'expérience est indifférent entre l'option 1 et l'option 2 du tableau 1, pourvu que l'événement soit E^*.

Remarquons que l'on a toujours, du moins si $u(\cdot)$ et E^* existent :

H1 $$x_1\, y_1 \approx E *\ y_1\, x_1$$

puisque $p(E^*) = p(\overline{E}^*)$ et puisque :

$u(x_1) + u(y_1) = u(y_1) + u(x_1)$.

Davidson et Suppes appellent H1 l'équivalence $x_1\ y_1 \approx E*\ y_1\ x_1$ que doit vérifier toute fonction d'utilité u(·) d'une personne rationnelle. La détermination de E* leur permet d'éviter le recours direct à la relation de préférence qui définit la « probabilité personnelle qualitative » chez Savage – tout au moins dans le cadre expérimental qu'ils se fixent, et qui consiste à tester la cohérence des choix, en relation avec la fonction d'utilité dont l'existence est supposée.

3.4. Des axiomes aux expériences

Une fois le cadre théorique présenté, il s'agit d'introduire une série d'axiomes et d'hypothèses formulés de telle sorte à ce que la théorie soit testable expérimentalement. Les hypothèses sont relatives à la fois aux utilités (3.4.1), conçues comme également espacées sur une échelle de mesure, ainsi qu'aux probabilités, déterminées une fois que l'événement E* aussi probable que sa négation est trouvé (3.4.2). Le cadre théorique, la construction d'hypothèses ainsi que les expériences participent d'une même cohérence, et sont donc nécessairement corrélés les uns aux autres.

3.4.1. Hypothèses et axiomes d'une structure d'utilité également espacée

Davidson et Suppes s'intéressent à ce qu'ils appellent une structure d'utilité également espacée, qui est formée par l'ensemble des issues envisagées, une relation de préférence binaire *P* sur ses éléments et une relation d'équivalence quaternaire $\approx E*$ telle qu'elle a été caractérisée ci-dessus, et qui remplace donc les probabilités subjectives (celles-ci pouvant être éventuellement calculées, en faisant appel à des hypothèses supplémentaires, non indispensables pour les expériences faites).
Ils définissent sur cette « structure » une série d'axiomes (six, pour être précis), qui sont soit triviaux - selon les auteurs - (symétrie, transitivité), soit découlent du principe de domination stochastique (pour que l'option 1 soit considérée comme équivalente à l'option 2, il faut qu'il y ait un événement de l'option 1 qui soit

préférée au correspondant de l'option 2, et un autre événement où c'est le contraire qui arrive) et qui se traduit, par exemple, par l'axiome (Davidson, Suppes, Siegel [1957], p.31) :

$$\text{si } x_1\, y_1 \approx E * x_2\, y_2 \text{ et si } x_1 P\, x_2, \text{ alors } y_2 P\, y_1$$

Un cas particulier dont Davidson et Suppes se servent beaucoup est celui où les issues sont des quantités de monnaie et où on a :

$$x_1\, y_1 \approx E * z\, z$$

Il résulte alors de l'hypothèse triviale que l'on préfère plus d'argent que moins, que z est forcément compris entre x_1 et y_1 (sinon, il y aurait domination stochastique).

Ces axiomes suffisent à Davidson et Suppes pour démontrer qu'il existe une fonction $u(\cdot)$ telle que :

- $x\, P\, y$ si et seulement $u(x) \geq u(y)$;
- $x_1\, y_1 \approx E * x_2\, y_2$ si et seulement si $u(x_1) + u(y_1) = u(x_2) + u(y_2)$;
- $u(\cdot)$ n'est définie qu'à une transformation affine (à coefficient strictement positif) près. Il suffit toutefois de se donner deux valeurs de référence (une échelle) pour disposer d'une mesure unique.

La fonction $u(\cdot)$ ressemble à la fonction d'utilité de vNM, sauf qu'elle se restreint à la comparaison d'issues équiprobables, de sorte qu'il n'y a pas besoin de faire apparaître des probabilités. En fait, l'égalité

$u(x_1) + u(y_1) = u(x_2) + u(y_2)$

peut se lire

$½\, u(x_1) + ½\, u(y_1) = ½u(x_2) + ½u(y_2)$,

puisque l'événement E^* est, par définition, tel que $p(E^*) = p(\overline{E}\ ^*)$.

L'égalité $u(x_1) + u(y_1) = u(x_2) + u(y_2)$ pouvant se mettre sous la forme :

(1) $$u(x_1) - u(x_2) = u(y_2) - u(y_1),$$

cela explique que Davidson et Suppes parlent de structure d'utilité également espacée, l'attention étant portée sur des couples d'issues (quantités de monnaie) auxquels correspondent des intervalles d'utilité égaux, lorsqu'on passe d'un couple à l'autre.

Le théorème de Davidson et Suppes va donc moins loin que le résultat de Savage concernant le passage des « probabilités personnelles qualitatives » à une loi de

probabilité « quantitative », et personnelle, puisqu'il se cantonne au cas d'options – les actions de Savage – considérées comme équivalentes, et donc aux seuls « intervalles d'utilité », sans qu'il leur soit affecté de probabilité (celles-ci n'apparaissent pas dans la formulation de base du résultat de Davidson et Suppes.). Comme chez Savage, mais non chez vNM, il est toutefois fait appel à deux relations de préférence (ou d'équivalence), l'une portant sur les issues « objectives » du jeu (les quantités de monnaie), l'autre sur l'appréciation concernant les aléas auxquelles elles sont soumises. L'approche est moins ambitieuse que dans le cas envisagé par Savage, mais elle a l'avantage de pouvoir être testée empiriquement.

On ne rentrera pas ici dans le détail de la démonstration d'existence, et de la forme particulière qu'elle peut prendre dans le cas où les issues sont des quantités de monnaie.

En raison du caractère expérimental de leur approche – tester la validité de la théorie – Davidson et Suppes s'intéressent tout particulièrement au cas où l'ensemble *K* présente certaines particularités, notamment lorsqu'il est formé par des quantités finies de monnaie. Ensemble discret, du moins en pratique, puisqu'il y a une unité monétaire (ici le centime). Davidson et Suppes vont donc envisager trois « hypothèses » telles que, si elles sont vérifiées, les axiomes le sont ici.

La première est, selon eux, « triviale » : le sujet préfère plus (de monnaie) à moins :

H0 : x_1 P y_1 si et seulement si $x_1 > y_1$

La seconde l'est beaucoup moins – et elle donnera lieu à des tests ; c'est la relation d'équivalence dont on a déjà parlé plus haut :

H1 $$x_1\, y_1 \approx E*\ y_1\, x_1$$

et qui peut s'interpréter en disant qu'il existe un événement *E** dont la probabilité subjective est égale à celle de son complémentaire ($p(E^*) = p(\overline{E}^*)$).

L'hypothèse suivante est plus compliquée, du moins en ce qui concerne son énoncé, qui comporte une série d'équivalences (13 exactement) qui apparaissent comme raisonnables. L'ensemble *K* est dans son cas un intervalle dont les éléments sont des sommes de monnaie, et qui est construit à partir de deux d'entre elles, *a* et *b*, avec *a*

$< b$, choisies à sa guise par l'expérimentateur et à partir desquelles d'autres sont construites en utilisant la relation d'équivalence $\approx E*$. Cette hypothèse, appelée H2 par Davidson et Suppes, consiste ainsi à supposer qu'il existe une somme de monnaie unique c telle que :

$$b, c \approx E* a, a,$$

ce qui implique que c est inférieur à a (puisque b est supérieur à a ; sinon l'option (b,c) serait strictement préférée à l'option (a,a), et il n'y aurait pas équivalence). En outre, si $u(\cdot)$ existe, il découle alors de la façon dont $E*$ a été définie (égalité (1) ci-dessus) que l'on a :

(2) $$u(b) - u(a) = u(a) - u(c),$$

ce qui signifie qu'aux couples de sommes monétaires (a,b) et (c,a) correspondent des intervalles d'utilité égaux.

Puisque c existe de façon unique, l'hypothèse H2 consiste à admettre aussi l'existence d'une somme d, définie par l'équivalence :

$$b, a \approx E* d, c,$$

qui implique que d soit supérieur à b (puisque $a > c$) et que :

$$u(b) - u(d) = u(c) - u(a).$$

Il résulte en outre de cette égalité et de (2) que :

$$u(b) - u(d) = u(a) - u(b),$$

et donc que le couple (b,d) donne lieu à un intervalle d'utilité égal à celui qui est associé au couple (a,b) (et donc au couple (c,a)).

De même, f et g sont définis, toujours selon l'hypothèse H2, par les équivalences :

$$d, f \approx E* a, a \qquad \text{et} \qquad g, c \approx E* b, b,$$

de sorte qu'on construit ainsi, de proche en proche, un intervalle dont les extrémités sont f et g, et tel que :

H2 $$f < c < a < b < d < g.$$

et tel que :

$$u(f) - u(c) = u(a) - u(b) \text{ ou } u(g) - u(d) = u(b) - u(a)$$

où d est l'élément « juste inférieur » à g, dans la liste des sommes envisagées. Pour simplifier, on appellera H2 la liste de sommes allant de f à g, et qui sont telles qu'elles donnent lieu à des intervalles égaux entre eux, pourvu qu'on prenne deux

éléments contigus dans cette liste. Cette liste est décrite par les treize équivalences suivantes :

i) $b, c \approx a, a$

ii) $b, a \approx d, c$

iii) $d, f \approx b, c$

iv) $b, f \approx a, c$

v) $d, f \approx a, a$

vi) $a, d \approx b, b$

vii) $a, f \approx c, c$

viii) $g, c \approx b, a$

ix) $g, f \approx d, c$

x) $g, c \approx d, a$

xi) $g, a \approx b, a$

xii) $g, c \approx b, b$

xiii) $g, b \approx d, d$

Le choix de K est donc ici très particulier, mais c'est celui qui intéresse Davidson et Suppes, qui vont l'utiliser dans leurs expériences. En effet, ils montrent que si les hypothèses H0, H1 et H2 sont vérifiées, alors il en est de même pour les axiomes qui permettent de dire que la fonction $u(\cdot)$ du théorème donné plus haut, et démontré par Davidson et Suppes, existe. Ces axiomes sont essentiellement ceux décrits dans l'article de Davidson et Suppes [1956] (comme par exemple l'axiome sur la relation J) présenté plus haut, nous ne reviendrons donc pas sur leur présentation.
Il ne le reste donc plus qu'à tester ces 3 hypothèses – en fait H1 et H2 puisque H0 est triviale (toute personne préférant disposer plus d'argent que moins). Pour cela, il est encore nécessaire d'aménager la théorie, après avoir remarqué qu'elle ne peut être testée qu'approximativement.

Mais avant de présenter les différents problèmes liés à l'expérimentation, il nous faut présenter deux autres hypothèses, directement liées à la détermination d'une mesure de la probabilité subjective.

3.4.2. Hypothèses et axiomes d'une structure faible de probabilité subjective

Il reste à décrire les hypothèses relatives à la mesure de la probabilité subjective à l'aide de la fonction d'utilité décrite par intervalles telle que dans H2.
L'hypothèse H3 a trait à l'équation de l'utilité espérée présentée plus haut :

$$p(E)u(x_1) + p(\overline{E})u(y_1) = p(E)u(y_1) + p(\overline{E})u(x_1),$$

Mais cette fois, on va supposer que l'on ne connaît que trois éléments pour tenter de déterminer le quatrième de manière unique.

H3 : Soient trois résultats *x*, *y* et *z* dans K l'ensemble des issues et un événement probable E dans l'ensemble S (avec *p* (E) + *p* (Ē) = 1) alors s'il existe un *w* dans K tel que

$$x, y \approx E \; z, w$$

Alors, *w* est unique.
L'hypothèse H4 impose que la mesure de la probabilité subjective soit indépendante des différents résultats utilisés.

H4 : Soit *u* (.) une fonction d'utilité déterminée par H2. Pour tous les résultats *x*, *y*, *z*, *w*, *x'*, *y'*, *z'* et *w'* dans K et pout tout événement probable E dans S, si $x, y \approx E \; z, w$, et $u(y) \neq u(w)$ et

$$\frac{u(x) - u(z)}{u(y) - u(w)} = \frac{u(x') - u(z')}{u(y') - u(w')}$$

Alors, $x', y' \approx E \; z', w'$.

3.5. Les problèmes liés à l'expérimentation

Davidson et Suppes constatent que dans les expériences menées sur la base de la théorie ci-dessus « la difficulté centrale est de déterminer empiriquement quand la relation $\approx E *$ est vérifiée » (Davidson, Suppes, Siegel [1957], p.40). Il faudrait pour cela que les personnes, auxquelles on présente plusieurs fois les options 1 et 2

choisissent dans (à peu près) la moitié des cas la première – et donc dans (à peu près) la moitié des autres cas la seconde. Or, il n'en est rien : une fois une option choisie parmi les deux, pourtant jugées équivalentes, les sujets des expériences s'y tiennent dans tous les cas où elles leur sont soumises (Davidson, Suppes, Siegel [1957], p 40). Davidson et Suppes expliquent ce comportement par le caractère très simple des options présentées, ce qui distingue leurs expériences de celle de Mosteller et Nogee (1951), qui utilisent une procédure plus compliquée de génération des événements, et qui portent sur des périodes beaucoup plus longues (Davidson, Suppes, Siegel [1957], p 41). A quoi s'ajoutent les variations entre les diverses options, qui sont relativement importantes et poussent à toujours s'en tenir à l'une d'entre elles, bien que d'autres lui soient jugées équivalentes.

Quoiqu'il en soit, après avoir constaté qu'ils sont bloqués dès le départ dans la détermination expérimentale de la probabilité, Davidson et Suppes vont proposer une approche approximative, en définissant une nouvelle relation binaire dont ils disent qu'elle est « analogue » à la relation d'indifférence $\approx E*$, relation qu'ils notent $\preccurlyeq E*$ et qui est telle qui est décrite dans les deux tableaux suivants :

	option 1	option 2
E*	x_1	y_1
$\overline{E}$ *	y_1	$x_1 + 1$

Tableau 4

et

	option 1	option 2
E*	$x_1 - 1$	y_1
$\overline{E}$ *	y_1	x_1

Tableau 5

C'est l'option 2 qui est choisie dans les deux cas. Rappelons que l'existence même d'une fonction d'utilité implique que l'on ait (cf. tableau 2 et la discussion qui le suit) :

$$x_1\, y_1 \approx E*\ y_1\, x_1.$$

Alors qu'ici on a :

$$x_1\, y_1 \preccurlyeq E * y_1\, x_1{+}1 \quad \text{et} \qquad x_1{-}1\, y_1 \preccurlyeq E * \; y_1\, x_1\,,$$

ce qui montre l'analogie avec $\approx E *$, qui est en quelque sorte « encadrée », du fait du centime rajouté dans l'option 2 ou enlevé dans l'option 1 (s'il n'y a pas ce centime – qui n'est toutefois pas négligeable dans les expériences menées, où les sommes dépassent rarement la dizaine de centimes – alors on retrouve la relation $\approx E *$ de l'hypothèse H1).

Davidson et Suppes vont alors montrer comment cette approximation de $\approx E *$ vérifie les axiomes ayant servi à établir l'existence de la fonction $u(\cdot)$ de leur théorème d'existence, même si cela revient à envisager des intervalles, plutôt que des sommes exactes, en ce qui concerne les sommes de monnaie (encadrées par une borne supérieure et une borne inférieure, que les expériences vont déterminer).

En ce qui concerne la relation $\approx E *$ approchée, Davidson et Suppes se servent de tableaux comme les 4 et 5, en prenant pour les couples (x_1, y_1) des valeurs telles que (5, - 5), (17,-10), (10,4), (- 4, -3), (-4, -13), etc. (Davidson, Suppes, Siegel [1957], p.57), comparés à des options où on ajoute 1, puis on enlève 1, à x_1 (dans le tableau 6 ci-dessous, les cas 1 et 5 correspondent aux tableaux 4 et 5 ; idem pour 2 et 13 ; 4 et 10 ; 6 et 9 ; … présentés « dans le désordre » pour que les sujets ne réagissent pas en choisissant la même option deux fois de suite, vu leur ressemblance).

Voyons comment Davidson et Suppes réussirent à trouver une procédure permettant de déterminer la relation $\approx E *$ approchée tout en respectant les treize équivalences exprimées plus haut (Davidson, Suppes, Siegel [1957], p.42).

La première étape consiste à trouver une quantité de monnaie c_l telle que :

(1) $b, c_l \preccurlyeq a, a$ et

(2) $a, a \preccurlyeq b, c_l + 1$

En fait, la quantité c_l correspond à la valeur ou borne « inférieure » de c. En effet, comme au départ, nous ne disposons que des quantités a et b, il s'agit de déterminer les utilités des autres quantités c, d, g, f expérimentalement par le jeu des équivalences. Mais comme la relation $\approx E *$ ne peut être qu'encadrée, il faut trouver les bornes inférieure et supérieure des quantités c, d, g, f pour elles aussi les encadrer.

Dans l'équation (2), la quantité c_l + 1 correspond à la quantité c_l à laquelle s'ajoute 1centime. Lorsque l'on remplace les quantités *a* et *b* par leurs utilités déterminées arbitrairement on obtient :

(3) $1 + u(c_l) \leq -2$ (puisque u (a) = - 1 et u (b) = 1) et

(4) $-2 \leq 1 + u(c_l+1)$

On peut donc dire, selon Davidson et Suppes, qu'il existe deux nombres non négatifs ε_1 et ε_2 tels que

(5) $u(c_l) + \varepsilon_1 = -3$

(6) $u(c_l+1) - \varepsilon_2 = -3$

Comme on le voit, les relations (1) et (2) correspondent à la première équivalence (i) présentée plus haut.

Les auteurs proposent d'appeler c_l + 1, la borne supérieure c_h que pourrait prendre *c*. L'idée étant que *c* se situe entre c_l et c_h.

A partir de ces éléments, on peut chercher les bornes inférieure et supérieure de la quantité *d*, en utilisant c_l et c_h pour déterminer d_h et d_l. L'idée étant de déterminer deux relations qui encadrent cette fois l'équivalence (2).

En utilisant c_l on doit trouver une quantité d_h telle que :

(7) $b, a \preccurlyeq d_h, c_l$

(8) $d_h - 1, c_l \preccurlyeq b, a$

De la même manière, on peut utiliser cette fois c_h pour déterminer la quantité d_l telle que

(9) $d_l, c_h \preccurlyeq b, a$

(10) $b, a \preccurlyeq d_l + 1, c_h$ où $d_l + 1$ correspond à d_h

En utilisant les équations (4) et (5) il est facile de voir que

(11) $u(d_h) \geq 3$ et $u(d_l) \leq 3$

En effet, comme $u(c_l) \leq -3$ en vertu de (3), il est possible de déduire que $u(d_h) \geq 3$ puisque lorsque l'on remplace les quantités de monnaie par les utilités dans (7) on obtient :

$$1-1 \leq u(d_h) + u(c_l)$$

D'où $\quad -u(c_l) \leq u(d_h)$

De là on tire, en multipliant des deux côtés par -1 dans l'inégalité (3) : $-u(c_l) \leq 3$

Donc $u(d_h) \geq 3$.

De la même manière qu'avec *c*, on peut donc dire, selon Davidson et Suppes, qu'il existe deux nombres non négatifs δ_1 et δ_2 tels que :

(12) $u(d_l) + \delta_1 = 3$ et $u(d_h) - \delta_2 = 3$

Il s'ensuite que

(13) $\varepsilon_1 \leq \delta_1$ et $\varepsilon_2 \leq \delta_2$ et que nous sommes parvenus à encadre la relation d'équivalence (ii).

Comme on peut s'y attendre, la précision de l'encadrement des valeurs s'amenuise à mesure que nous utilisons les bornes inférieure et supérieure des différentes valeurs pour déterminer celles d'une nouvelle valeur. La précision pour *d* est par exemple plus faible que celle de *c* comme le mentionnent les auteurs (Davidson, Suppes, Siegel [1957], p. 44).

L'étape suivante – une fois la détermination de *c*, *d* et *f* achevée – consiste à vérifier les valeurs trouvées avec des nouvelles équivalences. Par exemple, on peut utiliser les valeurs c_l et c_h pour vérifier les autres équivalences comme (iv) $b, f \approx a, c$. Cette procédure permettra de trouver une nouvelle paire c_l' et c_h' qui, du fait de la perte de précision évoquée, sera moins « nette » que la première :

(14) $c_l' \leq c_l \leq c_h \leq c_h'$

Autrement dit, comme l'expriment les auteurs, la première paire (c_l, c_h) sera « nichée » (*nest*) entre les valeurs (c_l', c_h').

Si cette (14) n'est pas vérifiée, alors il faudra reprendre le processus depuis le départ et obtenir de nouvelles bornes pour *c*, *d*, *f* et *g* puis revérifier.

Ce processus consistant à déterminer des nouvelles paires pour toutes les bornes de toutes les valeurs peut être répété indéfiniment.

Quoiqu'il en soit, l'important selon les auteurs, est d'avoir une représentation approchée d'une mesure de l'utilité par intervalles qui soit cohérente avec les hypothèses et les axiomes qui en découlent.

Enfin, les auteurs précisent que les équivalences (v), (vi), (vii), (xii) et (xiii) n'ont pas été utilisées pour déterminer les valeurs approchées car ces équivalences impliquaient des comparaisons des paris risqués et de valeurs sûres (*a* si E*, *a* si $\overline{E}$ ***c'est-à-dire *a* pour sûr). Or, de telles comparaisons constituaient comme on l'a vu l'une des critiques de Davidson, Suppes et Siegel à Mosteller et Nogee.

Au final, les auteurs sont parvenus à trouver expérimentalement – en partant de deux valeurs posées arbitrairement *u* (*a*) et *u* (*b*) – quatre quantités de monnaie *c*, *d*, *f* et *g* dont les utilités pour *u* (*c*), *u* (*d*) ,*u* (*f*) et *u* (*g*) sont également espacées sur une échelle d'intervalles et telles que *u* (*f*) = -5, *u* (*c*) = -3, *u* (*d*) = 3 et *u* (*g*) = 5 (Davidson, Suppes, Siegel [1957], p.26).

Parallèlement au processus d'encadrement des quantités monétaires et des utilités, il est possible d'utiliser une méthode similaire pour encadre la probabilité subjective (Davidson, Suppes, Siegel [1957], p.47).

Les auteurs proposent une « théorie de l'approximation utilisée pour la mesure de la probabilité subjective d'un événement probable E' dont la probabilité objective est ¼ » (*ibid*.).

Le point de départ consiste à supposer l'hypothèse usuelle concernant les probabilités :

p (E) + *p* (Ē) = 1. On considère en outre les valeurs *a*, *b*, *c*, *d*, *f*, et *g* également espacées sur une échelle d'utilité comme on l'a montré plus haut ainsi que leurs bornes inférieure et supérieure.

L'idée des auteurs est que si une mesure parfaite est possible et que la probabilité subjective qu'un sujet accorde à l'événement E' est ¼, alors les relations suivantes sont valides :

(15) *d*, *a* ≈ (E') *c*, *b*

(16) *g*, *c* ≈ (E') *a*, *a*

(17) *f*, *d* ≈ (E') *b*, *b*.

Mais comme nous ne pouvons avoir accès à la relation d'équivalence – comme on l'a expliqué plus haut – nous devons comme pour les utilités faire une approximation de cette relation.

A partir de (15) il est possible de trouver une quantité *a'* telle que :

(18) d_h , *a'* ≼ (E') c_l , b_l et c_l , b_l ≼ (E') d_h , *a'*+1

En transcrivant cette écrire sous la forme de l'utilité espérée on obtient :

(19) *p* (E') *u* (d_h) + [1 - *p* (E')] *u* (*a'*) ≤ *p* (E') *u* (c_l) + [1 - *p* (E')] *u* (b_l)

A partir des valeurs trouvées pour *u* (d_h), *u* (c_l) et *u* (b_l) (*u* (b_l) ≤ 1 puisque arbitrairement *u* (*b*) = 1), on peut en déduire que :

(20) p (E') $\leq \frac{1\text{-u (a')}}{7\text{-u (a')}}$

Symétriquement, en utilisant

(21) $c_h, b_h \preccurlyeq$ (E') d_l , a'' et d_l , a''- 1 $\preccurlyeq$ (E') c_h, b_h

On en déduit :

(22) $\frac{1\text{-u (a'')}}{7\text{- u (a'')}} \leq p$ (E')

D'où l'on tire :

(23) $\frac{1\text{-u (a'')}}{7\text{- u (a'')}} \leq p$ (E') $\leq \frac{1\text{- u (a')}}{7\text{- u (a')}}$

Il ne reste qu'à déterminer les valeurs de u (a') et de u (a'') dont on sait déjà quelles sont fonction de a (les auteurs évoquent la possibilité de procéder à des interpolations linéaires (Davidson, Suppes, Siegel [1957], p. 48).

Il reste à présent à détailler le protocole expérimental (constitué d'une étude pilote et d'un protocole final) pour présenter en suite les résultats de ces expériences.

3.6. Protocole expérimental

Les sujets étaient dix neuf sujets hommes tirés au sort au sein du service de l'emploi étudiant de l'Université de Stanford et testés individuellement, sans se connaître les uns les autres. Les sujets étaient choisis parmi ceux qui répondaient à une proposition du service emploi de travailler à des tâches non attractives comme tondre la pelouse ou faire du travail de bureau pour 1$ de l'heure (Davidson, Suppes, Siegel [1957], pp.49-50). Il était annoncé à ceux qui étaient tirés qu'ils allaient agrafer des papiers de telle sorte à ce qu'il n'y ait aucune manière, pour le sujet, d'être volontaire ou de choisir d'être le sujet de l'expérience.

La première session avec chaque sujet dura deux heures et à la fin de la première session, les expérimentateurs prenaient les coordonnées du sujet en lui disant qu'il serait rappelé s'il le voulait. La plupart des sujets étaient rappelés pour une deuxième et troisième sessions, les sessions étant espacées entre plusieurs jours et un mois.

Quand une personne tirée au sort venait pour travailler, elle était informée qu'un des sujets pour une expérience avait annulé son rendez-vous et qu'elle pouvait, si elle le désirait, servir de sujet expérimental au lieu de faire le travail qu'elle devait faire à l'origine. Le seul désagrément qu'impliquait cette nouvelle occupation est que le sujet potentiel aurait à parier avec son salaire et pourrait tout perdre, ce qui était notifié à la personne considérée. On lui disait aussi que si elle perdait tout son salaire avant que le temps s'achève, elle aurait à travailler le reste du temps (l'expérience durait deux heures) sans être payer. Pour rassurer les sujets potentiels, les expérimentateurs indiquaient que sur la base des expériences passées, les chances du sujet paraissent bonnes.

Si l'individu était intéressé, le jeu lui était expliqué et il devenait un sujet. S'il refusait, il agrafait pour 1$ de l'heure. Seulement une seule personne avait refusé d'être un sujet. Sans connaître les sujets, une promesse était faite au service de l'emploi qu'aucun de sujet ne devrait atteindre une moyenne de gain inférieur à 2$ pour une session de deux heures.

Quand quelqu'un acceptait de devenir un sujet, les expérimentateurs lui précisaient que l'expérience était faite pour tester le comportement humain dans des situations de paris et que toutes ses réponses du sujet allaient être également valorisées par les expérimentateurs. De même, ces derniers assuraient aux sujets qu'aucun comportement particulier n'était recherché ou attendu d'eux et qu'en aucun cas il s'agissait d'un test d'intelligence. Enfin, les auteurs expliquaient aux sujets que les jeux auxquels ils allaient jouer dureraient deux heures et que le taux de paye serait d'1$ par heure et que le paiement des 2$ s'effectuerait au début de la session et que cet argent servirait à participer aux jeux.

En général, d'après les auteurs, le jeu se déroulait de la façon suivante : le sujet était assis en face du « croupier » (l'expérimentateur), celui-ci lui présentait verbalement les options parmi lesquelles il devait choisir. On demandait toujours au sujet de choisir une option parmi les deux qui lui étaient présentées. Pendant quelques sessions, une tierce personne était présente pour enregistrer le temps pris pour prendre chaque décision. On recommandait au sujet de réfléchir à chaque choix autant de temps qu'il le voulait.

La première étape de l'expérience nécessitait que, pour chaque sujet, un événement E* soit trouver de telle sorte que :

$x, y \preccurlyeq (E^*)\ y + 1¢, x$ et $x, y - 1¢ \preccurlyeq (E^*)\ y, x$

où *x* et *y* sont considérés comme plusieurs pairs de conséquences différentes. En réalité, un seul événement probable était trouvé durant les études pilotes qui satisfaisait toutes les conditions pour chaque sujet.

Cet événement probable n'était pas pour autant facile à trouver. Une pièce était tirée et on donna au sujet l'opportunité de parier sur face (E) ou pile ($\overline{E}$), un dé fut utilisé avec des nombres pairs pour E et des nombres impairs pour $\overline{E}$, puis deux pièces furent tirées…Dans chaque cas, la plupart des sujets montrait une préférence pour E ou $\overline{E}$. Finalement, un événement fut trouvé qui satisfaisait les conditions. Cet événement fut produit au moyen d'un dé spécialement crée. Sur trois faces du dé étaient inscrits la syllabe dénuée de sens « ZOJ » et sur les trois autres, la syllabe « ZEJ ». Deux autres dés furent construits avec « WUH » et « XEQ », « QUG » et « QUJ » à la place de numéros. Ces syllabes sont celles qui, selon Glaze [1928] et d'autres n'ont pratiquement aucune valeur associative. L'espoir était que ces sujets n'aient aucun préjugé en faveur de l'une ou l'autre de ces syllabes. Tous ces dés furent testés avec chacun des sujets, et dans chaque cas, l'hypothèse H1 tenue (à l'intérieur des limites de 1¢ avec la méthode décrite), en considérant comme l'événement E* le jet d'un dé de telle sorte que ce soit la face « ZOJ » (ou « WUH » ou « QUG ») qui vint et $\overline{E}$ *le jet du dé de telle sorte à ce que ce ne soit pas « ZOJ » (ou « WUH » ou « QUG ») qui survint. Notons qu'en expliquant le jeu au sujet, $\overline{E}$ * était indiqué positivement en termes de syllabes restantes du dé).

Avant que le jeu commence, voici comment le jeu fut expliqué au sujet :

> « Le jeu auquel nous voulons jouer va prendre la forme suivante. Vous allez secouer ce dé dans le gobelet prévu à cet effet et vous allez le lancer sur la table. Comme vous pouvez le voir, il y a différentes syllabes sur le dé. Trois faces du dé ont ZOJ sur leur face, et sur les trois autres côtés, il y a la syllabe ZEJ. Parfois, nous jouerons avec ce dé et parfois avec un autre dé comme celui-ci mais avec d'autres syllabes ; à savoir WUH et XEQ ou QUG et QUJ.

Ces dés ont été faits spécialement pour nous et ils sont parfaitement équitables et non pipés.

Avant de secouer le dé, je vais vous présenter deux issues et vous devrez choisir entre l'une d'elles. Vous avez ici un crayon et un papier que vous pouvez utiliser. Prenez tout le temps que vous voulez pour faire votre choix.

Ici par exemple, c'est le dé ZOJ et ZEJ. Si vous voulez parier sur ZOJ, vous gagnerez cinq ¢ si vous avez raison (c'est-à-dire si ZOJ survient quand vous allez lancer le dé) et vous perdez cinq cents si vous vous trompez (c'est-à-dire si ZEJ survient). Si vous voulez parier sur ZEJ, vous gagnez sur six cents si vous avez raison, et vous perdez cinq cents si vous avez tort. Quel est votre choix ? [le sujet choisit]OK.

Avant de commencer, je voudrais vous suggérer que vous notiez chaque offre sous cette forme :

ZOJ	ZEJ
+5	+6
-5	-5

Cela signifie : si vous pariez sur ZOJ, vous allez gagner cinq cents si cela arrive et perdre cinq cents si c'est le contraire. Si vous pariez sur ZEJ, vous gagner six cents si cela arrive et perdre cinq cents si c'est le contraire. Etes-vous prêt à continuer ? » (Davidson, Suppes, Siegel [1957], p. 53).

Face à cette façon de procéder, l'étude pilote a révélé bon nombre de difficultés, les solutions à celles-ci furent incorporées dans le schéma final de l'expérimentation. Nous insistons sur ces difficultés que les auteurs rencontrent (3.6.1) et qui les conduisent à mettre en place un protocole final (3.6.2), puis présentons les résultats de l'expérience (3.6.3).

3.6.1 Difficultés rencontrées dans l'étude pilote

(a) Contrôle des « effets de renforcement cumulatifs et directs ».

Les auteurs se sont aperçus que gagner ou perdre plusieurs fois de suite par session rendait les sujets optimistes ou pessimistes, ce qui en retour pouvait avoir des effets sur les réponses des sujets à des offres similaires. De la même manière plus l'enjeu monétaire devenait important ou symétriquement plus il devenait faible, plus des effets de distorsions observés sur les choix pouvaient survenir (Davidson, Suppes, Siegel [1957], p. 53).

La difficulté était non seulement de repérer ces effets sur les sujets mais aussi d'éviter que ces effets ne contaminent durablement tous leurs choix.

Pour pallier cette difficulté, les auteurs mirent en place une procédure particulière.

Lorsque par exemple il s'agissait de tester l'hypothèse H1, le croupier-expérimentateur offrait près de 15 paris au sujet. Au départ, chaque pari pris par le sujet était suivi d'un jet de dé et du règlement de la récompense (par l'expérimentateur) - ou de la dette (par le sujet) - relative au pari lui-même. Pour contrôler les effets de renforcement, il était proposé au sujet, de faire plusieurs paris à la suite avant que le dé ne soit lancé – tout ceci avec son accord. La justification donnée au sujet était qu'il s'agissait pour lui de gagner du temps. C'est-à-dire qu'une paire d'issues était présentée à lui et le sujet faisait son choix, qui était noté, après quoi une autre paire lui était immédiatement présentée. On disait au sujet que les choix qu'il faisait étaient liés et que lorsqu'il faisait ses choix pour deux ou trois paires d'issues successivement, il ferait rouler le dé une fois pour chacun d'eux, après quoi il ne pourrait plus changer d'avis. En d'autres termes, le sujet était engagé dans le choix qu'il faisait quand les issues lui étaient proposées, même s'il ne faisait pas rouler le dé pour voir si un choix particulier était payant jusqu'à ce que tous ses choix soient faits. L'expérimentateur donna au sujet des paires d'issues en groupes de trois ou quatre jusqu'à ce que les épreuves pour tester H1 soient terminées.

Une fois cette mesure prise, l'expérimentateur était capable de mesurer l'utilité du sujet (H2).

A ce point, l'expérimentateur disait au sujet qu'il y avait seulement 25 ou 30 choix à faire encore, et il lui demandait s'il voulait les faire tous avant que le dé ne soit jeté pour voir quelle serait le résultat final[86].

Ayant recueilli l'accord de tous les sujets, aucune impression de retard dans la récompense ne semblait être ressenti par les sujets si bien que ces derniers avaient toujours l'impression de parier même si l'issue des paris n'était connue que de manière décalée (Davidson, Suppes, Siegel [1957], p. 54).

(b) L'effet de récence (*recency*)

Un deuxième problème expérimental concerne l'effet de récence (*recency*). L'expression la plus simple de cet effet est que les sujets avaient plus de facilité à se rappeler des dernières syllabes du dé lors des derniers jets plutôt que des syllabes obtenus dans les tous premiers.

Cet effet se matérialisait, selon les auteurs, lorsque les sujets voyaient une syllabe particulière d'un dé arriver (sortir) ou non suite à de nombreux jets. Si la même syllabe venait plus de trois fois de suite par exemple, la probabilité subjective baissait temporairement pour la plupart des sujets.

Le problème fut effectivement résolu en utilisant trois dés au lieu d'un. Le dé utilisé était changé après chaque jet. De cette manière, aucune syllabe particulière ne pouvait gagner ou perdre, ou même être choisie, deux fois de suite. Le temps que le dé soit réutilisé, le sujet avait oublié ou n'était pas influencé par son expérience précédente (Davidson, Suppes, Siegel [1957], p.55). L'expérimentateur n'annonçait pas les événements correspondants avant que le dé soit jeté.

Cet effet de récence rappelle les effets observés par Preston et Baratta [1948] évoqués plus haut. La particularité du modèle de 1957 de Davidson et *al* est que les auteurs ont tenté de corriger ces effets pour qu'ils ne puissent pas amoindrir ou remettre en cause les résultats de l'expérience.

[86] Les auteurs mentionnent qu'en fait, « il y avait le plus souvent plus de deux fois plus de paris restant mais aucun sujet ne l'avait noté » (Davidson, Suppes, Siegel [1957], p. 54).

(c) Distorsions dans la détermination des quantités de monnaie

Le troisième problème est relatif aux distorsions initialement rencontrées lors de la détermination expérimentale de la quantité de monnaie *c* comme cela a été mentionné plus haut. Ce problème, de l'avis des auteurs, ne reçut qu'une solution partielle (Davidson, Suppes, Siegel [1957], p.55).

Lors du test de H2, les auteurs ont procédé ainsi : partant de $a = -4¢$ et $b = 6¢$, arbitrairement choisis, un montant *c* était trouvé tel que $b, c \approx a, a$ (ou son approximation) ; sur la base de *a*, *b et c*, un autre montant est trouvé et ainsi de suite.

Dès lors, la valeur de chaque point sur l'échelle d'utilité dépend des points trouvés auparavant.

Une distorsion dans la valeur de *c* est donc cruciale, puisque cela va distordre toutes les autres valeurs. La distorsion initiale de *c* se répercute car elle constitue le point de départ de l'approximation des autres valeurs. Un problème essentiel réside aussi dans la relation (i), la première des treize équivalences, $b, c \approx a, a$. En effet, si le sujet choisit l'option sûre (a, a) [87] il est nécessairement perdant puisque l'issue consiste par hypothèse à perdre 4 centimes. Comme le mentionnent les auteurs : « Si le sujet a une utilité positive pour le pari, on devra lui offrir un *c* faussement bas avant qu'il reporte son choix sur l'option (a, a) ; cela va en retour rendre la valeur de *d* faussement haute et ainsi de suite [88] » (1957, p. 55).

La méthode employée pour compenser une distorsion possible se présente comme suit[89] : le point de départ est une distorsion qui existe si

- d'une part, on sait que à partir de *a* et *b* données par la relation (i), on trouve un *c*, puis avec (ii) un montant *d* est trouvé ; en utilisant *d* et (iii), *f* est trouvé.
- Et d'autre part, si grâce à la relation (iv) et les mêmes *f, b, d,* on trouve cette fois une nouvelle valeur *c'* telle que $b, f \approx a, c'$.

---

[87] Les auteurs reconnaissent donc de leurs propres aveux qu'ils n'ont pas pu éviter de comparer des issues sûres à des paris risqués comme ils l'avaient avancé au départ (Davidson, Suppes, Siegel [1957], p. 55).

[88] D'ailleurs, seule la relation (i) a été considérée par les auteurs puis le problème a été volontairement négligé pour les autres équivalences (Davidson, Suppes, Siegel [1957], p. 55).

[89] Il est à noter que seul le cas de la mesure parfaite est traité, mais pas celui de la mesure approchée (*ibid.*).

Si *c'= c*, aucune distorsion due à l'utilité du pari n'est entrée apparemment dans la détermination originale de *c* (les auteurs supposaient au cours de cette discussion qu'aucune distorsion due à l'utilité du pari n'entrait dans les choix entre deux options dont chacune est un pari).

Si *c' < c*, on peut raisonnablement penser que le *c* original est trop bas. L'expérimentateur augmentait alors le *c* original d'1¢, puis déterminait un nouveau *d* et un nouveau *f* (appelés *c+1¢*, *d'* et *f'*) et vérifiait si *b, f'≈ a, c+1¢* ; s'il en est ainsi, *c+1¢* est la valeur correcte. Sinon, un autre ajustement était indiqué. Une procédure similaire était suivie si *c < c'*.

Pour 1/3 de nos sujets, les auteurs ont eu besoin de faire des compensations. Mais pour seulement un seul d'entre eux il était nécessaire de faire des compensations à deux reprises. Davidson, Suppes et Siegel pensaient qu'il était possible d'interpréter ces résultats comme montrant que la distorsion due à l'utilité du pari n'était pas aussi forte ou aussi fréquente que ce qui avait été présumé, au moins pour de faibles montants de monnaie.

(d) Rapidité et simplicité dans le recueil des données

Le dernier problème mentionné par les auteurs est qu'ils se sont sentis obligés d'être prompts à obtenir les données relevant de la mesure de l'utilité d'une seule session puisqu'ils ne pouvaient pas être sûrs de la stabilité de la fonction d'utilité du sujet sur des périodes plus longue que celle de la session de test. Les auteurs ont donc considéré que la rapidité et la simplicité valaient quelque sacrifice en termes de complétude de certains aspects de l'expérience, au moins jusqu'à ce qu'il y ait une preuve claire que la fonction d'utilité était relativement stable au cours du temps. Dans la conception finale, les auteurs étaient capables de rassembler l'information nécessaire pour H2 en moins d'une heure en moyenne.

3.6.2. Le protocole expérimental final

Le protocole expérimental final effectivement utilisé par Davidson tente de surmonter les difficultés énoncées au paragraphe précédent.

Les sujets étaient placés dans une situation expérimentale et instruits de leur rôle comme cela a été mentionné.

La première séquence d'offre avait un double objectif : tester l'hypothèse H1 et familiariser le sujet avec le jeu. Le tableau 6 montre une séquence d'ouverture typique.

Numéro de l'offre	Evénements	Option 1	Option 2	Choix correct pour confirmer H1
1	ZOJ ZEJ	5¢ -5	-5¢ 6	Option 2
2	QUG QUJ	17 -10	-11 17	Option 1
3	ZOJ ZEJ	24 -3	13 5	
4	WUH XEQ	10 4	3 10	Option 1
5	ZEJ ZOJ	4 -5	5 -5	Option 2
6	QUG QUJ	-4 -3	-2 -4	Option 2
7	ZOJ ZEJ	13 -2	-3 13	Option 1
8	XEQ WUH	7 -6	-6 8	Option 2
9	QUJ QUG	-2 -4	-4 -1	Option 2
10	WUH XEQ	5 10	10 4	Option 1
11	ZOJ ZEJ	-4 13	13 -3	Option 2
12	XEQ WUH	-3 -5	8 -19	
13	QUJ QUG	-9 17	17 -10	Option 1

14	ZOJ ZEJ	15 -10	22 -19	
15	WUH XEQ	-6 9	8 -6	Option 1

Tableau 6

Dans cette séquence, toutes les offres - sauf 3, 12 et 14[90]- testaient H1, celles-ci étaient insérées uniquement dans le but d'exposer le sujet à des offres du type de celles qu'il trouvera dans une partie ultérieure de l'expérience.

Pour vérifier H1, le sujet n'avait qu'à faire les choix marqués dans la dernière colonne. Comme dans l'étude pilote, on donna des jetons aux sujets d'une valeur de 2$ au début du jeu en leur assurant qu'à la fin de la session, les jetons en leur possession seraient échanger avec de la monnaie.

Ce n'est qu'une fois que H1 avait était testé de manière adéquate, que l'expérimentateur faisait les offres nécessaires pour tester H2 et déterminer (si H2 était vérifiée) la courbe d'utilité du sujet pour les montants de monnaie impliqués.

La séquence indiquée dans la section 3.6.1 ci-dessus était suivie. Les valeurs de base $a = -4¢$ et $b = 6¢$ étaient choisies de telle sorte à ce que la totalité des offres tendent à avoir une valeur espérée proche de 0¢ et ceci afin que le sujet sente que ses pertes et ses gains seraient probablement égales (même si le dé n'était lancé qu'après une session de plusieurs paris). Les montants *a* et *b* étaient légèrement orientés pour rendre la valeur espérée de la totalité des offres légèrement positive (au moins pour la plupart des sujets) et éviter les possibles effets de distorsion qui pourraient résulter d'une symétrie autour de 0¢.

La méthode finalement utilisée pour déterminer expérimentalement les bornes inférieures et supérieures des valeurs monétaires se présentait comme suit :

Supposons que pour un sujet donné nous disposions des bornes inférieure et supérieure de *c* et de *d* ; par exemple $c_l = -11¢$, $c_h = -10¢$, $d_l = 11¢$ et $d_h = 12¢$. Si l'on cherche la valeur de *f* telle que l'équivalence (iii) $d, f \approx b, c$ soit vérifiée, il nous faut

[90] Les réponses à ces offres n'étaient pas directement appropriées pour tester H1 ou H2.

déterminer les bornes inférieure et supérieure de f à partir d'inégalités comme $d_h, f_l \preccurlyeq b, c_l$. En essayant plusieurs valeurs de x dans le tableau suivant :

Option 1	Option 2
12¢	6¢
X	-11¢

Tableau 7

Les auteurs découvrirent que le sujet commençait à préférer l'option 2 à l'option 1 à partir du moment où $x > -18¢$. De là, ils concluaient que $f_l = -18¢$. Cette procédure était suivie pour toutes les autres valeurs (Davidson, Suppes, Siegel [1957], pp. 58-59).

Là encore, les sujets ne devaient pas s'apercevoir que les expérimentateurs cherchaient le point où la préférence basculait d'une option à une autre. Il fallait donc à la fois multiplier les offres pour masquer cette recherche mais aussi compter sur l'habileté (*shrewdness*) de l'expérimentateur.

Il restait alors à mesurer la probabilité subjective avec la méthode décrite plus haut. Pour ce faire, les auteurs introduisaient deux différences intervenaient pas rapport aux précédentes expériences afin de calibrer l'échelle de probabilités de quart en quart.

La première était que le dé utilisé[91] n'avait cette fois-ci que quatre faces, chacune portant une syllabe dénuée de sens : ZEJ, WUH, XEQ, et VAF[92]. L'opportunité était donnée aux sujets de parier sur l'une de ces faces, ou de parier contre celle-ci, c'est-à-dire pour les trois autres restantes ; ce qui constituait la seconde différence avec les premières expériences décrites.

Restaient alors à suivre les étapes décrites au paragraphe précédent.

[91] Celui-ci était toujours équilibré et conçu spécialement pour l'occasion.

[92] On retrouve ici l'objectif de mesurer la probabilité subjective lorsque la probabilité objective est égale à ¼.

3.6.3. Résultats

Selon Davidson, Suppes et Siegel, les dix neuf sujets testés passent l'épreuve de l'hypothèse H1 :

Si une fonction d'utilité $u(\cdot)$ existe, elle doit vérifier les égalités des intervalles d'utilité entre deux éléments consécutifs de *K*. Davidson et Suppes constatent que le comportement de 15 des 19 sujets est « compatible avec l'affirmation selon laquelle il existe une fonction d'utilité $u(\cdot)$ unique à une transformation affine croissante près définie sur les issues de base envisagées » (Davidson, Suppes, Siegel [1957], p.61), les valeurs de *c*, *d*, *f* et *g* vérifiant les inégalités de H2[93].

Dans ces conditions, « si cette fonction d'utilité existe et si on lui donne les valeurs $a = -4$ et $b = 6$, $u(-4) = -1$ et $u(6) = 1$, alors il découle de la théorie (*equally spaced utilities*) que $u(c) = -3$, $u(d) = 3$, $u(f) = -5$, $u(g) = 5$ » (ibid., p. 61). L'écart des utilités entre *a* et *b* étant égal à 2, il en est de même des écarts entre *a* et *c*, entre *c* et *f*, entre *b* et *d* et entre *d* et *g*.

Il est alors possible de construire une courbe qui joigne les points de la forme $(x, u(x))$, où *x* est donnée – outre par *a* et *b* - par les valeurs trouvées expérimentalement *f, c, d, g*, les valeurs de $u(\cdot)$ étant connues en ces points (en supposant la théorie vraie). En fait, cette courbe s'avère « floue », puisqu'on a vu que les valeurs déterminées expérimentalement pour les sommes monétaires ne sont qu'approchées. La courbe d'utilité est ainsi elle-même encadrée par une courbe « basse » et « haute ».

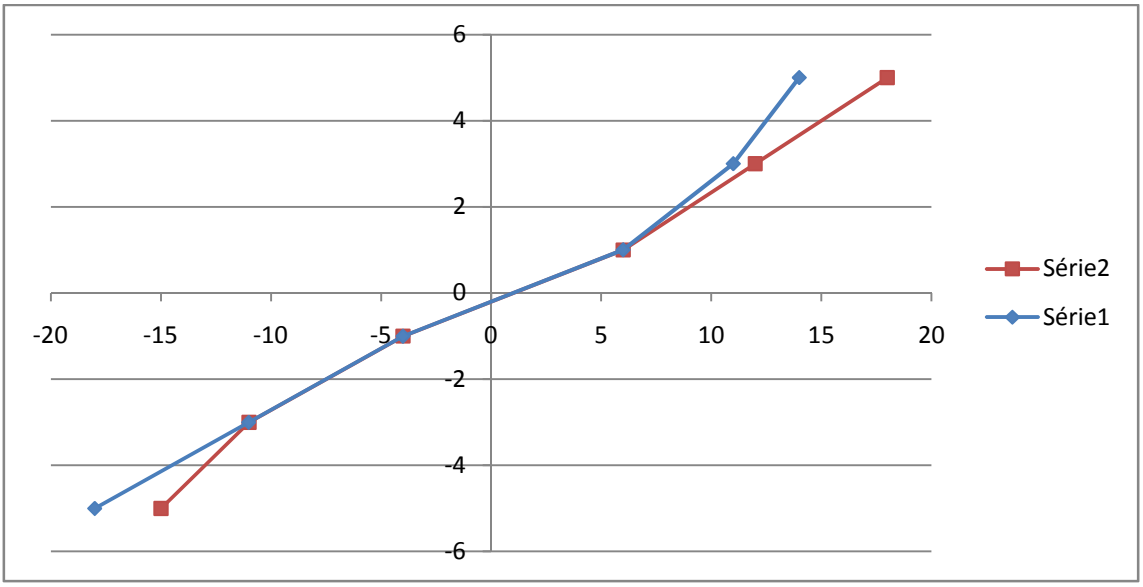
Elles varient évidemment d'un individu à l'autre.

[93] Pour les quatre sujets pour lesquels H2 n'étaient pas vérifiée, Davidson, Suppes et Siegel avaient observé soit une tension nerveuse particulière dans le simple fait de jouer soit des problèmes de compréhension des options dus au fait que certains sujets ne maîtrisaient pas parfaitement l'anglais (Davidson, Suppes, Siegel [1957], p.66).

Un exemple tiré des expériences pour l'un des sujets peut être présenté :

Bornes pour *f* avec *u (f)* = -5	Bornes pour c avec *u (c)* = -3	*u (a) = -1*	*u (b) = 1*	Bornes pour *d* avec *u (d) = 3*	Bornes pour *g* avec *u (g) = 5*
-18 à -15¢	-11 à -10¢	- 4¢	6¢	11 à 12¢	14 à 18¢

Graphiquement, deux courbes d'utilité entre lesquelles la véritable courbe d'utilité prend place peuvent être représentées.



Bornes pour la courbe d'utilité du sujet 1

Ici sont représentées les deux courbes d'utilité inférieure et supérieure qui bornent la courbe d'utilité du sujet 1. La série 1 représente les bornes inférieures et la série 2 les bornes supérieures pour tous les points d'utilité.
Les abscisses représentent les cents alors que les ordonnées représentent les utilités.
La même représentation peut être fournie pour tous les sujets.
Remarquons, comme Davidson, Suppes et Siegel l'indiquent ([1957], p. 72) que cette courbe ressemble à la courbe hypothétique de Friedman et Savage [1948].

Les résultats pour les autres sujets sont regroupés[94] dans le tableau suivant :

[94] Il ne figure dans ce tableau que les mesures pour 15 sujets car pour les 4 autres, cette mesure ne put être achevée (Davidson, Suppes, Siegel [1957], p. 66).

Sujets	Bornes pour *f* avec *u (f)* = -5	Bornes pour c avec *u (c)* = -3	*u (a) = -1* (établi arbitrairement)	*u (b) = 1* (établi arbitrairement)	Bornes pour *d* avec *u (d) = 3*	Bornes pour *g* avec *u (g) = 5*
1	-18 à -15¢	-11 à -10¢	- 4¢	6¢	11 à 12¢	14 à 18¢
2	-34 à -30	-12 à -11	-4	6	12 à 18	31 à 36
3	-18 à -11	-8 à -7	-4	6	10 à 13	14 à 22
4	-29 à -24	-15 à -14	-4	6	14 à 17	25 à 31
5	-21 à -14	-10 à -9	-4	6	10 à 12	16 à 24
6	-25 à -21	-14 à -13	-4	6	13 à 15	19 à 23
7	-18 à -7	-7 à -6	-4	6	7 à 14	10 à 23
8	-25 à -21	-14 à -13	-4	6	14 à 17	23 à 28
9	-35 à -29	-12 à -11	-4	6	16 à 18	43 à 50
10	-26 à -20	-15 à -14	-4	6	14 à 15	20 à 27
11	-22 à -19	-14 à -13	-4	6	11 à 13	18 à 22
12	-21 à -13	-12 à -11	-4	6	8 à 12	11 à 15
13	-34 à -23	-14 à -13	-4	6	13 à 17	23 à 32
14	-16 à -13	-10 à -9	-4	6	12 à 15	20 à 24
15	-12 à -8	-8 à -7	-4	6	8 à 10	11 à 15

Tableau 8

Pour le test de H3 et H4, seuls sept sujets (avec lesquels les auteurs avaient pu mesurer l'utilité) furent choisis. H3 fut vérifiée pour tous ces sujets. En revanche, H4 ne fut vérifiée que pour cinq sujets. Pour ces derniers, les auteurs tirèrent la conclusion que leur comportement était cohérent avec l'hypothèse qu'il existe une

fonction de probabilité subjective qui a les propriétés recherchées[95] (Davidson, Suppes, Siegel [1957], p.68).

Certains sujets furent mêmes testés à nouveau plusieurs semaines après les premières expériences. Pour trois sujets, les auteurs observèrent une reproduction exacte des premiers choix[96]. Pour les autres, soit les sujets se rapprochaient de la courbe d'utilité linéaire en monnaie (graphiquement elle est représentée par la première bissectrice) soit ils s'écartaient substantiellement de leurs choix précédents sans qu'aucune explication de ce phénomène ne puisse être apportée (Davidson, Suppes, Siegel [1957], p. 69).

3.7. Critiques de la théorie de Davidson, Siegel et Suppes (1957)

L'axiomatique, l'expérimentation et les résultats de la théorie de Davidson, Suppes, Siegel offrent de nombreux apports pour la théorie de la décision. Le modèle construit par les auteurs est d'abord un modèle riche qui emprunte à nombres de théoriciens de la décision et qui est en ce sens représente l'un, sinon le plus complets des années 1950. Les auteurs n'ont ainsi pas reculé devant la complexité d'une procédure visant à vérifier l'hypothèse d'utilité espérée de vNM et Friedman et Savage, l'utilisation de la méthode opérationnelle de Ramsey, la loi de probabilités subjectives initiées par Ramsey et Savage. Cette ambition particulièrement audacieuse a toutefois ses revers.

La présentation du modèle et de la procédure expérimentale est très aride. Les auteurs eux-mêmes y voient des limites pesant sur la théorie dans son ensemble (3.7.1). D'autres considèrent que la théorie de l'utilité espérée entière devrait être remise en cause (3.7.2).

---

[95] Voir présentation de l'article de 1956.

[96] Certains sujets furent mêmes testés une troisième fois et les auteurs remarquèrent que les choix étaient les mêmes que pour la seconde fois alors que celle-ci divergeait de la première (Davidson, Suppes, Siegel [1957], p.69)

3.7.1. Des critiques adressées par les auteurs eux-mêmes

Deux problèmes sont soulevés par les auteurs.

Le premier problème réside dans le fait que les issues sont des quantités de monnaie et qu'il semble donc difficile, selon les auteurs, d'étendre les résultats à des issues d'un autre type (Davidson, Suppes, Siegel [1957], p. 78).

Le second problème est que la méthode utilisée suppose que les choix entre les options ne changent pas d'une session à l'autre (Davidson, Suppes, Siegel [1957], p. 79). Les auteurs considèrent sur ce point que la théorie de Mosteller et Nogee est supérieure à leur modèle dans la mesure où elle est fondée sur une approche statistique des préférences. Comme on l'a vu, l'idée de cette approche est que l'on considère l'indifférence comme la situation où le sujet, lorsqu'il est face à deux options, choisit dans la moitié des cas l'une de ces options, et l'autre le reste du temps (*ibid.*).

Si les auteurs soulignent le manque de généralité de l'analyse due l'usage de la monnaie, ils en proposent aussitôt une solution. Pour pallier quelques imperfections du modèle de base faisant usage de quantités de monnaie, Davidson et Suppes proposèrent dans le même ouvrage de 1957 un modèle de programmation linéaire ayant le même objectif que le modèle précédent, à savoir mesure expérimentalement l'utilité cardinale et utiliser le calcul des utilités pour prédire d'autres choix ultérieurs.

Ce modèle est une réponse aux deux critiques internes de Davidson et Suppes par rapport à leur propre modèle. Plus précisément, l'idée était de trouver, d'une part, un modèle qui n'est pas limité par la nécessité de trouver des résultats également espacés sur une échelle d'utilité (et donc ne nécessite plus l'usage de quantités de monnaie) et, d'autre part, un modèle qui ne soit pas incompatible avec des incohérences dans les choix (rappelons que dans le premier modèle, une incohérence dans le choix menait à rejeter l'ensemble des réponses du sujet plutôt que de considérer que le sujet était relativement à une réponse, irrationnel).

Le choix de la méthode de la programmation linéaire[97] est sans doute lié au contexte institutionnel dans lequel ont eu lieu les recherches de *Decision Making*. Comme nous l'avons souligné, l'armée américaine[98] finançait largement, dans les années 1950, les recherches portant sur la théorie des jeux et sur la théorie de la décision individuelle. Les recherches de Davidson et Suppes ne faisaient pas exception puisque comme on l'a vue, les deux articles de 1955 et 1956 étaient directement liés à des projets de recherche de l'Office of Naval Research comme les auteurs le mentionnent eux-mêmes dans la préface de l'ouvrage de 1957.

Sans entrer dans les détails de cette expérience, nous pouvons en présenter quelques caractéristiques saillantes.

Le premier élément frappant de cette expérience utilisant la méthode de la programmation linéaire est que les objets sur lesquels portent les préférences sont des morceaux de musique (Davidson, Suppes [1957]).

L'idée centrale de ce modèle consiste à construire toute une série d'inégalités du type $p(E^*)u(x_1) + p(\overline{E}^*)u(y_1) = p(E^*)u(y_1) + p(\overline{E}^*)u(x_1)$ à l'aide de plusieurs paires d'issues (des morceaux de musique). A partir de trente cinq inégalités et six variables, il était possible, selon les auteurs, de trouver une solution au système.

La procédure expérimentale de Davidson et Suppes se décomposent en trois sessions expérimentalement reliées.

L'objectif ultime était de comparer les résultats trouvés avec un modèle ordinal simple dans la lignée de l'article de 1955 où il s'agissait de déterminer quelle mesure de la valeur était la plus forte théoriquement. Ce modèle ordinal consistait simplement dans un classement des morceaux de musique du moins apprécié au mieux apprécié sans aucune indication sur les intensités de ces préférences. Le modèle cardinal, faisant usage des inégalités exprimant des différences d'utilité, avait pour objectif de décrire quantitativement ces intensités.

Au cours des trois sessions de l'expérience, les sujets font face de manière aléatoire à des paires d'options entre lesquelles ils doivent choisir, chaque session étant séparée

---

[97] Méthode qui n'est pas présentée par les auteurs, ceux-ci la considérant comme connue. Pour une présentation d'un exemple de programmation linéaire où la méthode est clairement présentée, on peut se référer à Duncan Luce et Raiffa [1957], pp.17-19.

[98] Les modèles de programmation linéaire[98] semblent en effet tout à fait appropriés pour des problématiques de l'armée puisqu'il s'agit avant tout de réduire les coûts pour un maximum de résultats.

l'une de l'autre par plusieurs jours d'intervalle. Les récompenses des différents paris étaient les morceaux de musique eux-mêmes, que les sujets se voyaient offrir à la fin de l'expérience.

Les résultats de cette expérience vont dans le sens d'une supériorité du modèle cardinal par rapport au modèle ordinal essentiellement en termes de nombre de prédictions correctes (Davidson, Suppes [1957], p. 92). Cependant, la principale difficulté réside dans l'une des améliorations qu'ont voulu introduire les auteurs dans ce nouveau modèle par rapport à celui n'impliquant que des quantités de monnaie. En effet, en ouvrant la possibilité d'incohérences comme la non-transitivité des préférences[99], les auteurs se heurtèrent à un problème de comparaison avec le modèle ordinal car le pouvoir de prédiction de ce modèle s'en trouvait amoindri (Davidson, Suppes [1957], p.100).

Il n'est pas encore là de motif déterminant la construction par Davidson d'un nouveau modèle mais les défauts que relèvent les auteurs eux-mêmes témoignent de leur scepticisme permanent.

3.7.2. La théorie de l'utilité espérée doit-elle être remise en cause ?

Outre les critiques internes faites par les auteurs eux-mêmes à leurs propres modèles, il existe tout un courant critique de la théorie de l'utilité espérée qui peut aussi s'appliquer à l'approche de Davidson, Siegel et Suppes.

L'une des particularités de *Decision Making* est qu'il est rarement fait mention des différentes critiques expérimentales antérieures adressées à la théorie de l'utilité espérée. Par exemple, il n'est pas fait référence (dans la bibliographie de l'ouvrage de 1957) à Maurice Allais et à son fameux paradoxe de 1953 alors que Davidson, Suppes et Siegel évoquent plus volontiers les expériences conduites par Edwards entre 1953 et 1954. Et pourtant, Allais était connu des auteurs puisqu'il est mentionné dans l'article de 1955 même si ce n'est qu'en note bas de page et pour

[99] La non transitivité correspond, comme on la vu, à la situation où le sujet, face à trois issues A, B, et C, préfère A à B, B à C et C à A. Nous avions évoqué ce problème dans notre présentation de Davidson, McKinsey, Suppes [1955].

insister sur les indications bibliographiques relatives aux critiques de la « validité explicative » de la définition d'un classement rationnel de préférences (voir note 10 p.155).

Ward Edwards est davantage cité par les auteurs de *Decision Making*. Edwards est l'un des auteurs les plus actifs de la psychologie expérimentale aux Etats-Unis dans les années 1950[100].

Comme nous l'avons mentionné plus haut, l'influence d'Edwards en psychologie expérimentale est considérable. Ainsi, comme le mentionnent Phillips et von Winterfeldt [2006], en publiant deux articles séminaux majeurs en 1954 et 1961, Edwards fonde la théorie de la décision comportementale qui constitue l'ancêtre de la théorie du prospect de Kahneman et Tversky [1979], dont Edwards fut d'ailleurs le directeur de thèse. En 1961, un nouveau champ émerge, né de diverses expériences mettant en évidence des violations du modèle de l'utilité espérée. Ce champ est géographiquement situé à l'université du Michigan où Edwards constitue un véritable groupe de travail composé de jeunes chercheurs comme Amos Tversky ou encore Paul Slovic. Edwards, fasciné par les travaux de Savage, incite même ce dernier à le rejoindre à l'université du Michigan, ouvrant la voie à une réflexion profonde sur les théories bayésiennes, réflexion prenant notamment la forme d'un article publié en 1963 (Bayesian Statistical Inference fot psychological research).

Davidson, Suppes, Siegel citent Edwards deux fois, d'abord lors d'une discussion de méthode pour se prémunir d'une distorsion qui pourrait intervenir dans les expériences et remettre en cause les résultats de celle-ci ([1957], p. 17). Cette distorsion est relative à la préférence observée par Edwards pour des paris d'une forme particulière (des paris composés d'issues pondérées par les probabilités ½, ½) par rapport à d'autres. Autrement dit, la forme des paris et les probabilités impliquées dans le pari ont une incidence, selon Edwards, sur le choix des sujets.

La deuxième fois qu'Edwards est évoqué (Davidson, Suppes, Siegel [1957], p.25), c'est pour signaler qu'à la différence de lui [1954a] qui propose une analyse des

[100] En 1954, Edwards propose de fournir un tour d'horizon de la littérature théorique relative aux expériences d'économie expérimentale en s'adressant à un public de psychologues (Edwards [1954, 1967], p. 13).

préférences (lors d'une expérience) sans faire usage de considérations relatives à l'utilité, les auteurs de Stanford supposent que l'utilité est linéaire en monnaie (Davidson, Suppes, Siegel [1957], p.25).

Mais la référence à Edwards ne s'arrête pas, en ce qui concerne Davidson particulièrement. Le philosophe cite en effet Edwards dans son article « La croyance et le fondement de la signification » publié en 1974 soit prêt de vingt ans après les expériences de Stanford.

Davidson n'a pas encore proposé sa théorie unifiée mais il s'intéresse toutefois au lien, voire à l'interconnexion entre théorie de la décision et théorie de l'interprétation du langage. Davidson porte un intérêt particulier à l'« effet de présentation », défini par Edwards et qui servirait de clé de voûte pour mettre en évidence l'interconnexion mentionnée. L'effet de présentation constitue en effet une critique de la théorie de l'utilité espérée, qui n'est pas déliée de la théorie de l'interprétation.

Pour comprendre cette critique, nous allons présenter dans un premier temps certains travaux réalisés par Edwards dans les années 1950 afin de saisir le type de distorsion qui peuvent affecter la théorie de1957 (3.7.2.1).

Puis nous insisterons sur l' « effet de présentation » lui-même le reliant au problème « d'interprétation » qu'il révèle (3.7.2.2).

3.7.2.1 Les expériences de Ward Edwards : 1953 et 1954a

Le point de départ des expériences proposées par Ward Edwards dans les années 1950 à l'université John Hopkins est l'hypothèse selon laquelle les sujets, même s'ils ont pour but de maximiser leurs gains espérés tout en minimisant leurs pertes, font des choix qui ne s'accordent pas avec un tel objectif (Edwards [1953], p. 349). En se focalisant sur les choix des sujets parmi des paris spécifiques, Edwards tente de montrer que les sujets se démarquent du modèle qu'il qualifie d'objectif[101] c'est-à-dire du modèle où les sujets choisissent les paris dont la valeur espérée (c'est-à-dire le produit des sommes monétaires et des probabilités) est la plus élevée. Edwards

[101] Dans le modèle objectif, la valeur espérée (EV) est égale à la moyenne de la récompense par jeu pondérée par les probabilités : $EV = p_1r_1 + p_2r_2 + \ldots + p_nr_n$ où $(p_1, \ldots, p_n)$ représente les probabilités et $(r_1, \ldots, r_n)$ représente les sommes monétaires.

n'utilise pas le terme « utilité » mais plutôt celui de « valeur ». Il considère en effet que le terme « utilité » est celui utilisé par les économistes. Or, comme nous le verrons, Edwards défendra l'idée que l'utilité, considérée comme la valeur subjective qu'accordent les individus à la monnaie, n'est pas une donnée pertinente pour expliquer les choix.

La stratégie d'Edwards est d'une part de montrer que les expériences de la théorie de l'utilité espérée comme celles de Preston et Baratta, et de Mosteller et Nogee prouvent que les sujets ne suivent pas le modèle objectif et, d'autre part, que les explications de cette déviance par rapport au modèle objectif recouvrent essentiellement une préférence des sujets pour certaines valeurs prises par les probabilités comparativement à d'autres. Plus précisément, Edwards propose plusieurs hypothèses permettant d'expliquer la déviance par rapport au modèle objectif (Edwards [1953], p.350):

i) La première hypothèse est qu'il se peut que les sujets essayent de maximiser leurs gains monétaires sans savoir comment le faire.

ii) La seconde est qu'ils interprètent mal les probabilités ou les sommes monétaires ou les deux.

iii) Troisième hypothèse : les sujets ne savent pas comment combiner des probabilités et des sommes monétaires pour déterminer la meilleure mise.

Mais aucune de ces pistes n'est suivie par Edwards. Ce dernier propose une expérience où le modèle objectif ne s'applique pas car les valeurs espérées de tous les paris sont toutes égales. Dès lors, il n'est plus possible de choisir, comme le suggérait le modèle objectif, les paris disposant des valeurs espérées les plus élevées. Cette procédure a pour but de mettre à jour les véritables déterminants des choix (Edwards [1953], p.351) et donc les variables qui influencent les sujets et les empêchent de prendre des décisions cohérentes avec le modèle objectif.

Les résultats de l'expérience indiquent, d'une part, une tendance générale à prendre ou à éviter les gros risques et, d'autre part, un ensemble de préférences pour certaines probabilités plutôt que d'autres[102].

[102] De même, Edwards remarque les paris les plus appréciés par les sujets étaient ceux où figuraient le verbe « gagner » et ceux qui étaient les moins appréciés étaient ceux où figure le verbe « perdre » (Edwards [1953], p.359).

En effet, Edwards remarque que deux probabilités particulières (4/8 et 6/8) cristallisent, respectivement une préférence spécifique et une aversion pour le risque. Autrement dit, pour ces deux probabilités, l'auteur remarque un comportement particulier des sujets, comportement qui relève selon lui d'une « préférence pour des probabilités ». L'idée d'Edwards est donc que les préférences sur les probabilités déterminent les choix dans cette expérience. Cette idée a, selon l'auteur, des conséquences significatives sur la possibilité d'une mesure de l'utilité comme celle proposée par vNM et expérimentée par Mosteller et Nogee. Autrement dit, si les intuitions d'Edwards sont correctes, cela voudrait dire que la possibilité de représenter les choix des sujets par une courbe d'utilité est remise en cause car si les sujets ont une préférence pour certaines probabilités, préférence qui détermine leurs choix, alors, l'utilité comme outil de mesure se trouve reléguer au second plan voire exclu de l'analyse (Edwards [1953], p. 363). La conclusion centrale d'Edwards est donc celle-ci : ce n'est pas la valeur subjective – l'utilité – qu'accordent les sujets aux sommes monétaires qui détermine et expliquer leurs choix. Même si nous voulions, selon Edwards, utiliser le choix pour mesurer à la fois les utilités et les probabilités[103], on devrait nécessairement connaître l'une des valeurs ou en fixer une ; mais cela ne permettrait pas de prendre en compte les préférences particulières sur les probabilités.

L'enjeu de cet article d'Edwards est fondamental car il laisse présager les développements ultérieurs de la théorie de l'utilité non-espérée qui se chargera de regrouper tous les phénomènes impliqués dans les préférences sur les paris de manière à proposer un modèle intégrant les différentes « anomalies » ou remises en cause empiriques de la théorie de l'utilité espérée (Edwards [1953], p. 354).

En 1954, Edwards généralise les conclusions de son article de 1953 puisqu'il propose une expérience où il sera proposé aux sujets de comparer des paris dont les valeurs espérées sont différentes. Cette fois encore, même si les préférences pour certaines probabilités constituent un phénomène moins saillant, elles expliquent une partie non négligeable des choix sur des paris dont les valeurs espérées diffèrent (Edwards [1954a], p. 66).

[103] Comme le suggère la méthode de Ramsey mais sans que cela soit évoqué par Edwards.

C’est au cours d’une autre série d’expériences qu’Edwards tentera de mettre en évidence d’autres éléments qui influencent, selon lui, le comportement de choix (Edwards [1954b], p. 68). Cette fois, Edwards évoque des problèmes méthodologiques. Il mentionne notamment la question de savoir si les résultats peuvent être reproduits, si les résultats antérieurs ont un effet sur le comportement de choix et plus encore s’il existe des effets de formulation (*wording* effect) des paris qui ont une influence sur les choix.

Nous allons nous intéresser plus spécifiquement à ce dernier effet car il est mentionné par Davidson lui-même comme un biais limitant la théorie d’utilité espérée (Davidson, 1974, 1993b).

3.7.2.2 Effet de formulation

En 1953, Edwards avait déjà mentionné certains éléments relatifs aux réactions des sujets aux paris durant les sessions d’expériences et après celles-ci. Mais Edwards ne les avaient pas exploités dans son modèle.

En 1954, il évoque ce qu’il appelle des « variables superflues » (*extraneous variables*) (Edwards [1954b], p. 76).

De nouvelles expériences étaient proposées aux sujets de manière à évaluer les rôles respectifs de la formulation des paris, ou encore des précédents résultats (fournis par la boule du flipper).

L’auteur mentionne que lors des expériences faisant usage de valeurs espérées équivalentes comme en 1953, les sujets semblaient considéré que les paris dont les valeurs espérées étaient nulles ressemblaient formellement à ceux dont les valeurs étaient positives (Edwards [1954b], p. 77). Les sujets semblaient davantage porter leur attention sur les formulations ou expressions de ce qu’ils pouvaient gagner plutôt que sur celles relatives à ce qu’ils pourraient perdre. L’auteur avance même l’idée que les sujets semblaient se focaliser sur le montant qui était présenté en premier dans le pari, que ce soit un gain ou une perte.

Pour rendre compte de ces effets, Edwards proposa d’inverser les formulations. Il observa cependant que cela ne faisait pas réellement de différences (*ibid.*). Dès lors,

les effets de formulation ne constituaient pas, selon Edwards, une remise en cause de l'hypothèse centrale selon laquelle les sujets ont des préférences pour certaines probabilités plutôt que pour d'autres. Davidson envisage différemment les conséquences de ces effets de formulation.

CONCLUSION

La théorie de Davidson, Siegel et Suppes qu'ils conçoivent comme la somme d'un modèle axiomatisé et d'une procédure expérimentale est une théorie de l'utilité espérée qui se présente immédiatement sous forme expérimentale. Grands lecteurs des théoriciens de la décision des années 1940 et 1950, Davidson, Siegel et Suppes y ont puisé nombres de concepts, méthodes et outils pour construire une théorie complexe dont l'objectif est la détermination simultanée des utilités et probabilités subjectives, autrement dit des désirs et des croyances, à partir de « données » comportementales.

Pour Davidson, cependant, cette théorie n'est pas sans critiques. Les effets de formulation, et plus généralement tous les effets que l'on peut regrouper sous la catégorie « effets de présentation » constituent des problèmes que la théorie de la décision standard ne permet pas de régler en l'état (Davidson [1976, 1993b]). Comme nous le verrons dans la partie qui suit, les objections soulevées par Edwards, comme par d'autres constituent non pas un motif pour rejeter définitivement la théorie de l'utilité espérée mais pour la remodeler très profondément.

Deuxième Partie :

Surmonter les défaillances théoriques et expérimentales de la théorie de la décision : l'introduction de la signification dans la théorie unifiée (Davidson, 1980)

Introduction

Au cours des années 1960-70, après la publication du modèle de 1957, Davidson publie nombre d'articles relatifs à la logique, la théorie de l'action et la philosophie du langage. Ceux-ci pourraient signer l'abandon de ces premiers travaux en théorie de la décision.

Pourtant il n'en est rien. D'une part, en effet, ces différents articles philosophiques présentent des critiques importantes de la théorie de la décision des années 1950 et témoignent ainsi de la préoccupation constante de Davidson pour cette dernière.

D'autre part, on assiste dans les années 1980-1990 à la publication d'une série d'articles ([1980], [1985], [1990]) dans lesquels Davidson propose explicitement de construire une nouvelle théorie de la décision.

C'est pourquoi nous commencerons par présenter les critiques émises par Davidson dans ces travaux philosophiques des années 1970 (chapitre 1) afin de montrer comment, s'additionnant aux critiques internes de l'économiste (partie I, chapitre 3), elles ont conduit à la nécessité de refondre et enrichir le modèle initial de 1957. Selon Davidson, la théorie de la décision formelle ne dit rien *à propos du monde* ; sa structure abstraite n'offre pas d'interprétation significative des termes qu'elle utilise comme le terme « préférer à » (Davidson [1999], p.32). Autrement dit l'une des critiques les plus importantes adressées à la théorie de la décision des années 1950 est relative à sa façon d'éluder les significations. Davidson choisit une nouvelle fois de s'appuyer sur Ramsey [1926]. Ce dernier cherchait, en effet, selon lui « avant tout à fournir une assise dans le comportement à l'idée qu'une personne accorde un certain degré de créance à une proposition » (Davidson [1974, 1993a], pp. 312-313). Or, il est d'usage de considérer que les propositions correspondent aux significations des énoncés d'un locuteur quelconque. L'idée est, comme on le verra, d'intégrer les propositions au cœur de la théorie de la décision de manière à avoir accès ou du moins parvenir à déterminer les significations qu'accordent les sujets à leurs phrases. Cette idée est la conséquence de l'utilisation de travaux de Quine par Davidson. Quine proposait en effet d'avoir accès

simultanément aux croyances et aux significations dans un processus de traduction radicale.

La théorie que Davidson construit dans les années 1980 prend appui à la fois sur la méthode de Ramsey permettant de déterminer simultanément les désirs et les croyances et sur les travaux de Quine relatifs à l'interconnexion entre croyances et significations. Il s'agit pour Davidson d'aller plus loin que le modèle de 1957 et d'intégrer des données mentales (les significations) qui jusque là, étaient négligées par le modélisateur et l'expérimentateur.

Pour ce faire, Davidson adopte une démarche singulière puisqu'il cherche à analyser dans une même théorie le triplet désirs-croyances-significations et ses relations internes. Dans la mesure où les désirs dépendent étroitement des deux autres pôles - les croyances et les significations - la théorie de la décision économique de Davidson ne peut donc être étudiée sans faire appel aux théories des croyances et des significations qu'il propose.

Après une explicitation détaillée de cette démarche particulière dans l'optique de construire une théorie unifiée de la pensée, de la signification et de l'action (chapitre 2), nous présenterons les axiomes formels du modèle de décision présentée dans les années 1980 que Davidson considère comme une expression de la théorie unifiée (chapitre 3). Nous montrerons en particulier qu'une grande partie de ce modèle est emprunté, selon les termes de Davidson ([1980, 2004], pp. 160-161) à la théorie de Jeffrey (1965,1983).

Enfin, nous chercherons à déterminer les apports et limites (chapitre 4) de ce nouveau modèle

- à la fois vis-à-vis de celui de 1957
- au regard des critiques internes et externes soulevées par Davidson. Il parle lui-même de leçons « tirées » de ces expérimentations ((Davidson [1999], p.32).

Chapitre 1. Les critiques de Donald Davidson à la théorie de la décision

Dans l'ouvrage *Decision Making* (1957), Davidson, Suppes et Siegel avaient pointé toute une série d'insuffisances et de défauts de la théorie qu'ils avaient testée. Comme cela a été souligné dans le dernier chapitre de la partie I, la solution behavioriste du modèle de 1957 se heurtait à des critiques méthodologiques et analytiques fortes relatives notamment à l'absence d'analyse des significations qu'accordaient les sujets aux différentes issues. Toutes ces considérations se trouvent renforcées lorsque Davidson coécrit avec Jacob Marschak l'article « Experimental tests of a stochastic decision theory » où les deux auteurs testent notamment la cyclicité des choix et ce, en mêlant à la fois le modèle de 1957 de Davidson et le modèle de Marschak de 1959 initié par Papandreou (1919-1996). Près de 15 ans après ces expériences de 1959, Davidson avoue, avoir ressenti un certain scepticisme quant à la possibilité pour la théorie de la décision de devenir une « théorie scientifiquement respectable » (Davidson 1974, 1993a, p. 313). Pour Davidson, une théorie scientifiquement respectable est une théorie qui formule des lois qui ont la même validité que les lois de la physique, à savoir être vraies en tout temps et en tout lieu. Comme il le souligne : « l'hypothèse selon laquelle une personne donnée, ou tout le monde, agit rationnellement au sens de la théorie de la décision, est une hypothèse empirique, testable, et probablement *fausse* » (Davidson [1976,1993], p.355, nos italiques).

Les nombreux articles qu'il publie entre 1963 et 1984 présentent ainsi toute une série de critiques adressées à la théorie de la décision et à son modèle de 1957 dont on peut distinguer trois grands types que nous détaillons dans ce chapitre :

- La première est relative à la conception spécifique des préférences de Davidson. Selon lui, la théorie de la décision n'offre pas de solution pour les cas de conflit entre les désirs. Une personne peut avoir une raison de préférer A à B et une autre raison de préférer B et A. La théorie de la décision standard fait l'impasse sur ce problème et ne nous dit pas pourquoi une issue se trouve préférée à une autre. Cette critique illustre le scepticisme de Davidson quant au modèle de 1957 (1.2).
- La deuxième critique, corrélée à la première, est que la théorie de la décision décrit une situation statique à savoir, la trame des attitudes et des croyances d'une personne à un moment donné (Davidson [1974,1993a], p.316). L'auteur est amené à se demander si l' « on peut accepter la théorie de la décision comme théorie scientifique du comportement et la placer sur le même plan qu'une théorie physique ? » (Davidson [1974,1993a], p.313) Cette critique renvoie directement aux expériences menées avec Jacob Marschak en 1959 (1.3).
- Davidson montre enfin que la théorie de la décision conçue en 1957 élude les significations. En effet, cette théorie présuppose que l'on peut identifier et individuer les propositions orientées vers les attitudes propositionnelles le désir et la croyance. Cependant, « notre aptitude à identifier ce qu'un agent désire ou croit ne doit pas être séparée de notre aptitude à comprendre ce qu'il dit (Davidson [1980,2004]) Cette critique renvoie directement à l'apport spécifique de Davidson à la théorie de la décision, apport présenté dans ses articles de 1980 et 1985 (1.4).

Puisque c'est en écrivant sur la philosophie de l'action que Davidson met au jour ces différentes critiques, plus encore puisque que c'est la première qui permet de comprendre ces dernières, avant de présenter ces trois grandes critiques, nous faisons un détour par sa conception de l'action (1.1)

1.1 Une conception de l'action qui éclaire les critiques faites à la théorie de la décision

La philosophie de l'action de Davidson – telle que développée notamment dans Davidson [1963,1993a], [1971,1993a], [1978,1993a] – est fondée sur une posture analytique singulière : les actions sont des événements (1.1.1). Pour comprendre cette position, nous présentons la théorie des événements développée par Davidson (a) et plus

précisément les deux types de descriptions (ou les deux niveaux d'analyse) des actions comme évènements, physique (par exemple la mise en action de certains neurones) et mental (par exemple une croyance ou un désir) (b). Ces deux niveaux de l'analyse de l'action par Davidson sont imbriqués analytiquement et cette imbrication renvoie à une autre relation, la relation de causalité, qui permet d'analyser les raisons de l'action comme les causes de celle-ci (c).

Cette présentation en trois parties de la théorie de l'action de Davidson – qui d'une certaine manière est l'expression du holisme de Davidson, nous y reviendrons – nous permettra ensuite d'en saisir les analogies et différences d'ordre analytique et opératoire avec la théorie de la décision (1.1.2). La théorie de l'action bâtie par Davidson dans les années 1960-70 permet en effet de mettre en lumière par le biais d'une analyse comparée, les défauts de la théorie de la décision (en particulier en matière de conflit des désirs et d'absence de significations.

1.1.1 Les actions comme évènements

La théorie de l'action de Davidson se fonde sur une thèse ontologique – c'est-à-dire une analyse de la nature et de l'identité de l'action - singulière : les actions sont des événements.

L'explication de cette position théorique tient d'une part à la théorie des événements de Davidson (1.1.1.1) et, d'autre part, à sa conception des relations entre mental et physique (1.1.1.2). C'est en particulier à partir de l'application de la théorie des événements aux événements mentaux que nous pourrons mettre en évidence la théorie de l'action de Davidson (1.1.1.3).

1.1.1.1 La théorie des événements de Davidson[104]

La théorie des événements repose sur deux éléments.

Elle postule d'abord que l'on peut décrire les évènements comme on décrit les objets physiques, les individus, les entités aussi appelées « substances ». De la même façon que l'on peut distinguer deux objets matériels, on peut distinguer deux évènements.

[104] Je suis ici la méthodologie d'Engel présentée dans [1993], [1994], [1994b] et [1997].

Davidson qualifie les évènements de « particuliers fondamentaux » (Davidson [1969, 1993a], p. 224), sans pour autant les définir avec précision. Il explique simplement que « dans la majeure partie de notre parler courant […] il existe une référence explicite à des entités [des évènements] ou à des quantifications sur elles » (Davidson [1969, 1993a], p. 224), quantifications qui en font « d'authentiques particuliers » (*ibid.*, p. 224). Autrement dit, « les événements sont une catégorie ontologique fondamentale […] Ils peuvent être dénombrés aussi aisément que les stylos, les pots et les gens » (Davidson [1969, 1993a], pp. 242-243).

Notons que le fait de considérer les événements comme des particuliers a une implication significative en matière d'analyse des significations des phrases : « sans événements il ne semble pas possible de rendre compte de façon naturelle et acceptable de la forme logique de certaines phrases de type usuel ; c'est-à-dire qu'il ne semble pas possible de montrer comment les significations de telles phrases dépendent de leur composition » (Davidson [1969, 1993a], p. 223). Autrement dit, la notion d'événement donne accès à une analyse de la signification d'une expression linguistique en se focalisant sur la détermination des significations de ses parties[105].

Le second élément de la définition des évènements repose sur l'idée de changement qui lui est associée. On peut en effet selon Davidson admettre sans difficulté que les entités physiques (substances) soient dotées de certaines propriétés (un cheval blanc par exemple). Pour Davidson, un évènement existe lorsqu'il est question de changements notamment dans ses substances[106]. Les changements comme par exemple des chutes de pierre ou des avalanches sont des événements, c'est-à-dire un certain type de choses qui arrivent.

Notons que pour Davidson, les entités matérielles ou physiques (ou substances)[107] et les changements (événements) sont interdépendants : « On ne peut concevoir ni la catégorie de substance ni la catégorie de changement indépendamment l'une de l'autre » (Davidson [1969, 1993a], p. 236).

---

[105] C'est ce qu'on appelle le principe sémantique de compositionnalité (Engel [1994], p.8).

[106] « De nombreux événements sont des changements au sein d'une substance » (Davidson [1969, 1993], p. 232).

[107] Cette distinction entre substances et propriétés est traditionnelle en philosophie depuis Aristote et plus tardivement Leibniz, comme le souligne Pascal Engel [1994b]

Enfin, il est essentiel de distinguer types d'événements et descriptions d'événements. Ainsi une éruption volcanique en Italie et une autre en Amérique du Sud correspondent à une même classe d'événements, celle des éruptions volcaniques. Parallèlement, un même événement peut être décrit de plusieurs manières : cette éruption est le résultat d'un mouvement des plaques tectoniques, elle a complètement détruit tous les villages voisins, elle causa la disparition de Pierre. Il y a donc une distinction entre événements et descriptions.

1.1.1.2 La théorie des événements appliquée au domaine du mental

Une fois précisée la théorie des évènements de Davidson, nous allons montrer qu'elle s'applique directement aux « événements mentaux ».

Pour ce faire, revenons sur la relation entre évènements mentaux et physiques.

Selon Davidson, un certain nombre de « théories de l'identité de l'esprit et du cerveau requièrent l'identification des événements mentaux à certains événements physiologiques ; si ces théories (…) ont un sens, les événements doivent être des individus [des entités physiques ou substances] » (Davidson [1969, 1993a], p.222). Autrement dit pour Davidson ces théories font de tous les évènements, même mentaux des entités physiques.

Mais il ne considère pas qu'il existe des lois psychologiques ayant la même valeur scientifique que des lois de la nature. Ainsi, même s'il accepte que les événements mentaux soient des événements physiques, il refuse de considérer que leurs lois respectives sont équivalentes. Cette position est claire dans l'exemple suivant :

Imaginons par exemple qu'un individu faisant du vélo, lève le bras pour indiquer qu'il va tourner à gauche. Cet événement peut être décrit en termes purement physiques comme l'activation de cellules dans le cerveau du cycliste, activation dont le résultat se manifestera par un bras qui se lève. Ce même événement peut être décrit en termes mentaux comme une croyance qu'en levant le bras, le cycliste pourra indiquer qu'il souhaite tourner à gauche.

Ces deux descriptions renvoient au même événement car elles sont elles-mêmes « instanciés »[108] par le même événement. L'événement physique et l'événement mental sont donc identiques puisqu'ils décrivent le même événement. Mais la description physique et la description mentale - qui sont elles-mêmes des événements en tant qu'elles sont des changements – ne sont pas identiques.
Les théories monistes évoquées par Davidson qui identifie évènements mentaux et physiques diraient qu'on a ici un seul et même évènement mais aussi une seule et même description de l'évènement.

Toutefois, Davidson ne souscrit pas à la thèse selon laquelle ces deux descriptions sont identiques c'est-à-dire que même si les deux types d'événements exemplifient le même événement, cela ne veut pas dire que la description mentale est équivalente à la description physique. La raison pour cela est que les lois qui régissent les événements physiques – autrement dit les lois de la nature – ne sont pas semblables à d'éventuelles lois mentales car selon Davidson ces dernières n'existent pas (la psychologie n'est pas une science pour l'auteur). Selon lui, en effet, il est possible pour un chercheur : « quand le monde entre en contact avec un individu, ou quand il se meut pour modifier son environnement, [d'] enregistrer et [de] codifier les interactions selon des procédures qui ont été raffinées par les sciences sociales et par le sens commun » (Davidson [1974, 1993a], p. 306). Toutefois, « on ne peut pas en tirer les lois strictes quantitatives qui figurent dans les théories sophistiquées et fiables que nous pouvons attendre de la physique, mais seulement des corrélations irréductiblement statistiques qui résistent, et résistent par principe, aux tentatives que nous faisons sans cesse pour les améliorer » (*ibid*.). Davidson refuse donc de réduire les événements mentaux aux événements physiques.

Cette thèse correspond à ce que Davidson appelle « le monisme anomal » décrit dans plusieurs articles (Davidson [1970] notamment). Nous allons l'expliquer en détail dans le paragraphe qui suit.

[108] Je reprends ici les termes de Pascal Engel [1994].

1.1.1.3 Les actions comme événements

La théorie de l'action proposée par Davidson au début des années 1960 repose comme nous l'avons dit sur la théorie des événements exposée au paragraphe (1.1.1.1).
Toutefois il n'y a pas un lien d'identité stricte entre une action et un évènement ; pour Davidson une action correspond à une sous-classe d'événements.

Comment alors reconnaître une action au sein des évènements ? Selon Davidson, il faut être capable de répondre à trois questions fondamentales : « Quels sont les événements qui, dans l'existence d'une personne, signalent la présence de l'agir ? A quoi reconnaît-on ses actes ou les choses qu'il a faites par opposition aux choses qui lui sont simplement arrivées ? Quelle est la marque distinctive de ses actions ? » (Davidson [1971, 1993a], p. 67).

Les réponses à ces questions, qui permettent de distinguer ce qui, dans la classe des évènements, constituent une action, résident dans l'analyse combinée de l'intention et de la sémantique des phrases d'action comme l'explique Davidson : « une personne *est l'agent* d'un événement si et seulement s'il existe une description de ce qu'il a fait qui rende vraie une phrase qui dit qu'il l'a fait intentionnellement » (Davidson [1971, 1993a], p. 71, nos italiques).
Pour résumer, en tant qu'évènement une action peut être décrite de plusieurs façons. Cependant ces descriptions se distinguent de celles d'autres évènements (qui ne sont pas des actions) par le fait qu'elles sont intentionnelles et exprimées par des phrases.

1.1.2. Théorie de l'action et théorie de la décision : analogie et différences

La théorie l'action développée par Davidson dans les années 1960-1970 possède de nombreux points communs avec la théorie de la décision, en particulier dans sa structure.

Les objectifs de ces deux théories sont d'abord liés puisqu'ils peuvent être présentés l'un en termes de raisons d'agir, l'autres en termes de choix d'actions : la théorie de l'action cherche à donner une explication des raisons d'une action isolée, alors que la

théorie de la décision apporte une explication au fait qu'un agent fasse un choix entre plusieurs actions possibles, qu'il a toutes de bonnes raisons de faire.

Pour remplir ces objectifs respectifs Davidson propose en outre deux théories aux raisonnements théoriques et schéma conceptuel proches.

En premier lieu, théorie de l'action et théorie de la décision prennent toutes deux appui sur une analyse des rôles respectifs des désirs (préférences) et des croyances (probabilités) (1.1.2.1). Elles sont, ensuite, construites à partir d'une même conception singulière du syllogisme pratique d'Aristote et s'inscrivent ainsi dans une perspective « téléologique » (1.1.2.2). Par l'adoption d'une méthode similaire, ces deux théories font en outre face à des limites communes comme l'incapacité pour les deux théories de fournir des lois du comportement qui aient valeur de lois scientifiques comme dans les sciences physiques (Davidson [1974, 1993a]) (1.1.2.3).

Outre ces similitudes, la théorie de la décision semble cependant occuper dans l'œuvre de Davidson une place différente de celle de la théorie de l'action car la première parvient selon lui à faire un pas supplémentaire en direction de la « respectabilité scientifique », et ce parce qu'elle fait usage d'une structure formelle composée d'axiomes qui permettent, par le biais d'une méthode opérationnelle du type de celle de Ramsey, de quantifier la force ou l'intensité des utilités et des degrés de croyance. C'est cette structure qui relie directement les causes de l'action à l'action elle-même (1.1.2.4).

1.1.2.1 Deux concepts fondamentaux au cœur de la théorie de la décision et de l'action: le désir et la croyance.

Dans l'œuvre de Davidson, comme nous l'avons dit, les croyances et les désirs occupent une place centrale (au côté des significations). L'utilisation de ces concepts dans la théorie de l'action comme celle de la décision pour décrire et expliquer une action constitue donc un point commun particulièrement important des deux théories.
Dans le cas de la théorie de l'action, Davidson mobilise les désirs et les croyances de manière systématique. Il explique par exemple que « chaque fois que quelqu'un fait quelque chose pour une raison, on peut dire *a)* qu'il avait une sorte de pro-attitude [un

désir] à l'égard d'actions de ce type, et *b)* qu'il croyait (ou savait, percevait, remarquait, se rappelait) que cette action était de ce type » (Davidson [1963, 1993a], p.16).
En théorie de la décision, comme dans le modèle de 1957 de Davidson, nous l'avons montré, les choix sont, à la suite de Ramsey, expliqués de manière simultanée par les utilités cardinales, qui représentent des désirs, et les probabilités subjectives, représentant les croyances.

1.1.2.2 Un schème explicatif commun

Les théories de l'action et de la décision sont construites autour d'un raisonnement analytique très proche.

Selon Davidson, on peut expliquer des choix en théorie de la décision de façon assez parallèle à la manière dont on explique les actions par des raisons.
Ces deux théories s'appuient, en effet, sur une forme singulière du raisonnement pratique qui représente les désirs et les croyances d'un agent comme les prémisses d'un syllogisme pratique[109] dont la conclusion est une action, ou un choix.

Au sein de la théorie de l'action, d'abord, le syllogisme pratique permet de mettre en évidence les composantes (mentales par exemple) d'une action. On peut illustrer ce raisonnement par l'exemple donné par Davidson dans « La faiblesse de la volonté » (Davidson [1970, 1993a]) : Imaginons un individu qui souhaite savoir l'heure. Son action peut être expliquée en la décrivant comme « le désir de savoir l'heure » et ce désir peut être exprimé par une proposition du type « Il serait bon pour moi de connaître l'heure » ou « Tout acte venant de moi qui a pour effet de connaître l'heure est désirable ». Le désir, conçu comme un principe d'action, est la prémisse majeure du syllogisme dont la conclusion est une action. La prémisse mineure étant la croyance de l'agent du type « Regarder ma montre aura pour effet que je connaîtrai l'heure ». La conclusion de cet exemple est qu'« en subsumant le cas sous une règle, l'agent accomplit l'action désirable : il regarde sa montre » (Davidson [1970, 1993a], p. 51).
Au sein de la théorie de la décision, le syllogisme pratique permet de mêler, d'une part, un ensemble de préférences rationnelles (c'est-à-dire conformes à certains réquisits

[109] Dont l'origine remonte à Aristote.

comme la transitivité) portant sur des résultats ou conséquences et, d'autre part, une estimation subjective des probabilités (ou degrés de croyance) portant sur l'occurrence d'un événement qui conditionne, en fonction de sa survenance, la réalisation de résultats ou conséquences. Ainsi, on peut considérer que les utilités et les probabilités constituent les prémisses d'un raisonnement dont la conclusion est une décision.

1.1.2.3 Limites communes

Outre partager concepts, démarches et méthodes, ces deux théories partagent aussi des défauts selon le point de vue de Davidson. Pour lui, en effet, ni l'une, ni l'autre ne peuvent prétendre constituer des théories scientifiques puisqu'elles ne peuvent produire des lois générales mais uniquement des corrélations entre des concepts.

En effet, la théorie de l'action telle qu'elle est conçue par Davidson, repose sur deux notions centrales : la notion de causalité et la notion de rationalité. Comme Davidson le mentionne lui-même : « Une raison est une cause rationnelle. L'une des façons dont la rationalité intervient est claire : la cause doit être un désir et une croyance à la lumière desquelles l'action est raisonnable » (Davidson [1974, 1993a], p.311). Autrement dit, l'usage mêlé de ces deux notions nous permet de décrire et d'expliquer de manière convaincante - sans supposer, selon Davidson, ce que l'on voudrait expliquer – l'action. Cependant, « le prix que nous payons pour avoir des explications est justifié : nous ne pouvons pas transformer ce mode d'explication en quelque chose qui aurait des allures plus scientifiques» (*ibid.*, p. 311). En effet, le raisonnement impliqué ici ne fait pas état d'observations quelconques, ou de manière plus neutre encore, de manifestations claires de ces deux notions de causalité et de rationalité. Autrement dit, la simplicité relative de la théorie de l'action l'empêche d'être rattachée à un ensemble d'observations détaillées ou à un ensemble d'analyses qui se recouperaient et ceci, notamment du fait de l'assimilation des actions à des événements. Ainsi, le raisonnement n'est valide que selon une certaine description, en ne tenant compte que de certains facteurs causaux et d'une certaine trame rationnelle qui correspond à la description de l'action. Il semble délicat, selon Davidson, d'étendre cette description à un cas général ou à une généralisation de l'explication des actions, ce qui n'empêche pas, toutefois, la description initiale d'être valable et précieuse.

La théorie de la décision telle que la conçoit Davidson rencontre le même problème et ceci du fait de sa vérification expérimentale puisque, rappelons-le, il explique : « l'hypothèse selon laquelle une personne donnée, ou tout le monde, agit rationnellement au sens de la théorie de la décision, est une hypothèse empirique, testable, et probablement fausse » (Davidson [1976, 1993a], p.355). Le fait que la théorie de la décision soit « probablement fausse » pour Davidson renvoie directement aux différents arguments utilisés par l'auteur notamment au sujet des prédictions faites par la théorie de la décision telle qu'il l'a testée mais aussi aux cas où la théorie n'apporte pas de réponse comme par exemple le conflit des désirs (voir section 1.2).

1.1.2.4 La théorie de la décision comme théorie sophistiquée des attitudes propositionnelles

La théorie de l'action développée par Davidson dans les années 1960-1970 procède d'une description et d'une explication de l'actions par des raisons qui en sont les causes. En 1976, Davidson décrit la théorie de la décision comme « une théorie sophistiquée des explications par les raisons » ([1976, 1993a], pp. 354-355). Ceci pourrait constituer une différence centrale entre théorie de l'action et théorie de la décision.
Afin de comprendre la spécificité de la théorie de la décision par rapport à la théorie de l'action, nous nous attardons sur l'usage du terme « sophistiqué » utilisé à plusieurs reprises par Davidson[110]. Nous identifions au moins deux explications de cet usage dans son travail :

i) La première est que contrairement à la théorie de l'action, la théorie de la décision est de part en part axiomatisée à l'aide d'outils mathématiques. Plus précisément, les concepts fondamentaux comme les désirs et les croyances sont insérés dans une structure axiomatique cohérente et traduits sous forme numérique. Cette structure axiomatisée a de nombreux attraits comme le souligne Davidson : « en posant des conditions formelles sur des concepts simples et sur leurs relations les uns aux autres,

[110] Le terme « sophistiqué » se retrouve au moins dans deux articles : « La psychologie comme philosophie » p.312, et dans « Explication de l'action selon Hempel » (1976), p. 355.

une structure puissante peut être définie » (Davidson [1999], p.32). La théorie de l'action n'a pas de tels attraits ; ses prémisses et ses raisons se prêtent moins facilement à une telle axiomatique. Une étape structurelle semble donc ne pouvoir être franchie par la théorie de l'action.

ii) La seconde raison repose sur les difficultés liées à la quantification des prémisses et raisons de la théorie de l'action alors que la théorie de la décision dans la version de Ramsey repose sur une quantification de la force de la préférence et du degré de croyance. Comme le dit Davidson, « La théorie [de l'action] ne part pas du principe que nous pouvons évaluer les degrés de croyance ou établir des comparaisons de valeur directement. Elle postule plutôt qu'il existe une structure de préférences raisonnables parmi les diverses actions possibles, et elle montre comment on peut construire un système de croyances quantifiées afin d'expliquer les choix » (Davidson [1975, 1993b] p. 236). Au contraire, la théorie de la décision de Davidson construite autour des modèles et expérimentations de 1957 et 1959 a montré que les désirs et les croyances pouvaient être identifiés, isolés et calculés, ce qui n'est pas possible en théorie de l'action : « Etant donné les conditions idéalisées que postule la théorie [de la décision], la méthode de Ramsey permet d'identifier et d'individualiser les croyances et désirs pertinents. Plutôt que de dire qu'il s'agit d'un postulat, il vaudrait peut-être mieux dire ceci : dans la mesure où nous pouvons considérer que les actions d'un agent correspondent à une certaine structure (rationnelle), nous pouvons expliquer ces actions dans les termes d'un système de croyances et de désirs quantifiés » (Davidson [1975, 1993b], p. 236).

La théorie de l'action ne permet pas un tel raffinement en raison de sa structure et des concepts sur lesquels elle s'appuie et en particulier le concept de rationalité : « Le concept d'action effectuée pour une raison (et par conséquent le concept de comportement en général) fait intervenir deux notions : la notion de cause, et la notion de rationalité. Une raison est une cause rationnelle. L'une des façons dont la rationalité intervient est claire : la cause doit être un désir et une croyance à la lumière desquelles l'action est raisonnable. Mais la rationalité entre en ligne de compte de façon plus subtile, parce que la manière dont le désir et la croyance fonctionnent afin de causer l'action doit satisfaire d'autres conditions, qui ne sont pas spécifiées. L'avantage de ce

mode d'explication est clair : nous pouvons expliquer le comportement sans avoir à connaître trop de choses sur la manière dont il a été causé. Et le prix que nous payons pour avoir des explications est justifié : nous ne pouvons pas transformer ce mode d'explication en quelque chose qui aurait des allures plus scientifiques » (Davidson [1974, 1993a], p. 311). Autrement dit, la structure qui sert de base à la théorie de l'action fait appel à une conception de la causalité et de la rationalité permettant de décrire des corrélations qui ne s'assimilent pas à des raisons nécessaires et suffisantes.

La différence centrale entre la théorie de l'action et celle de la décision chez Davidson tient au fait que la théorie de l'action proposée par l'auteur est une forme particulièrement simple d'explication par les raisons qui ne tient pas compte de la manière dont un agent fait des choix parmi plusieurs actions, actions « qu'il a toutes de bonnes raisons de faire » (Davidson [1976, 1993a], p. 354).

La théorie de l'action de Davidson semble occuper une position antérieure dans la chronologie d'un choix quelconque puisqu'elle se focalise essentiellement sur la manière dont un agent décide de faire telle action et les mécanismes mentaux qui sont en jeu à ce stade.

De même, la théorie de l'action n'intègre pas les variations dans la force des désirs ou dans les degrés de croyance, car ces questions relèvent de la théorie de la décision.

Toutefois, pour Davidson, la théorie de la décision est une théorie sophistiquée de l'action parce qu'elle fait un pas en direction de la « respectabilité scientifique » : « Elle abandonne le projet d'expliquer les actions une à une en faisant appel à quelque chose de plus fondamental, et postule à la place une structure sous-jacente au comportement à partir de laquelle on peut inférer les croyances et les désirs » (Davidson [1974, 1993a], p.313).

En effet, ce qui semble constitué un atout majeur de la théorie de la décision, selon Davidson, est que celle-ci utilise un mécanisme d'attribution d'attitudes comme des désirs et des croyances d'une portée empirique significative. Autrement dit, la théorie de la décision a accès à des données psychologiques sophistiquées comme des intensités de préférences qui sont significatives (ont du sens) selon Davidson car elles ne présupposent pas ce qu'elles veulent expliquer, elles sont accessibles immédiatement à l'observation durant une expérience (attention pas à l'observation pure). Ainsi, la

théorie de la décision, telle que la perçoit Davidson, postule une structure explicative reliée à l'observation expérimentale. Davidson poursuit : « On écarte du même coup toute nécessité d'établir l'existence des croyances et des attitudes indépendamment du comportement, et on prend en compte (comme relevant d'une construction théorique) l'ensemble du réseau pertinent des facteurs cognitifs et motivationnels » (*ibid*.). En ce sens, la théorie de la décision est – comme la théorie de l'action - intensionnelle[111] puisqu'elle est fondée sur le mode de la compréhension et à partir de notions qui expliquent et rationalisent le comportement.

En outre, ce qui différencie une fois de plus théorie de l'action et théorie de la décision est l'idée de « mesure » comme l'explique Davidson : « la théorie [de la décision] assigne des nombres afin de mesurer les degrés de croyance et de désir, ce qui est indispensable si elle doit produire des prédictions adéquates, bien qu'elle ne fasse ces mesures que sur la base de données purement qualitatives (des préférences ou des choix entre des paires d'options) » (*ibid*.). Nous reviendrons dans les chapitres qui suivent sur l'importance de cette idée de mesure pour Davidson.

Reposant sur des concepts et notions construits au cours des années 1970, la théorie de l'action de Davidson, que nous venons de présenter succinctement, en raison de son point de vue externe, permet de révéler trois grandes critiques adressées à la théorie de la décision que nous détaillons maintenant.

[111] L'intension est un concept logique qui s'oppose à l'extension. Toute classe d'éléments peut être définie en extension (en nommant ou en désignant chaque individu qui en fait partie) ou en intension, par une description (spécification d'un certain nombre de prédicats) qui définit la classe.

1.2 L’absence d’analyse des conflits entre les désirs en théorie de la décision

Comme nous venons de le montrer, on peut expliquer des choix particuliers en théorie de la décision de façon assez parallèle à la manière dont on explique les actions par des raisons. La différence principale entre la théorie de l’action et la théorie de la décision est que pour cette dernière, les désirs de l’agent deviennent comparatifs et quantitatifs.

Mais, comme le souligne Davidson [1976, 1993a], même si la théorie de la décision peut-être considérée comme une théorie sophistiquée de l’action, elle fait face à une difficulté majeure : l’absence d’analyse des conflits entre les désirs.

Une personne peut avoir une raison de préférer A à B et une autre raison de préférer B à A. La théorie de la décision « fait l’impasse sur ce problème, parce qu’elle ne nous dit en rien pourquoi un résultat principal se trouve préféré à un autre, parce que la théorie écarte toute possibilité de constater un conflit dans le comportement » (Davidson [1976, 1993a], p. 356).

Dans les années 1970, Davidson traite des conflits de désirs et il parvient à une analyse de ce problème en renouant avec la notion d’akrasie d’Aristote. Cette analyse prend place dans la théorie de l’action de Davidson comme nous le montrons dans un premier temps. (1.2.1). Ce traitement philosophique des conflits de désir rend d’autant plus saillant l’absence d’un tel traitement dans la théorie de la décision de 1957 (1.1.2).

1.2.1. Qu’est-ce qu’un conflit des désirs ?

L’absence de conflits de désir en théorie de la décision, notamment dans le modèle de 1957, apparaît d’autant plus saillante dans le travail de Davidson que le philosophe traite cette question lorsqu’il propose une analyse de l’akrasie - ou faiblesse de la volonté.

En 1970, en effet, l’auteur décrit ce qu’il appelle « un agent incontinent » : « On considère souvent comme une condition de l’action incontinente que celle-ci soit accomplie en dépit du fait que l’agent sache qu’une autre action est meilleure. […] Si un homme estime qu’une certaine ligne de conduite est, tout bien considéré, la meilleure

ou celle qui est correcte, ou la chose qu'il lui faut faire, et s'il fait malgré cela quelque chose d'autre, il agit de manière incontinente » (Davidson [1970, 1993a], pp. 37-38). Autrement dit, l'incontinence se manifeste, apparemment, par une contradiction dans l'évaluation de l'agent. Mais en réalité, il ne s'agit pas d'une contradiction logique[112] mais d'une rupture entre raison et cause comme on va le voir.

Pour préciser sa description de l'agent akratique, Davidson ajoute : « je dirais aussi qu'il agit de manière incontinente pourvu qu'il estime qu'une certaine ligne de conduite est globalement meilleure que celle qu'il choisit ; ou que, se trouvant placé devant une alternative entre une ligne de conduite qu'il croit à sa portée et l'action qu'il accomplit, il juge qu'il devrait accomplir cette autre action » (*ibid.*). L'incontinence se manifeste à la fois dans les jugements évaluatifs de l'agent mais aussi au niveau de ces choix. La particularité de l'incontinence est qu'elle n'est pas durable à moins que l'individu ne succombe continuellement au plaisir de l'action incontinente mais celle-ci sera alors un vice (Davidson [1970, 1993a], p. 42). Autrement dit, un agent est akratique, incontinent ou sujet à la faiblesse de sa volonté si, de manière ponctuelle, il agit contre son meilleur jugement et en cela, contre ses propres normes de rationalité.

Cette irrationalité chez l'agent incontinent pose problème à la théorie causale de l'action évoquée plus haut. Un agent qui accomplit l'action *x* l'a fait parce qu'il avait une certaine raison, cette raison se décompose, comme on l'a vu en un désir et une croyance. Or la raison (primaire) d'une action est sa cause. Si l'agent a accomplit *x* c'est qu'il juge qu'il serait meilleur de faire *x* plutôt que *y* par exemple. Or, l'agent incontinent juge, à la fois, qu'il est meilleur de faire *x* et qu'il est meilleur de faire *y*. Ainsi, pour reprendre la formulation du problème par Pascal Engel, on peut se demander comment concilier cette contradiction avec l'idée que ce sont les meilleures raisons qui causent l'action ? (Engel [1991], p. 12).

Davidson propose d'intégrer l'akrasie au sein de la théorie intentionnelle de l'action en révisant le raisonnement pratique pour y introduire une nouvelle dimension, les jugements conditionnels. En effet, en 1963, Davidson suggérait qu'une « personne, quand elle agit, connaît ses propres intentions de façon infaillible, sans recourir à l'induction ou à l'observation, alors qu'aucune relation causale ne peut être connue de

[112] Comme le mentionne Pascal Engel [1991], p. 11.

cette manière » (Davidson [1963, 1993a], p. 34). En 1970, dans son article « Comment la faiblesse de la volonté est-elle possible ? », Davidson avance l'idée qu'il existe « un fragment de raisonnement pratique présent dans le conflit moral, et par conséquent dans l'incontinence, et qu'[il a] jusqu'alors complètement négligé » (Davidson [1970, 1993a], p. 54). Ce fragment réside dans la relation qui relie les raisons à l'action. Cette relation est liée aux données dont dispose l'agent, aux jugements conditionnels qu'il produit en faisant le choix d'une ligne d'action. Ainsi, selon le raisonnement de Davidson, nous jugeons généralement que notre action est la meilleure tout bien considéré : « on caractérise l'*akratès* comme soutenant que, tout bien considéré, il serait meilleur de faire *b* que de faire *a*, quand bien même il fait *a* et non pas *b*, en ayant une raison de le faire » (Davidson [1970, 1993a], p. 61). L'agent incontinent qui fait *y* alors qu'il juge qu'il serait meilleur de faire *x* peut croire inconditionnellement que *x* est meilleur tout en faisant *y* car il estime *prima facie* (à première vue) que *y* est meilleur : « Le raisonnement pratique parvient néanmoins—souvent à des jugements inconditionnels selon lesquels une action donnée est meilleure qu'une autre, car si ce n'était pas le cas, on ne pourrait pas agir en ayant des raisons » (Davidson [1970, 1993a], p. 61).

L'article publié par Davidson en 1970 propose donc de considérer l'irrationalité de l'agent incontinent non pas comme une contradiction logique mais comme une déconnexion entre raison et cause : « Si *r* est la raison qu'a quelqu'un de soutenir que *p*, alors le fait qu'il soutient que *r* doit être, je pense, une cause du fait qu'il soutient que *p*. Mais, c'est ici le point crucial, le fait qu'il soutient que *r* peut être la cause du fait qu'il soutient que *p* sans que *r* soit sa raison pour cela » (Davidson [1970, 1993a], p. 63).

Plus précisément, en agissant contre son meilleur jugement, l'*akratès* ne parvient pas à agir sur la base de toutes les raisons qu'il considère comme pertinentes mais agit plutôt en se basant sur une raison qui n'est pas remise en cause par toutes les autres. Autrement dit, cette raison qui le pousse à agir contre son meilleur jugement n'est pas suffisamment contrebalancée par toutes les autres raisons et en particulier la raison qui serait la cause de son action. On comprend dès lors que Davidson considère que l'agent

incontinent est « sourd »[113] relativement à son propre comportement et qu'il ne parvient pas à se comprendre lui-même[114].

Précisons pour conclure sur ce point qu'il faut garder à l'esprit que l'essai « Comment la faiblesse de la volonté est-elle possible ? » s'intéresse essentiellement à un certain type d'actions que l'on peut considérées comme incontinentes. Davidson n'y fait référence à une attitude cognitive qu'indirectement. En effet, ce n'est que plus tardivement, au cours des années 1980, que Davidson évoquera plus spécifiquement ce qu'il appelle la « duperie de soi » en faisant référence explicitement à une attitude cognitive contradictoire qui se traduit par le fait que l'agent croit que *p* et croit, en même temps, que non-*p*. Il s'agirait ici d'un conflit des croyances et non plus d'un conflit des désirs[115] mais cette question est aussi éludée dans le modèle de 1957, ne serait-ce que du fait de l'utilisation de la méthode de Ramsey qui consiste à fixer les croyances (les probabilités grâce à l'événement aussi probable que sa négation) de manière à déterminer les intensités de préférences. Nous y reviendrons dans le chapitre 4 lorsque nous analyserons les apports et les limites du modèle de 1980. Pour le moment, nous allons voir si une analyse des conflits des désirs est intégrée dans le modèle de 1957.

1.2.2 L'absence de conflits de désirs dans les modèles des années 1950

S'il est largement question des conflits de désirs dans les écrits de Davidson des années 1970 et 1980 notamment lorsqu'il parle de faiblesse de la volonté, cette question demeure traitée dans le cadre de la théorie de l'action. Au contraire, la théorie de la décision s'avère, selon lui inapte, à prendre en compte ce genre d' « incontinences ».

En effet, l'une des seules irrationalités dans les préférences mentionnée par Davidson dans ces travaux en théorie de la décision dans les années 1950 est l'intransitivité. Comme cela a été mentionné dans la chapitre 3 de la partie I, l'argument de la pompe à

[113] Davidson [1969, 1993a], p. 65.
[114] Davidson fait d'ailleurs souvent référence à un plaisir qui résisterait aux voix de la raison comme dans Davidson [1970, 1993a], p. 49.
[115] Même si ces deux niveaux sont liés du fait de l'interdépendance des désirs et des croyances comme on le verra dans le chapitre 2.

finance décrit la situation où un individu n'ayant pas de préférences transitives pourrait se voir proposer par un joueur malin, une suite de paris qui le conduirait à sa ruine.

Mais dans ce cas, l'individu est considéré comme irrationnel puisqu'il entretient des préférences incompatibles entre elles. Nous sommes donc loin de la situation où un individu considère qu'il serait meilleur de faire *x* plutôt que *y* tout en faisant *x* avec une raison de le faire.

Ce genre de problème ne trouve apparemment pas de solution en théorie de la décision car cela voudrait dire que l'individu préfère l'action *x* à l'action *y* mais pourtant fait *x* alors que le contexte dans lequel a lieu cette décision est le même. Plusieurs exemples présentés dans la littérature permettent de défendre l'idée de préférences dépendantes du contexte[116]. Mais dans le problème présenté par Davidson, l'individu se place dans un seul et même contexte et agit contre son meilleur jugement avec une raison de le faire.

Une piste pour intégrer ce genre de d'irrationalité apparente pourrait être la proposition de Mosteller et Nogee selon laquelle les sujets d'une expérience sont indifférents entre deux issues lorsqu'ils choisissent soit l'une soit l'autre dans 50% des cas. Mais là encore, la théorie ne parvient pas à décrire la situation où l'individu agit contre son meilleur jugement. Que l'individu choisisse de faire *y* une fois sur deux et *x* dans les autres ne signifie pas qu'il estime *x* plus désirable que *y* puisqu'au contraire il les considère comme équivalents.

Ainsi, la théorie de la décision, dans sa forme standard, et telle qu'elle fut testée expérimentalement par Davidson ne parvient pas à intégrer au modèle, selon l'auteur, des situations de contradiction telle que celles de l'incontinence.

[116] Voir Sen [1993] ou encore Tversky et Simonson [1992, 2000].

1.3 La théorie de la décision comme théorie statique des préférences

C'est dans un article coécrit avec Marschak en 1959 que Davidson avait mis en lumière pour la première fois ce qu'on peut reconnaître comme une difficulté majeure de l'économie expérimentale de la décision : alors que la théorie de la décision est censée prédire les choix que ferait un individu dans telle circonstance particulière, elle ne parvient pas, dans sa forme générale statique, à prendre en compte les cas où une personne ne fait pas toujours le même choix lorsqu'elle est confrontée aux mêmes options, même quand les circonstances du choix sont les mêmes.

Les auteurs montraient que plusieurs stratégies avaient été proposées pour surmonter cette difficulté et en identifiaient quatre dans la littérature[117] :

1. La première stratégie consiste à analyser cette difficulté comme une déviance individuelle : il s'agit donc d'insister sur le statut normatif de la théorie et d'interpréter un écart par rapport à la norme qui voudrait qu'une personne fasse les même choix quand elle est face aux mêmes options, comme une déviance, comme la manifestation d'une erreur du sujet.
2. On peut défendre aussi l'exactitude descriptive de la théorie et arguer qu'elle a été mal interprétée par le sujet, que ce dernier a par exemple mal identifié deux options comme étant les mêmes (en disant par exemple que gagner 1$ au temps t revient à gagner 1$ au temps t+10 minutes). Cette stratégie permet ne pas, selon Davidson et Marschak, de remettre en cause la validité descriptive de la théorie.
3. La troisième stratégie consiste à interpréter chaque cas d'incohérence (c'est-à-dire un cas où une personne ne fait pas le même choix face aux mêmes options) comme un cas d'indifférence : si le sujet a choisi *a* plutôt que *b* puis peu après *b* plutôt que *a*, cela peut être interprété comme de l'indifférence entre ces deux

[117] Sans pour autant préciser quels auteurs avaient proposé ces différentes solutions.

objets. Toutefois, dans les applications empiriques, cette approche ferait de l'indifférence un élément central, faisant figurer la préférence stricte comme un cas marginal.

4. Une approche alternative aux trois précédentes suppose de définir la préférence et l'indifférence en termes de probabilités de choix dans le temps. Mosteller et Nogee, en testant les axiomes de vNM, considéraient un sujet comme indifférent entre deux options quand il choisissait chaque option la moitié du temps.

La stratégie qu'adoptent Davidson et Marschak, s'inspire en partie de cette quatrième posture puisqu'elle consiste à décrire l'ensemble des choix des individus à partir de leurs probabilités de manière à mettre en évidence une théorie stochastique de la décision, ou plus précisément une transitivité stochastique.

Cette stratégie s'inscrit directement dans la lignée du programme de recherche initié par Papandreou[118] (1919 - 1996). En effet, peu après avoir obtenu son doctorat d'économie à Harvard, Papandreou ouvre la voie, avec d'autres chercheurs comme Leonid Hurwicz (prix Nobel d'économie en 2007), à des travaux portant sur la programmation linéaire et non linéaire et ses applications en théorie économique. Ces différents travaux prenaient place au sein de la Cowles Commission, dirigée entre 1943 et 1948 par Jacob Marschak, et dont le mot d'ordre pouvait s'exprimer en une phrase : « la science est mesure ». Cette institution prestigieuse fondée par Alfred Cowles en 1932 et basée initialement à Colorado Springs, a connu de nombreux représentants tous très influents au sein de la sphère des théoriciens de la science économique. On peut citer par exemple Tjalling Koopmans (1910-1985), prix Nobel d'économie en 1975 et directeur de la Cowles Commission de 1948 à 1954 ; ou encore James Tobin (1918 -) prix Nobel d'économie en 1981 et directeur de la Cowles Commission de 1955 à 1967. C'est aussi au sein de cette Commission que furent publiées les monographies de Gérard Debreu, *Theory of Value: An axiomatic approach* et d'Harry Markowitz, *Portfolio Selection* en 1959.

L'article de Davidson et Marschak, publié la même année, prenait aussi place dans les recherches de cette commission. Un article collectif publié par l'équipe de Papandreou (Owen H. Sauerlander, Oswald H. Brownlee, Leonid Hurwicz et William Franklin) en

[118] Fondateur du Parti Socialiste grec puis dans les années 1980, Premier Ministre grec, mais aussi auteur d'études expérimentales dans les années 1940-1950 aux Etats-Unis.

1957, faisait déjà référence au terme « stochastique » au sens de formulation probabiliste de la théorie des choix, idée proposée en 1936 par Georgescu-Rogen (Moscati [2004], p. 19). En effet, selon ce dernier, définir les préférences en termes de fréquence des choix permettait d'éviter de considérer celles-ci comme une relation invariable.

Fechner (1876) avait déjà proposé des analyses portant sur le problème de la transitivité des préférences. Mais les premiers modèles de théorie de la décision moderne traitant de ces questions apparurent essentiellement dans le courant des années 1950 en même temps que plusieurs expériences furent menées pour évaluer et expliquer les cas d'intransitivités des choix. Tel fut par exemple l'objet de celles de Papandreou (1953) et May (1954).

Tous ces modèles (de Davidson, Marschak, Papandreou et Georgescu-Rogen) ont d'ailleurs une conception commune de la transitivité stochastique. Celle-ci sera dite forte si la probabilité que A soit préféré à B et B soit préféré à C est supérieure ou égale à ½, alors la probabilité que A soit préféré à C est supérieure ou égale à la plus grande des deux premières. La transitivité stochastique sera dite faible si l'on demande seulement que les trois probabilités soient supérieures à ½.

Si Davidson et Marschak choisissent de participer à l'élaboration de cette théorie stochastique générale du choix c'est qu'elle permet d'obtenir selon eux – à condition d'imposer certaines conditions sur les probabilités des choix - une forme de mesure qu'il qualifie de « plus forte » que la simple relation d'ordre[119] (Davidson, Marschak [1959], p.236). Elle constituerait une seconde méthode, en plus de celle de Ramsey, pour établir des échelles d'utilité (Davidson [1974, 1993a], p. 359). En effet, quand les conditions fortes de transitivité stochastique sont satisfaites, il est possible, selon les auteurs, d'interpréter une comparaison de probabilités comme une comparaison de différences de valeur subjective ou d'utilité[120]. Cette mesure, comme celle de Ramsey, est considérée comme étant « plus forte » car plus précise et riche que la simple information délivrée par un classement ordinal. C'est cette posture théorique qui est

[119] La raison pour cela est la même que celle mentionnée dans le chapitre 3 de la partie I relativement à une mesure par intervalles.

[120] On comprend ici qu'une théorie stochastique des préférences est compatible avec le théorème de représentation usuel de la théorie de l'utilité espérée.

notamment défendue en 1955 par Davidson dans son article coécrit avec McKinsey et Suppes (voir le chapitre 2 de la partie I).

Le modèle axiomatisé proposé par Marschak et Davidson et les résultats de leurs expériences regroupés dans Davidson et Marschak [1959] sont détaillés ici pour plusieurs raisons :
- ils mettent d'abord en lumière l'une des critiques fondamentales de Davidson à la théorie de la décision : c'est une théorie statique. En cela, elle ne permet pas – au moins dans sa version canonique - de prendre en compte et d'éviter les effets de mémoire et effets de contexte. Les conclusions de Davidson sur ce point sont proches de celles de Tversky proposées en 1975.
- ils permettent en second lieu de faire le lien entre le modèle de 1957 de Davidson et les critiques de ce dernier en tant que philosophe dans les années 1970. Cette expérience avec Marschak cristallise en effet les différentes critiques que Davidson fait à la théorie de la décision et ce, en particulier parce qu'elle constitue le dernier travail de recherche en économie expérimentale de Davidson.

Afin d'éclairer la position de Davidson relative aux effets de présentation et au problème de l'interprétation, nous présenterons dans un premier temps le cadre analytique proposé par Davidson et Marschak (1.3.1) pour ensuite mettre en évidence leurs résultats expérimentaux (1.3.2). Il s'agira enfin de présenter l'analyse que fait Davidson, en tant que philosophe, de ces expériences (1.3.3).

1.3.1 Cadre analytique

Etablissant d'abord ce qu'ils appellent des « hypothèses primitives » (*primitives*) (Davidson et Marschak [1959], p. 234) dont ils déduisent des « définitions » (*ibid.*), Davidson et Marschak construisent leur modèle de choix stochastique de choi autour d'hypothèses et axiomes relatifs à la transitivité stochastique et à la mesure de l'utilité par intervalles. Ils procèdent, pour cela, en plusieurs étapes.

Premièrement, ils énoncent deux premières hypothèses primitives :

Hypothèse Primitive 1. Supposons un ensemble A des issues. A, incluant des paris (des choix impliquant des risques) aussi bien que des revenus certains.

Hypothèse Primitive 2. Soit la probabilité P (a, b) d'un sujet qui, forcé de choisir entre a et b, choisit a. Alors

(a) P (a, b) + P (b, a) = 1

Cette hypothèse primitive impliquant que lorsqu'on demande à un sujet de choisir entre a et b, il choisit toujours a ou b[121].

(b) P (a, b) s'étend sur un intervalle ouvert (0,1).

Ces deux hypothèses primitives permettent aux auteurs de poser plusieurs définitions qui décrivent la construction théorique visant à surmonter la difficulté des choix intransitifs[122] mentionnée plus haut :

- **Définition 1**. a est absolument préféré à b si et seulement si P (a, b) = 1. Ce concept correspond à ce que les psychologues appellent la "parfaite discrimination".
- **Définition 2**. a est stochastiquement préféré à b si et seulement si ½ < P (a, b) < 1.
- **Définition 3**. a et b sont stochastiquement indifférents si et seulement si P (a, b) = ½.
- **Définition 4**. c est un point stochastique médian entre a et b si et seulement si P (a, c) = P (c, b).

Ces définitions, en particulier la première, présentent toutefois immédiatement une difficulté : la préférence absolue décrite par la définition 1 entre en effet en contradiction avec l'hypothèse primitive 2b (selon laquelle P (a, b) ne peut être égale à 1 et P (b, a) ne peut pas être égale à 0). Cette contradiction est particulièrement saillante dans le cas où a et b décrivent respectivement "recevoir m dollars" et "recevoir n dollars" alors m > n implique P(a,b) = 1. Plus généralement, si m_1 et n_1 sont des quantités d'une denrée et m_2 et n_2 sont des quantités d'une autre denrée, et $m_1 > n_1$, $m_2 \geq n_2$, alors l'issue consistant à

[121] Cette hypothèse ne sera pas testée par les auteurs.
[122] C'est-à-dire des choix successifs du type : a ≻ b, b ≻ c et c ≻ a.

recevoir m_1 et m_2 sera absolument préférée à l'issue consistant à recevoir n_1 et n_2. Mais la primitive 2a est essentielle à la construction du modèle de Davidson et Marschak si bien que les auteurs ont choisi de modifier l'ensemble A des issues de telle sorte à éviter la situation où une issue est absolument préférée à une autre.

Les auteurs ajoutent une quatrième « définition » tout à fait particulière, essentielle, selon eux, à la construction d'une théorie stochastique du choix. Cette définition suppose d'interpréter une comparaison de probabilités comme une comparaison de différences de valeurs subjectives ou d'utilités.
Utilisant une méthode similaire à celle de Fechner en psychophysique (1859), Davidson et Marschak construisent plus précisément une échelle subjective sur la base de « la fréquence des différences discriminées » (Davidson et Marschak [1959], p. 236). Ce glissement des probabilités aux utilités est décrit par la définition suivante :

- **Définition 5** : Pour un sujet donné, une fonction de valeur réelle u est appelée une fonction d'utilité dans A (dans le sens de la définition 5) si et seulement si, pour tout a, b, c et d dans A, $P(a,b) \geq P(c,d)$ si et seulement si $u(a) - u(b) \geq u(c) - u(d)$.

Pour éviter le problème des préférences absolues, les auteurs ajoutent à cette définition 5, la restriction suivante : « à condition que : ni P(a,b) ni P(b,a) ne soient égales à 0 ou à 1 » (Davidson, Marschak [1959], p. 237).

Les définitions suivantes découlent de la dernière. La définition 5 conduit en effet à l'analyse de plusieurs cas en fonction des conditions portant sur l'ensemble A des issues, analyse qui permet dans chaque cas de prouver l'existence d'une fonction d'utilité.
Chaque cas présenté constitue une variation de l'ensemble des issues sur lesquelles portent les choix mais également les conditions d'expérimentation.

- ***Cas (a)*** Si l'ensemble contient un nombre connu fini n d'issues : a_1, …, a_n, et il est toujours possible de stipuler les conditions des probabilités $P(a_i,a_j)$ nécessaires et suffisantes pour l'existence d'une fonction d'utilité.

- ***Cas (b)*** Si, cette fois, l'ensemble A contient un nombre arbitraire d'issues qui sont également espacées en termes d'utilité (de telle sorte que pour tout a, b, c et d dans A, si a et b sont adjacents en utilité[123] et c et d sont adjacents, alors, P(a,b) = P(c,d), on se trouve alors dans le cadre proposé par Davidson, Suppes et Siegel (1957). Pour analyser les choix, il suffira donc de se reporter à cet ouvrage présenté dans la partie I.

- ***Cas (c)*** Pour présenter ce cas, deux nouvelles définitions sont nécessaires selon les auteurs et permettent d'établir un théorème qui relie les probabilités et la fonction d'utilité :

  - **Définition 6**. On peut définir un ensemble A d'issues, stochastiquement continues si et seulement si les trois conditions suivantes sont réunies :

(i) il existe un point médian stochastique entre a et b.
(ii) si $P(c,d) > P(a,b) > ½$, alors, il existe un g tel que $P(c,g) > ½$ et $P(g,d) \geq P(a,b)$.
(iii) condition d'Archimède. Si $P(a,b) > ½$, alors pour toute probabilité q telle que $P(a,b) > q > ½$, il existe un entier positif n tel que : $q \geq P(a,c_1) = P(c_1,c_2) = \ldots = P(c_n,b) > ½$.

Cette définition revient à présenter les préférences à partir des probabilités et à reformuler l'idée de centre de gravité déjà présente cher vNM mais cette fois en l'appliquant à la dominance stochastique.

  - **Définition 7**. La condition portant sur le quadruplet a, b, c et d (quadruple condition) est satisfaite si et seulement si, pour tout a, b, c et d dans A, $P(a,b) \geq P(c,d)$ implique $P(a,c) \geq P(b,d)$.

Le théorème suivant est alors proposé :

[123] Soit $P(a,b) \geq ½$; alors a et b sont dit adjacents en utilité si $P(a,b) \leq P(a,c)$ pour tout c avec $P(a,c) \geq ½$.

Théorème I : Si A est stochastiquement continu alors il existe une fonction d'utilité si et seulement si la condition portant sur le quadruplet est satisfaite.

- ***Cas (d).*** Un résultat similaire a été obtenu par Dedreu (1958-1959) sous une définition différente des propriétés de continuité stochastique. Debreu a montré qu'il existe une fonction d'utilité dans A si les conditions suivantes sont satisfaites :

(i) si a, b, c sont dans A et $P(b,a) \geq q \geq P(c,a)$ alors il existe un d tel que $P(d,a) = q$

(ii) la condition portant sur le quadruplet tient pour A.

- ***Cas (e).*** L'ensemble A contient un nombre inconnu (peut-être fini) d'issues. Dans ce cas, aucun système d'axiomes n'est connu et il a été conjoncturé par Scott et Suppes (1958) que sous certaines restrictions naturelles sur la forme des axiomes, aucune axiomatisation n'est possible.

Les auteurs font remarquer que dans les cas (b), (c), et (d), les systèmes d'axiomes nécessaires pour prouver l'existence d'une fonction d'utilité (dans le sens de la Définition 5) permettent aussi de prouver que chacune de ces fonctions est unique à une transformation linéaire près (c'est-à-dire l'existence d'utilités cardinales).

Davidson et Marschak proposent de soumettre au test empirique l'hypothèse selon laquelle une fonction d'utilité existe dans A[124] ainsi que la transitivité stochastique.

- Pour tester l'existence de la fonction d'utilité, ils font l'hypothèse que celle-ci existe pour un ensemble contenant des paris monétaires.
  Si A est stochastiquement continu (c'est-à-dire si les conditions de continuité des probabilités sont satisfaites) comme dans les cas (c) et (d), que la condition portant sur le quadruplet s'applique pour T (qui est un échantillon de A), et donc pour A,

[124] Pour cela, il faut rajouter des conditions spécifiques concernant l'ensemble A. Si, par exemple, le nombre d'éléments de A est connu, on se situera dans le cas (a), ou si l'on considère que A est stochastiquement continu, on se situera dans les cas (c) et (d).

les auteurs affirment qu'il existe une fonction d'utilité dans A, en vertu du Théorème 1.

- Ce n'est pas la condition portant sur le quadruplet en elle-même qui est testée, mais les implications de celle-ci (Davidson, Marschak [1959], p.239). Ces implications renvoient à la transitivité stochastique telle que testée, non pas sur des quadruplets mais sur des triplets[125]. Comme cela a été mentionné plus haut, différents types de transitivité stochastique sont testés par Davidson et Marschak et ils sont décrits comme suit :

Condition 1.

(a) transitivité stochastique faible : tient dans A si et seulement si, pour tout a, b et c dans A, si $P(a,b) \geq ½$ et $P(b,c) \geq ½$ alors $P(a,c) \geq ½$.

(b) transitivité stochastique forte ; tient dans A si et seulement si, pour tout a, b et c dans A, si $P(a,b) \geq ½$ et $P(b,c) \geq ½$ alors $P(a,c) \geq \max [P(a,b), P(b,c)]$.

La condition 1(b) implique 1(a) mais 1(a) n'implique pas 1(b).

Notons en outre que les tests de l'existence d'une fonction d'utilité et de la transitivité stochastique sont intimement liés. En effet, les deux conditions sont des conditions d'existence d'une fonction d'utilité.

La condition 1(b) est équivalente a :

Condition 2. Si $P(a,b) \geq ½$, alors $P(a,c) \geq P(b,c)$[126].

---

[125] Les auteurs ne précisent pas si les résultats obtenus avec des triplets sont identiques à ceux obtenus avec des quadruplets.

[126] Pour montrer que la condition 1(b) implique la condition 2, supposons que $P(a,b) \geq ½$ et montrons que, par la Condition 1(b), $P(a,c) \geq P(b,c)$ pour chacun des trois cas :

1. $P(b,c) \geq ½$, alors $P(a,c) \geq \max [P(a,b),P(b,c)] \geq P(b,c)$.
2. $P(b,c) < ½ < P(a,c)$; alors $P(a,c) \geq P(b,c)$.
3. $P(b,c) < ½$, $P(a,c) < ½$; alors $P(c,a) > ½$, d'où $P(c,b) \geq \max[P(c,a), P(a,b)] \geq P(c,a)$, $P(c,a) \geq P(b,c)$.

On pourrait tenter, selon les auteurs, de prouver l'inverse, que la Condition 2 implique la Condition 1(b). Considérons trois issues fixées, a_1, a_2, a_3, et marquer les trois probabilités pertinentes : $P(a_1,a_2) = p_1$, $P(a_2, a_3) = p_2$, $P(a_3,a_1) = p_3$. Les deux types transitivité s'appliquent à un ensemble contenant a_1, a_2, a_3 peuvent être exprimés de manière symétrique par la condition 3 :

Condition 3.

(i) Transitivité faible : p_1, p_2, p_3 ne sont pas tous supérieurs ou égaux ou inférieurs ou égaux à ½ à moins qu'ils soient égaux à ½

Dans l'expérience rapportée dans l'article de Davidson et Marschak, seuls des triplets étaient utilisés. Les auteurs notent alors que si l'ensemble des issues consiste simplement en trois éléments a, b, et c, alors la condition de transitivité stochastique forte est une condition nécessaire et suffisante pour qu'il existe une fonction d'utilité.

Un autre type de test sera opéré en utilisant cette fois l'événement aussi probable que sa négation, décrit dans le chapitre 1. Cet événement correspond à l'événement éthiquement neutre de Ramsey (voir Partie 1 chapitre II), dont la probabilité est de ½. Le test de ce type de spécificité a bien entendu nécessité que les auteurs ajoutent des conditions supplémentaires pour intégrer cet événement particulier.
L'introduction de l'événement aussi probable que sa négation implique en particulier de présenter deux hypothèses primitives additionnelles :

- **Hypothèse primitive 3** : Soit un ensemble X d'états du monde. Les sous ensemble de X sont appelés des événements, notés E, F et formant un ensemble ε.

- **Hypothèse primitive 4** : Si a, b sont dans A et E est dans ε alors aEb est un pari qui consiste à obtenir a si E survient et b si E ne survient pas.

En intégrant cette dernière hypothèse primitive aux définitions 1 à 4, l'hypothèse sur les probabilités P (aEb, cFd) = 1/2 signifie que aEb et cFd sont stochastiquement indifférents.

L'évènement E peut donc être défini plus explicitement par Davidson et Marschak selon la « définition » suivante :

(ii) Transitivité forte :
$p_1 \geq ½$ si et seulement si $p_2+p_3 \leq 1$,
$p_2 \geq ½$ si et seulement si $p_1+p_3 \leq 1$,
$p_3 \geq ½$ si et seulement si $p_1+p_2 \leq 1$

- **Définition 8**. Un événement E dans ε est un événement aussi probable que sa négation si et seulement si pour tous a et b dans A, (aEb, bEa) = ½.

Dans cette veine, si un événement E est aussi probable que sa négation – si sa probabilité est ½ - on peut décrire le pari aEb comme un pari dont les issues sont conditionnées à la même probabilité (*even-chance wager*).

L'introduction de l'évènement entraine par ailleurs l'introduction d'une neuvième « définition » concernant les choix des sujets :

- **Définition 9** : Le sujet sera dit non-biaisé si et seulement si, pour toute paire d'événements aussi probables que leurs négations E et F, et pour tous a et b dans A, P(aEb, aFb) = ½. Si cette condition est satisfaite, il existe une fonction d'utilité *u* dans A telle que pour toute paire d'événements aussi probables que leurs négations E et F, et pour tous a et b dans A, *u* (aEb) = *u* (bEa) = *u* (aFb) = *u* (bFa).

La définition 9 nécessite d'imposer une condition restrictive supplémentaire sur la définition 5 exposée plus haut.

- **Définition 10** : Une fonction à valeurs réelles *u* est une fonction d'utilité relative à des paris d'égales probabilités (*even-chance wager*) (ou une fonction d'utilité au sens de la définition 10) dans A si et seulement si :
  (i) *u* est une fonction d'utilité dans A dans le sens de la Définition 5 ;
  (ii) pour tout a et b dans A et tout événement E aussi probable que sa négation, *u* (aEb) = $[\frac{u\,(a)}{2} + \frac{u\,(b)}{2}]$.

La définition 10 exprime l'hypothèse d'utilité espérée sous forme stochastique[127].

1.3.2. Expérimentations et résultats

L'intérêt de la démarche de Davidson et Marschak a trait aux différentes expériences qu'ils ont menées pour tester la transitivité stochastique (et donc logiquement, comme nous l'avons expliqué, l'existence d'une fonction d'utilité). Nous commençons par présenter ici leurs procédures d'expérimentation en raison de sa complexité (mélange de cartes stimulus et dés) (1.3.2.1) avant d'en venir aux résultats (1.3.2.2).

1.3.2.1 Procédure d'expérimentation

L'expérience de Davidson et Marschak avait pour objectif de tester la validité de l'hypothèse selon laquelle, pour des individus donnés, il existe une fonction d'utilité dans le sens de la définition 10.

Dix-sept étudiants (six femmes et onze hommes) d'une classe de logique élémentaire de l'université de Stanford prirent part au test. Deux types de tests furent menés : la transitivité stochastique (faible et forte) d'issues présentées sous forme de triplets (interprétés comme des paris) et la transitivité stochastique (faible et forte) d'intervalles d'utilité présentés par groupes de six issues.

Ces expériences, comme celles de Davidson, Suppes et Siegel [1957], faisaient usage de cartes-stimulus (*stimulus cards*) et de dés. Ainsi pour tester la transitivité des choix parmi les issues, les sujets faisaient face à des cartes de ce type :

[127] A cette définition 10 s'ajoute des hypothèses supplémentaires concernant l'existence d'un point stochastique médian et des conditions permettant de décrire l'ensemble A comme étant stochastiquement continu (comme pour les définitions 6 et 7 ou pour les cas (c) et (d)), hypothèses sur lesquelles nous ne nous attarderons pas.

	A	B
ZOJ	-5¢	+36¢
ZEJ	-21¢	-38¢

Tableau 9 – Carte 1

	A	B
QUG	+36¢	-54¢
QUJ	-38¢	+22¢

Tableau 9 – Carte 2

	A	B
WUH	-54¢	-5¢
XEQ	+22¢	-21¢

Tableau 9 – Carte 3

L'événement consistant à obtenir ZOJ ou ZEJ était produit par un dé spécial dont trois des faces portaient le symbole ZOJ et les trois autres le symbole ZEJ comme dans l'expérience de Davidson en 1957[128]. De la même manière, trois dés différents étaient utilisés dont chacun avait une paire de syllabes différente. Les sujets devaient choisir la colonne des cartes-stimulus alors que le dé déterminait le rang.
On demandait plus précisément à chaque sujet de faire 319 choix, un choix consistant en une réponse verbale (A ou B) à la carte stimulus proposée. Suite à son choix, le sujet se voyait remettre ou devait payer le montant de son gain ou de sa perte.

Les auteurs proposèrent aussi un test de l'existence d'une fonction d'utilité en relation à des paris d'égales probabilités, en vérifiant si les événements crées par les trois dés étaient des événements aussi probables que leur négation. Il s'agissait de s'assurer que pour tous les m et n représentant des sommes monétaires, la condition suivante était respectée :

[128] Tout comme Davidson, Suppes et Siegel, Davidson et Marschak tentèrent d'éviter l'effet de mémoire (effect of memory) en ne proposant jamais deux fois les mêmes paris.

P(mEn, nEm) = ½ si et seulement si P(mEn, n-1Em) > ½ et P(mEn, n+1Em) < ½. Ceci correspond parfaitement à l'équation proposée par Davidson, Suppes et Siegel pour trouver un encadrement à la relation $\approx E *$ de l'hypothèse H1 (voir Partie I chapitre 2).

Comme cela a été mentionné plus haut, les sujets devaient choisir la colonne des cartes-stimulus alors que le dé déterminait le rang.

Si l'on pose à présent, comme l'expliquent les auteurs, qu'il existe trois probabilités reliées p_1, p_2, p_3 telles que définies par la condition 3 (c'est à dire la transitivité faible et la transitivité forte, voir note 22). Notons $p^i = < p_1^i, p_2^i, p_3^i >$ qui est un point du « cube unité » U (*unit cube*), cube dont tous les côtés ont une unité de longueur[129]. En réalité, il s'agit simplement de représenter les trois probabilités, en même temps, dans un espace à trois dimensions et c'est cette idée que recouvre l'expression « cube unité ».

Ainsi, Davidson et Marschak décrivent plusieurs sous régions de U :

La région W correspond à la condition 3 (i) c'est-à-dire la région de transitivité faible, de même la région S correspond à la condition 3 (ii). S étant inclut dans W.

Dès lors, si le sujet choisit la colonne A dans le tableau 1-carte 1, il y a plus de chance que $p_1>1/2$ au lieu de $p_1<1/2$; si le sujet choisit la colonne A dans le tableau 1-carte 3, il y a plus de chances que $p_3>1/2$ au lieu de $p_3<1/2$. Le triplet de probabilités a donc de grandes chances de se situer dans la région de transitivité forte.

Une observation est alors définie comme un triplet ordonné de réponses. Pour les trois cartes présentées plus haut, il y a 8 cas possibles :

$O_1 = < A, A, A>$

$O_2 = < A, A, B>$

$O_3 = < A, B, A>$

$O_4 = < A, B, B>$

$O_5 = < B, A, A>$

$O_6 = < B, A, B>$

$O_7 = < B, B, A>$

$O_8 = < B, B, B>$

[129] Cube à trois dimensions permettant de représenter les trois probabilités en même temps, et de mettre en relief par la même occasion les différentes « régions » dans lesquelles se situent les probabilités.

Par exemple l'observation O_1 se lit comme suit : le sujet a choisit l'option A lorsqu'il était face aux cartes 1, 2 et 3 du tableau 1.

Comme les auteurs le mentionnent, les observations O_1 et O_8 seraient considérées comme des cas d'intransitivité ; dans la théorie stochastique, elles renforcent la preuve de cette intransitivité stochastique. Ces observations sont caractérisées de cycliques. L'observation O_1 constitue précisément un cycle puisque l'on voit que le couple (-5¢, -21¢) est préféré au couple (+36¢, -38¢) dans la première carte, que le couple (+36¢, -38¢) est préféré au couple (-54¢, +22¢) dans la seconde carte, et qu'enfin, le couple (-54¢, +22¢) est préféré au couple (-5¢, -21¢) dans la troisième carte, ce qui constitue un cycle.

Il s'agissait ensuite, pour Davidson et Marschak, de calculer l'espérance du nombre d'observations cycliques lorsque l'on se situe dans l'une des régions du cube unité c'est-à-dire dans la région de transitivité forte ou dans celle de la transitivité faible. A partir du calcul de ces espérances – calcul réalisé à partir des observations – Davidson et Marschak étaient en mesure de comparer les observations cycliques récoltées pour chaque sujet aux observations cycliques espérées.

1.3.2.2 Résultats de l'expérience

Le tableau suivant – concernant uniquement le test de la transitivité sur des issues présentées sous la forme de triplets - retranscrit les résultats des expériences de Davidson et Marschak.

Les auteurs précisent que 76 observations furent faites pour chaque hypothèse.

Sessions	I	II	III	Total
Nombre de triplets offert	22	28	26	76

Espérance du nombre d'erreurs cycliques sous une distribution uniforme dans la région de :	Nombre d'observations cycliques

Cube unité	5.50	7.00	6.50	19.00
Transitivité faible	4.13	5.25	4.88	14.25
Transitivité forte	3.03	3.85	3.58	10.45

Sujets				
A	4	0	0	4
B	3	5	2	10
C	5	3	3	11
D	4	7	0	11
E	1	0	0	1
F	3	6	0	9
G	2	1	2	5
H	2	1	1	4
I	1	2	1	4
J	2	9	5	16
K	4	2	2	8
L	1	1	0	2
M	2	2	1	5
N	6	3	7	16
O	4	2	1	7
P	1	3	3	7
Q	7	5	2	14
Moyenne des observations cycliques	3.06	3.06	1.76	7.88

Proportion des observations cycliques	13.9%	10.9%	6.8%	10.4%

Tableau 10

Voici les informations que l'on peut tirer de ce tableau :

La première partie du tableau regroupe les prédictions, espérances de fréquence de préférences cycliques pour la transitivité forte et faible, la seconde regroupe les observations.

Le nombre de réponses cycliques est bien inférieur pour la majorité des sujets à l'espérance de la fréquence des préférences cycliques la transitivité stochastique forte.

Même si la moyenne du nombre de réponses cycliques pour les sujets de 10.4 est quasiment identique à la prédiction de 10.45 faite pour la transitivité stochastique forte, certaines évolutions du test furent difficiles à expliquer. Leurs prédictions ne semblent donc pas se vérifier avec autant d'envergure pour la transitivité faible notamment, comme si le modèle malgré ses précisions et l'accumulation de conditions ne pouvaient traiter pertinemment de l'évolution des choix dans le temps.

Par exemple, Davidson et Marschak s'étonnent du fait qu'au fur et à mesure des sessions, la proportion d'observations cycliques baissait, passant de 13.9% à 6.8% alors qu'ils avaient mis en œuvre plusieurs procédés pour éviter les effets comme l'effet de mémoire mentionné plus haut (Davidson, Marschak [1959], p.260).

Cependant cet étonnement n'entraine pas de tentative d'explication par les auteurs, Davidson et Marschak n'adhèrent pas à l'idée que ce serait la valeur actualisée des paris (cartes stimulus) ou le degré de risque qui expliquerait ce phénomène même si on se décidait, à la manière de Mosteller et Nogee (1951) à classer les individus selon leur caractère conservateur ou aventurier.

Si, pour Davidson, la question de la dominance stochastique n'était pas centrale en 1957, il essaie deux ans plus tard d'introduire cette analyse pour parvenir à une théorie des choix plus dynamique.

Pourtant, même le modèle stochastique qu'il essaie de construire avec Marschak ne donnent pas d'outils suffisant pour analyser les choix dans le temps, pour prendre en compte les effets d'apprentissage et de mémoire. C'est précisément ce défaut

d'explication qui sera pointé par Davidson dans les années 1970 et qu'il cherchera à pallier en 1980, dans un nouveau modèle.

1.3.3. Le regard de Davidson en 1974 sur l'expérimentation de 1959 : effets d'apprentissage et distorsions liées à l'expérimentation

Outre la présentation de l'axiomatique et les expérimentations de 1959, c'est surtout la façon dont Davidson les envisage en 1976 dans son essai « L'explication de l'action selon Hempel », qui nous intéresse ici. Il revient sur le caractère non dynamique de la théorie de la décision qu'il a lui-même testé et ses implications (1.3.3.1), une analyse très proche de celle de Tversky qui parle d'effet de certitude (1.3.3.2) et de problème d'interprétation (1.3.3.3).

1.3.3.1 Des expériences initiales qui révèlent le caractère statique de la théorie de la décision

Selon lui, le modèle proposé dans l'article écrit en 1959 avec Marschak constitue une seconde méthode permettant d'établir des échelles d'utilité – la première méthode étant celle proposée par Ramsey et expérimentée dans le modèle de 1957. Cependant, le jugement que Davidson porte sur les expériences de 1959 est plutôt négatif puisqu'en 1974 il écrit : « Je ne décrirai pas l'expérience, qui était très compliquée, ni les résultats, qui étaient indéchiffrables » (Davidson [1976, 1993a], p.359).

Alors que le modèle de 1957 ne permettait pas de résoudre les conflits de désirs chez un sujet (voir section 1.2), le modèle de 1959 pose lui aussi problème puisqu'il ne permet pas de constituer une théorie dynamique satisfaisante des préférences des individus. Davidson estime simplement que le modèle de 1959 était incapable de rendre compte ou prédire les évolutions des choix dans le temps comme il le dit en 1976 : « comment pouvions-nous dire que les sujets n'étaient pas influencés dans leurs préférences par l'expérience elle-même – que leurs préférences ne changeaient pas au fur et à mesure que nous avancions dans nos expériences ? » (Davidson [1976, 1993a], p. 360).

Autrement dit 15 ans après l'écriture du modèle de 1959, le contrôle total des conditions d'expérimentation et de l'influence des paris eux-mêmes sur les sujets lui semblaient être une chimère : « *La théorie de la décision entend décrire une situation statique* : la trame des attitudes et des croyances d'une personne à un moment donné. Nous faisions tous nos efforts pour ne pas changer les préférences d'un sujet au cours d'une session. Si vous offrez à un sujet un pari, et qu'il le prend, et s'il voit ensuite une pièce virevolter ou un dé rouler, cela peut changer ses attentes à la fois suivante. Si l'argent change de mains, il peut être « rationnel » de sa part de considérer le gain ou la perte suivante sous un nouveau jour » (*ibid.*, *nos italiques*). Autrement dit, la procédure expérimentale en elle-même pouvait selon Davidson constituer ce que l'on pourrait appeler une distorsion ou un biais dans l'expression des préférences des sujets.

En 1976, Davidson identifie plus précisément ces différents problèmes à des effets de « conditionnement ou d'apprentissage », problèmes que la théorie de la décision standard ne permet pas selon lui de régler en l'état.

Davidson expliquait en effet dès 1957 que sa théorie ne constituait en rien une théorie particulière de l'apprentissage (Davidson, Suppes et Siegel [1957], p.79). Il expliquait, avec Suppes et Siegel, que le cadre théorique dans lequel interviennent les intensités de préférences et les probabilités subjectives était trop étroit mais qu'à partir du moment où l'on essayera de proposer une théorie de la décision dynamique alors elle devra être insérer dans une théorie de l'apprentissage et de la motivation (Davidson, Suppes et Siegel [1957], p. 80).

Bref, la théorie standard de la décision étant une théorie statique, il n'est pas possible, selon Davidson, pour les expérimentateurs d'empêcher les sujets de s'appuyer ou de se remémorer leurs choix antérieurs même si tout était fait pour maîtriser et limiter cet effet de mémoire ou d'apprentissage en 1957 et 1959. Un test empirique représente donc une coupe instantanée des préférences d'un ensemble de sujets. Or il semble clair que les sujets apprennent à jouer et à parier au fur et à mesure des expériences comme Davidson l'avait lui-même suggéré lors des expériences de 1957. En effet, les tests permettant de vérifier les hypothèses du modèle expérimental étaient, comme on l'a

vue, reproduits[130] quelques jours ou quelques semaines après les premiers tests. Sans conclure sur un véritable effet d'apprentissage (seul un des sujets semblait se rapprocher significativement de la linéarité en monnaie), les auteurs avaient évoqué une certaine duplication dans les choix des sujets. Pour la plupart d'entre eux, les choix étaient identiques d'une session à une autre et cet effet de mémoire[131] « étonnait » (*astonishing* feat) les auteurs.

Davidson mentionne d'ailleurs une autre expérience - qui ne semble pas avoir été publiée – menée avec Merrill Carlsmith (Davidson [1974, 1993a], p. 314). Les sujets devaient examiner un certain nombre d'options et faire un choix entre plusieurs paires d'options qui leur étaient proposées. Ces options étaient présentées plusieurs fois aux sujets afin de tester la cohérence de leurs choix. Pour s'assurer qu'aucun effet d'apprentissage ou de conditionnement n'entrait en jeu, les options avaient une allure suffisamment complexe pour que les individus ne remarquent pas qu'elles étaient identiques d'une session à une autre. Seulement, l'auteur s'aperçut que les cas d'intransitivités étaient progressivement éliminés. Ce constat constituait pour Davidson une preuve que la théorie de la décision, en tant que théorie statique, ne pouvait faire des prédictions viables. En effet, si, comme on l'a vu, la théorie de la décision standard s'appuie sur des résultats expérimentaux pour prédire d'autres résultats hypothétiques, elle ne peut être en mesure d'intégrer des effets d'apprentissage au cœur de la théorie puisqu'elle prédit qu'un individu qui manifeste des préférences et des choix pouvant être représentés par la théorie de l'utilité espérée, fera des choix semblables s'il se trouve dans une situation similaire dans la mesure où il est rationnel, au sens de la théorie, de se comporter ainsi. Dès lors, si la théorie prévoit une continuité dans les préférences et les choix d'un individu, des effets de mémoire ne sont pas envisageables[132]. Or, l'expérience menée avec Carlsmith montre, selon Davidson, que « le simple fait de faire des choix (sans récompense ni effet de retour) change les choix ultérieurs effectués par les agents » (Davidson [1974, 1993a], p. 314). Il existe certainement, selon l'auteur, des « valeurs sous-jacentes » qui, au fur et à mesure des expériences, se révélaient et agissaient sur les choix des sujets (*ibid.*).

---

[130] Parfois même deux fois ou trois fois (Davidson, Suppes, Siegel [1957], p. 68).
[131] Même si le terme n'est pas utilisé par les auteurs.
[132] Sauf si, comme Davidson et *al.* le mentionnent en 1957, la théorie prend place dans une théorie plus large des attitudes et des motivations de l'action.

On voit ici se préciser un argument qui sera utilisé ultérieurement par Davidson, à savoir le fait que la théorie de la décision telle que testée par Davidson, néglige une donnée mentale qui entre en ligne de compte dans les choix des individus : les significations.
Comme Davidson le mentionne lui-même en 1976, cette analyse critique de l'absence de prise en compte de certains déterminants du choix est très proche de celle menée par Tversky l'année précédente.
En effet, dans son article de 1975, Tversky met en évidence deux limites – intimement liées – de la théorie de l'utilité espérée. La première est relative à ce qu'il appelle « l'effet de certitude » (1.3.3.2) et la seconde à un problème d'interprétation inhérent à la théorie (1.3.3.3).

1.3.3.2 L'effet de certitude

Selon Tversky, la théorie de la décision peut être appréhendée au moins de deux manières différentes.
Soit l'on considère que cette théorie est descriptive, et il s'agira alors de vérifier empiriquement si la théorie parvient à fournir un modèle de prise de décision qui décrit fidèlement la manière dont des sujets, placés dans des conditions d'expérimentation, font des choix parmi des paris risqués.
Soit l'on considère la théorie comme ayant un statut normatif dans son essence et il s'agira alors de décrire la manière dont les sujets devraient choisir, s'ils étaient rationnels, parmi des paris risqués.

Choisissant d'adopter dans un premier temps la première posture, Tversky présente toute une série de résultats expérimentaux – récoltés en collaboration avec Kahneman.
Comme il est habituel de le faire, Tversky propose de considérer la théorie de la décision comme un choix entre des paris ou loteries. Les paris consistent essentiellement en des choix parmi des revenus monétaires, revenus soumis à des probabilités (objectives).
L'un des résultats les plus significatifs de l'auteur porte sur les choix observés face à la situation suivante :

Imaginons que les sujets soient face à la situation suivante

Choix I	A = (1000, ½, 0)	B = (400)
Choix II	C = (1000, 1/10, 0)	D = (400, 1/5, 0)

Kahneman et Tversky observent que la plupart des gens préfèrent 400$ pour sûr, soit l'option B plutôt que le pari A (gagner 1000 euros avec une probabilité de ½ ou gagner 0 euro avec la même probabilité) et que la majorité des sujets préfère B à A et C à D.
Pourtant, de tels choix sont incompatibles avec la théorie de l'utilité espérée. Pour en faire la démonstration, il suffit de noter que B est choisi par rapport à A, et en posant que U(0) = 0, on obtient que U (400) > ½ U(1000). D'un autre côté, C est choisi par rapport à D, on obtient 1/10 U (1000) > 1/5 U(400), ce qui entre en contradiction avec le premier choix. D'autant que C peut être exprimé comme (A, 1/5, 0) alors que B peut être exprimé par (B, 1/5, 0). C'est-à-dire que C et D peuvent être représentés comme des paris consistant à obtenir respectivement A et B avec une probabilité de 1/5.
La théorie de l'utilité espérée suggère que si A est préféré à B, tout mélange de A avec l'issue consistant à ne rien obtenir - mélange conditionné à n'importe quelle probabilité – sera toujours préféré à tout mélange de B avec l'issue consistant à ne rien obtenir. C'est ce que qui est appelé la condition de substitution.
Or, en préférant B à A et C à D, les sujets violent cette condition car suivant celle-ci ils choisiraient D à C s'ils avaient choisit B à A.
En fait, selon Tversky, les données révèlent un effet de certitude positive c'est-à-dire une préférence pour les résultats obtenus avec certitude. L'utilité d'un résultat positif apparaît comme étant plus grande quand ce résultat est certain que lorsqu'il est enchâssé à un pari. Le même résultat est découvert pour des paris composés de revenus monétaires négatifs. On parlera alors d'effet de certitude négative.
Mieux, ces phénomènes appelés « effets de certitude » produisent de l'aversion pour le risque pour les paris à valeurs positives et du goût pour le risque pour les paris à valeurs négatives.

Tversky en conclut que ces effets de certitude ne sont pas compatibles avec la construction d'une fonction d'utilité sans une modification au préalable de la théorie. Cette difficulté expérimentale conduit à invalider la conception descriptive de la théorie de l'utilité espérée (Tversky [1975], p.168).

1.3.3.3 Un problème d'interprétation inhérent à la théorie

Si, cette fois, on considère la théorie de l'utilité comme une théorie normative, on suppose qu'elle prescrit les choix que devraient faire un agent plutôt que de les décrire.

D'autres avaient montré avant Tversky que les problèmes liés à cette perspective remettaient en cause soit l'axiomatique de la théorie de l'utilité espérée (Allais, 1953), soit les préférences des individus (Savage, 1954). Telle n'est pas la position de Tversky comme le montre son analyse du contre-exemple de Maurice Allais et la discussion qui en est faite par Savage.

- Allais et la remise en cause de la théorie

Dans l'expérience menée par Allais (1953), les sujets étaient amenés à faire des choix dans les deux situations suivantes :

Situation I : A 1000000 $ pour sûr
B 1000000 $ avec une probabilité de 0.89
5000000 $ avec une probabilité de 0.10
Rien avec une probabilité de 0.01

Situation II : C 1000000 $ avec une probabilité de 0.11
Rien avec une probabilité de 0.89
D 5000000 $ avec une probabilité de 0.10
Rien avec une probabilité de 0.90

Face à ces situations, la plupart des sujets choisissaient A par rapport à B et D par rapport à C. Mais ces choix étaient en contradiction à la théorie de l'utilité espérée. En effet, en préférant A à B et D à C, les sujets violaient l'axiome d'indépendance selon lequel les préférences d'un individu entre deux loteries ne sont pas modifiées lorsque celles-ci sont combinées à une troisième dans des proportions identiques. Savage lui-même reconnaît avoir commis une erreur lors de se rencontre avec Allais où celui-ci lui avait proposé les deux paris[133].

Selon Tversky, le même processus psychologique – que celui décrit dans (a) - est en jeu ici, c'est-à-dire l'effet de certitude. Allais fait simplement usage de sommes d'argent plus importantes et de probabilités plus extrêmes (Tversky [1975], p. 169).

Pour Allais, seuls deux couples de réponses sont compatibles avec l'hypothèse d'utilité espérée, à savoir (AC) et (BD), compatibles car ils assurent l'existence d'une fonction d'utilité.
Cependant, Tversky explique que selon Allais, les caractères parfaitement naturels des choix de A par rapport à B et celui de D par rapport à C impliquent de modifier les axiomes de la théorie plutôt que de changer de sujets.

- Savage et les préférences problématiques

Savage, dans son ouvrage *The Foundations of Statistics* (1954) déjà évoqué dans le chapitre I, a proposé une analyse particulière du contre-exemple de Allais.
L'auteur des *Foundations* propose alors de reformuler l'exemple d'Allais en découpant les options en 100 tickets de loterie numérotés, tickets dont chacun est associé à un lot en millions (*prize*).

[133] Savage [1954], pp.102-103.

		Numéro de ticket		
		1	2 à 11	12 à 100
Situation I	A	1	1	1
	B	0	5	1
Situation II	C	1	1	0
	D	0	5	0

En présentant l'exemple de Allais comme suit, Savage remarque que si l'on tire un ticket dont le numéro est compris entre 12 et 100, l'option A est identique à l'option B et l'option C identique à l'option D.

Si, cette fois, le ticket possède un nombre inférieur à 12, tout sujet qui choisirait A à B, si et seulement si il choisit en même temps C à D. Il est suggéré ici que les sujets n'auront pas de mal à constater que s'ils tirent un ticket dont le numéro est 3 par exemple, l'option A est identique à l'option C et l'option B et identique à l'option D. Dès lors, si le sujet choisit A par rapport à B, c'est qu'il choisit C par rapport à D.

Avec cette présentation du problème, Savage ne contredit plus l'axiome d'indépendance[134].

Cependant là où Allais propose de modifier les axiomes de la théorie de l'utilité espérée, Savage les défend et incrimine au contraire les préférences des individus qui posent problème.

- Tversky et l'interprétation

Pour Tversky, qui n'est d'accord avec aucun d'eux, le problème en jeu ici est un problème d'interprétation des conséquences (Tversky [1975], p. 171).

En effet, on pourrait très bien considérer selon l'auteur que dans l'option B de Savage, le résultat consistant à ne rien obtenir pourrait être interprété et décrit comme le fait de manquer l'occasion (et donc un regret pour le sujet) de gagner un million de dollars pour sûr. Ainsi, en considérant les conséquences non plus comme de simples valeurs

[134] Savage [1954], p. 103.

monétaires mais comme ayant aussi un contenu psychologique pour les sujets, les conclusions de Savage en termes de préférences ne tiennent ; pas plus que la remise en cause d'un axiome précis de la théorie de l'utilité espérée.

Ce qui doit être remis en cause selon Tversky, c'est la conception de l'individu autour de laquelle est construite la théorie de l'utilité espérée et en particulier le fait que cette dernière élude leurs manières d'appréhender, comprendre et interpréter le monde.

Selon Tversky, la théorie de l'utilité espérée néglige l'interprétation que les sujets attribuent aux conséquences. Cet examen attentif de la question du choix sous incertitude révèle donc qu'il n'est guère possible d'évaluer l'adéquation normative des axiomes de la théorie de l'utilité espérée sans spécifier, au sein de la théorie, une interprétation des résultats.

C'est le même constat que fait Davidson lorsqu'il présente dans la revue *Theory and Decision* un modèle de théorie de la décision (1980, 1985) incorporant une analyse des significations : « En testant la théorie de la décision sous incertitude, il est généralement supposé que les sujets comprennent les mots utilisés par l'expérimentateur pour décrire les options qui leur sont proposées. Ou plus exactement, l'expérimentateur suppose qu'il sait comment le sujet comprend ses mots. Etant donné la complexité des paris, l'obscurité de la connexion qui est supposée exister entre les états de la nature et les résultats, et les multiples incompréhensions qui peuvent survenir, l'hypothèse d'une communication parfaite entre le sujet et l'expérimentateur n'est généralement pas satisfaite » (Davidson [1985], p. 87).

Non seulement les modèles de 1957 et 1959 bâtis par Davidson sont inaptes à prendre en compte le temps et l'évolution des préférences, ils semblent, en plus, selon ses propres termes, poser des problèmes d'interprétation.

1.4 La théorie de la décision, une théorie qui fait l'impasse sur les significations

Dans son article « A new basis for decision theory »[135] publié en 1985 dans la revue *Theory and Decision*, Donald Davidson évoque une difficulté récurrente de la théorie expérimentale de la décision : l'expérimentateur considère comme acquise l'analyse des significations ou plus précisément les interprétations que les sujets attribuent aux différents objets de la théorie comme les paris, et les conséquences de ces derniers.

Comme nous venons de le voir, cette critique qu'adresse Davidson à la théorie de la décision trouve écho aux différentes remarques faites par Tversky en 1975 et est présentée de multiples façons dans plusieurs articles de Davidson.

Ainsi, dans l'essai « Expressing Evaluations » (1984), qui traite directement de cette impasse, Davidson décrit-il ce qui pose pour lui problème en ces termes :

> « les théories de la décision bayésienne ont un inconvénient fatal : elles supposent simplement qu'un interprète peut dire quelles propositions un agent évalue ou choisit, ou quelles phrases interprétées expriment les préférences de l'agent […].La théorie de la décision commence avec de simples préférences entre des propositions ; une fois qu'elles ont été identifiées, la théorie nous permet d'extraire les croyances et les désirs qui entrent dans et expliquent les préférences. Mais il n'est rien dit au sujet de ce qui détermine les objets des simples préférences originales. Les préférences sont, évidemment, manifestées dans le comportement de plusieurs manières. Mais cela ne nous dit pas comment le contenu de la préférence est fixé » (Davidson [1984, 2004], p.28).

Cette citation permet de montrer que l'impasse sur les significations pose en réalité au moins deux types de problèmes logiquement liés :

En faisant l'impasse sur les significations, la théorie de la décision néglige d'abord une partie essentielle des données mentales qui poussent un agent à faire un choix ou à prendre une décision. En négligeant (ou en considérant comme acquise) l'analyse des significations, la théorie de la décision se borne à une analyse behavioriste alors que le langage relie l'agent aux autres et au monde qui l'entoure (1.4.1).

Le deuxième problème, corrélé au premier, peut se résumer ainsi dans les termes de l'auteur : négliger l'analyse des significations c'est ne pas voir qu' « établir qu'une attribution de croyance ou de désir est correcte pose à peu près les mêmes problèmes

[135] Cet article est une version remaniée de Davidson [1980].

que montrer que nous avons compris les paroles de quelqu'un d'autre » (Davidson [1974, 1993a], p. 316). Plus précisément, la théorie de la décision et la théorie de l'interprétation constituent toutes deux des mesures du mental interdépendantes (du fait de l'interdépendance des contenus mentaux), mesures qui utilisent des méthodes similaires. Ainsi, en négligeant cette théorie de l'interprétation, la théorie de la décision se passe d'une mesure supplémentaire du mental qui permettrait pourtant de disposer d'une image enrichie de celui-ci (1.4.2).

1.4.1. La théorie de la décision comme théorie behavioriste du mental

Comme nous l'avions mentionné dans le chapitre 2 de la première partie, la théorie de la décision, du moins dans la version de celle-ci proposée par Savage, adopte clairement une vision behavioriste de l'individu car, plutôt que de demander aux gens la probabilité qu'ils attribuent aux événements, Savage préfère la déduire de leurs choix dans diverses circonstances.

De même la solution proposée par Davidson, Suppes et Siegel au problème de la mesure séparée de la loi de probabilité subjective et de l'utilité est une solution behavioriste (Davidson, Suppes, Siegel [1957], p. 12) comme Davidson l'explique lui-même : « Tout ce que nous avions à faire était de donner une interprétation behavioriste claire de 'X préfère A à B' » (Davidson [1976, 1993a], p. 358).

Or l'une des raisons pour lesquelles Davidson critique la théorie de la décision est que la solution proposée à la plupart de ses modèles est behavioriste et néglige une partie des données mentales qui déterminent et expliquent le comportement de choix.

Le rejet de Davidson du behaviorisme s'explique par le fait qu'il constitue un programme réductionniste c'est-à-dire une tentative de réduire les états mentaux à des manifestations purement physiques[136]. Nous avons montré que le monisme anomal de Davidson est en opposition avec une telle posture.

Pour Davidson, le behaviorisme pose d'abord problème dès lors qu'il considère que les états mentaux ne sont rien d'autre que ce que nous considérons comme preuves pour ces derniers. Autrement dit, le behaviorisme considère, selon Davidson, que les états

[136] On retrouve ici la thèse du monisme anomal.

mentaux se réduisent à leur manifestation phénoménale et rien d'autre. Ces états mentaux peuvent être définis explicitement qu'en termes comportementaux. Le problème est que le behaviorisme ne réussit pas à expliquer le fait que nous n'ayons pas besoin de preuves lorsque nous nous attribuons des attitudes à nous-mêmes. L'argumentation de Davidson se présente, comme c'est souvent le cas, sous la forme d'un raisonnement par l'absurde :

> « Supposez que nous essayions de dire, sans utiliser de concepts mentaux, en quoi consiste pour un individu le faire de croire qu'il y a de la vie sur Mars. On pourrait supposer ceci : quand un certain son est produit en présence de l'individu (« Y a-t-il de la vie sur Mars ? ») il en produit un autre (« Oui »). Mais bien entendu, cela ne montre qu'il croit qu'il y a de la vie sur Mars que s'il comprend le français, que l'émission du son produit soit intentionnelle, et qu'elle soit une réponse à des sons reconnus comme signifiant quelque chose en français, et ainsi de suite. A chaque fois qu'il manque quelque chose pour parvenir à la définition, il faut ajouter une clause restrictive. Et pourtant quelle que soit la manière dont nous rafistolons et adaptons les conditions mentales, nous avons toujours besoin d'une condition additionnelle […] qui a un caractère mental » (Davidson [1970b, 1993a], p. 291).

Autrement dit, il semble difficile de construire une théorie strictement physique des états mentaux : toute attribution d'attitude mentale comme une préférence ou une croyance nécessite de faire appel à des états mentaux et donc à des éléments qui ne se manifestent pas directement lors d'observations.

1.4.2. Théorie de la décision et théorie de l'interprétation du langage : deux mesures du mental.

En choisissant de renoncer au behaviorisme, il faut disposer d'autres données que celles issues de l'observation expérimentale du comportement. Ces données pourraient être accessibles par le langage et l'une des méthodes permettant d'y accéder consisterait à construire une théorie de l'interprétation de celui-ci et cette méthode – inspirée de Quine, va consister à séparer le rôle des croyances et des significations, de même qu'en

théorie de la décision on sépare les rôles des désirs et des croyances (1.4.1.1). Or, nous avions montré dans la première partie de la thèse que la méthode utilisée en théorie de la décision est celle de Ramsey et qu'elle constitue une mesure des désirs et des croyances. Pour Davidson, la méthode proposée par Quine visant à utiliser le principe de Charité pour fixer les croyances et déterminer les significations, va constituer non seulement une autre méthode mais aussi une autre mesure du mental.
Il s'agira à la fois de présenter à la fois l'analogie constante que défend Davidson entre les mesures d'éléments physiques comme le poids ou la température et les mesures de données mentales (1.4.1.2).

1.4.2.1. Théorie de la décision et théorie de l'interprétation, deux problèmes distincts ?

La théorie de la décision et la théorie de l'interprétation cherchent à résoudre deux problèmes, séparer le rôle des croyances et des désirs pour la première, séparer celui des croyances et des significations pour la deuxième.
Comme nous l'avons mentionné, Davidson remarque que la théorie de la décision présuppose que l'on peut identifier et individuer les propositions orientées vers les attitudes propositionnelles comme le désir et la croyance. Cependant, notre aptitude à identifier les propositions qui sous-tendent les attitudes qu'un agent nourrit n'est pas à séparer, selon l'auteur de notre aptitude à comprendre ce qu'il dit (Davidson [1980, 2004], p. 155). Davidson ajoute qu'on découvre généralement ce que quelqu'un veut, préfère ou croit, seulement en interprétant ces propos (Davidson [1990], p. 318). Pour l'auteur, il n'est pas plus facile d'établir qu'une attribution de désir est correcte que d'interpréter le discours de quelqu'un mais il faut aller plus loin et dire que les deux problèmes sont identiques (Davidson [1974, 1993a], p. 318). En général on ne peut pas déterminer des croyances sans maîtriser le langage de l'individu auquel on les attribue ; et on ne peut maîtriser le langage de quelqu'un sans savoir ce qu'il croit : « Pour interpréter le comportement verbal, nous devons être capables de dire quand un locuteur tient une phrase qu'il énonce pour vraie. Mais on tient des phrases pour vraies en partie en raison de ce que l'on croit, et en partie en raison de ce que l'on veut dire en énonçant des mots. Le problème de l'interprétation est donc celui de savoir comment nous

pouvons séparer simultanément les rôles de la croyance et de la signification de la structure des phrases auxquelles un locuteur souscrit au cours d'une certaine période. La situation est semblable à celle de la théorie de la décision : tout comme nous ne pouvons inférer les croyances des choix sans aussi inférer les désirs, nous ne pouvons pas établir ce que quelqu'un veut dire par ce qu'il dit sans en même temps construire une théorie de ce qu'il croit » (Davidson [1974, 1993a], p. 318).

1.4.2.2. Mesures physiques, mesures psychologiques

Notre présentation jointe d'un problème de décision et d'un problème d'interprétation a pour objectif de faire apparaître les deux mesures sous-jacentes des contenus mentaux qu'ils impliquent respectivement.

Après avoir expliqué que chez Davidson, il existe une analogie constante entre ces mesures des contenus mentaux et les mesures physiques, nous attardons sur les conditions nécessaires à l'obtention d'une mesure d'un contenu mental puis distinguons la mesure du mental de la théorie de la décision de celle de la théorie de l'interprétation.

- Une analogie entre mesures physiques et mesures des contenus mentaux

L'idée de mesure parcoure la réflexion de Davidson de part et part. A maintes reprises, l'auteur évoque les différents types de mesures en jeu dans les sciences physiques comme la mesure de la masse ou de la longueur. On peut remarquer que Davidson fait souvent l'analogie entre mesure en physique et mesure des contenus mentaux. Non pas pour dire que mesurer un désir revient à mesurer une longueur ou une température mais que la mesure d'un désir n'est pas déterminée de manière unique tout comme la mesure de la longueur n'est pas unique puisqu'on peut mesurer celle-ci par le pied ou le mètre par exemple.

Les travaux de Frank Ramsey qui, comme on l'a vu dans le chapitre I, constituent le point d'ancrage des recherches de Davidson en économie expérimentale, faisaient déjà état d'une attention particulière pour la mesure. Plus précisément, dans son article « Vérité et Probabilité » de 1926, Ramsey soutenait que le degré de croyance était comme un intervalle de temps (Ramsey [1926, 2003], p. 162).

Son idée était de dire que toute mesure, qu'elle soit dans le domaine physique ou dans un autre domaine, faisait nécessairement face à un double problème (*ibid.*). Le premier problème est relatif à la possibilité de trouver une mesure suffisamment exacte. Ce problème est ravivé du fait notamment du grand nombre potentiel de mesures permettant de mesurer la même chose et à la nécessité de spécifier, à chaque fois les correspondances entre les différentes mesures.
Le second problème est précisément celui des divergences qui peuvent survenir entre les différentes mesures. Autrement dit, comment être sûr par exemple, qu'une différence de valeur identifiée avec une certaine mesure sera équivalence pour une autre mesure relative au même objet.

Ramsey suggérait de combler ces deux « faiblesses » de la mesure par un ensemble de conditions formelles à satisfaire (*ibid.*). Ainsi proposa t-il, concernant les croyances,
- que le système de mesure rende possible le fait d'assigner à une croyance une position et donc une grandeur donnée sur une échelle d'ordre ;
- que deux croyances de mêmes degrés soient placées au même rang sur l'échelle.
Pour donner sens à cette échelle, il fallait ajouter ce qu'il appelle un peu de « fiction » (*ibid.*) de manière à rendre la mesure acceptable et compréhensible : il introduisit la méthode du pari pour calibrer l'échelle.

Cette analogie avec la mesure en physique sera mentionnée à plusieurs reprises par Davidson[137]. Ce dernier utilisera plus précisément une argumentation similaire à celle de Ramsey : « D'un point de vue formel, la situation est analogue à celle de la mesure fondamentale en physique, par exemple la mesure de la longueur, de la température ou de la masse. Assigner des nombres afin de mesurer ces quantités présuppose l'existence d'un ensemble de conditions très strictes » (Davidson [1974, 1993a], p.315). On retrouve la même idée de « conditions strictes » permettant de s'assurer que la mesure va avoir un sens et qu'elle pourrait être utilisée dans le cadre d'une théorie plus complexe. Davidson poursuit : « je pense que nous pouvons considérer les deux cas comme parallèles du point de vue suivant. Tout comme on peut considérer la

[137] Voir par exemple Davidson [1974, 1993a], [1989, 2001], [1997a, 2001], et [1997b, 2001].

satisfaction des conditions de mesure de la longueur ou de la masse comme étant constitutive du domaine d'application des sciences qui emploient ces mesures, on peut considérer la satisfaction des conditions de non contradiction et de cohérence rationnelle comme constitutive du domaine d'application de concepts comme ceux de croyance, de désir, d'intention, et d'action » (*ibid.*). Autrement dit, les critères nous permettant de dire lorsqu'une mesure physique est acceptable sont similaires à ceux utilisés en théorie de la décision lorsqu'il s'agit de décrire la rationalité des attitudes d'un individu.

Il existe donc pour Davidson, comme chez Ramsey, des conditions de rationalité nécessaires pour donner un sens à leur mesure, autrement dit il faut imposer une structure rationnelle (représentée sous forme d'axiomes), il faut rationaliser le processus de décision pour que la mesure prenne sens.

- Les conditions de rationalité imposées par Davidson

Comme on l'a vu, l'un des intérêts majeurs de la théorie de la décision est, selon Davidson, la structure de rationalisation qu'elle propose. Autrement dit, la théorie présente les différents états de l'individu qui l'ont amené, à la suite d'une délibération, à faire tel choix. Et la structure proposée combine ces états de manière à fournir une raison qui est la cause de son action, d'où l'analogie entre théorie de l'action et théorie de la décision évoquée plus haut.

Ces conditions rationnelles permettent à Davidson de définir l'irrationalité d'autres décisions ou actions. Ainsi concernant l'irrationalité associée aux préférences intransitives (qui vont donc à l'encontre des conditions de transitivité imposée par la structure axiomatique), Davidson écrit-il :

> « il n'est pas facile de décrire, avec suffisamment de détails, une expérience qui nous convaincrait que la relation *est plus lourd que* n'est pas transitive. Bien que le cas de la relation de préférence ne soit pas aussi extrême, je ne pense pas que nous puissions énoncer clairement à quelles conditions nous pourrions nous convaincre qu'un individu a, à un moment donné (et sans changement d'opinion) préféré *a* à *b*, *b* à *c* et *c* à *a*. Cela nous semble difficile parce que nous ne parvenons pas à attribuer de façon sensée une préférence autrement que par rapport à un arrière-plan d'attitudes cohérentes » (Davidson [1974, 1993a], pp.315-316).

Cette référence à des préférences est, là encore, due à Ramsey et à la pompe à finance évoquée dans le chapitre I. Mais l'argument de rationalité utilisé par Davidson va plus loin que celui de Ramsey. Il ne s'agit pas simplement de dire qu'un joueur malin pourrait proposer une suite de paris à l'individu qui entretient des préférences intransitives. Davidson ajoute que nous ne pouvons pas faire sens d'une quelconque irrationalité sans postuler que l'individu est fondamentalement rationnel. La raison pour cela est que l'erreur, conçue de manière générale, n'a de sens que par rapport à une norme. Ainsi par exemple, « si nous pouvons comprendre ce qui rend l'erreur possible, nous pouvons dès lors voir comment, étant donnée l'existence de la pensée, il se peut que plusieurs de nos croyances soient vraies et justifiées, et constituent alors la connaissance » (Davidson [1995, 2004], p.4). Autrement dit, l'intransitivité n'a de sens que par rapport à une norme de rationalité que serait la transitivité. Dans l'exemple de Ramsey, ce n'est que parce que l'agent rationnel maximise ses gains et minimisent ses pertes qu'il est considéré comme irrationnel s'il entretient des préférences intransitives qui le conduisent à être ruiné.

Les conditions d'adéquation d'une mesure en théorie de la décision imposent une structure rationnelle à l'échelle des degrés de croyance par exemple car on ne pourrait faire sens d'une échelle qui ne remplirait pas ces conditions.

- Les mesures du mental en théorie de la décision et en théorie de l'interprétation

Parmi ces mesures du mental proposée par Davidson, on peut distinguer celle de la théorie de l'interprétation qui est une mesure de la valeur de vérité d'une phrase et celle de la théorie de la décision qui est la mesure des désirs et des croyances. La différence formelle qui saute aux yeux immédiatement entre ces deux mesures est le caractère quantifiée, numérique de la seconde qui n'existe pas dans la première. Cela implique t-il pour autant que la première n'est pas contrainte par une structure rationnelle comme la première ? Ou encore qu'elle est moins objective que la seconde ?

i) Des fondements objectifs de la mesure de la théorie de l'interprétation

Fonder rationnellement les mesures des attitudes des individus permet de les rendre objectives selon Davidson car le concept d'objectivité est lié au concept de vérité. Intuitivement, le faux n'a de sens que par rapport au vrai.

En théorie de la décision, l'analogie est la suivante : l'objectivité – au sens de Davidson – est relative à la possibilité de l'erreur dans les préférences par exemple, et cette erreur peut être décrite par exemple par des préférences intransitives du type $A \succ B$, $B \succ C$ et $C \succ A$ qui constituent une erreur dans la mesure où elles contredisent les canons de la rationalité qui imposent la rationalité et la cohérence. Il s'agit donc essentiellement d'une erreur dans la mesure où l'on s'écarte de la norme mais non d'une erreur au sens où les individus sont dans l'erreur lorsqu'ils ont des préférences intransitives.

L'argument de l'objectivité est aussi utilisé lorsque Davidson analyse l'interprétation du langage. Pour l'auteur, l'existence de la pensée et de la communication – et donc de toutes les attitudes évaluatives – est le fait que deux créatures ou plus réagissent au monde extérieur et se répondent mutuellement (Davidson [1997, 2001], p. 83). Ainsi lorsqu'un interprète tente de comprendre un locuteur, il peut être amené à comparer ses propres états mentaux à ceux du locuteur le conduisant à un processus itératif entre ses valeurs et celles du locuteur de manière à constituer un ensemble de significations le permettant de comprendre l'autre. Cette procédure va le conduire, comme on va le voir dans le chapitre 2 à une certaine mesure des contenus mentaux du locuteur et cette mesure n'est, selon Davidson pas moins objective que la mesure en théorie de la décision car elle est le fruit de l'intersubjectivité, c'est-à-dire d'un certain accord et cet accord est de même valeur que celui qui nous mène à utiliser un même critère de rationalité.

On peut donc dire, selon Davidson, que la mesure en théorie de la décision et la mesure en théorie de l'interprétation sont fondées objectivement.

ii) Une mesure de la valeur de vérité des phrases soumises à des conditions de rationalité

Dans le cas du choix rationnel ainsi que dans celui des énonciations, il faut, selon Davidson, nécessairement avoir recours à une structure rationnelle : « Ce que je veux dire est que si nos attributions d'attitudes et de croyances sont sensées, ou que si nous

voulons pouvoir utilement décrire des mouvements comme du comportement, nous sommes obligés de déceler, dans la structure du comportement, de la croyance et du désir, une large mesure de rationalité et de cohérence » (Davidson [1974, 1993a], p.316).

De la même manière que l'expérimentateur ou le théoricien postule une trame rationnelle de préférences, un interprète lorsqu'il tente de comprendre un locuteur dont il ne connaît pas nécessairement le langage, doit considérer qu'il a des croyances « cohérentes et correctes selon [ses] propres critères » (Davidson [1974, 1993a], p.318). Comme on le verra dans le chapitre 3, cette position théorique renvoie directement au principe de Charité à l'œuvre dans toute interprétation.

Enfin, on peut noter qu'ici aussi, Davidson semble appliquer le monisme anomal. Ainsi, la mesure de facteurs psychologiques est une mesure physique dans la mesure où elle utilise notamment la classe des nombres (dans le cas de la théorie de la décision par exemple) mais la multiplicité des transformations affines de la fonction d'utilité par exemple, met en évidence une sorte de non réductionnisme du psychologique au physique, une idée finalement de non transposition à l'identique. Ainsi la mesure en théorie de la décision est une mesure physique mais elle ne s'y réduit pas car il n'est pas possible d'établir des lois en théorie de la décision qui aient la même valeur que des lois physiques et c'est la raison pour laquelle Davidson dit que la théorie de la décision est de toute évidence fausse (Davidson [1976, 1993a], p.355).

CONCLUSION

Nous avons débuté ce chapitre en présentant l'analogie défendue par Davidson entre théorie de la décision et théorie de l'action ainsi que ses limites. Le caractère quantifiable et sophistiqué de la théorie de la décision incitait à un certain optimisme quant à l'explication et la description de l'action par des lois. Seulement, la théorie de la décision n'y parvient pas plus que la théorie de l'action et produit essentiellement des

régularités statistiques qui ne peuvent s'ériger au rang de lois comme dans les sciences de la nature.

En allant plus loin dans l'analyse des critiques qu'adresse Davidson à la théorie de la décision, nous avons pu identifier trois problèmes majeurs selon lui.

Premièrement, l'absence d'analyse des conflits de désirs. C'est ici le caractère statique de la théorie de la décision qui est critiqué par l'auteur.

Puis, en présentant les expériences menées en collaboration avec Jacob Marschak en 1959, nous avons pu mettre en évidence deux éléments négligés, selon Davidson, par la théorie standard de la décision : l'effet de certitude et le problème de l'interprétation.

Enfin, la dernière critique adressée de l'auteur correspond à l'impasse faite sur les significations. Cette critique est centrale puisque c'est à partir de celle-ci que Davidson va construire une nouvelle théorie de la décision intégrant cette dimension à l'analyse.

Chapitre 2 : Définition de la théorie unifiée

Les critiques adressées par Davidson lui-même à partir des années 1970 à la théorie de la décision des années 1950 le conduisent dans les années 1980 à proposer une nouvelle théorie.

Le modèle initial de 1957 tel que décrit dans la première partie est alors enrichi et transformé au cours de multiples articles écrits dans les années 1970-80. Pour ce faire, Davidson inscrit ce modèle enrichi au sein d'une théorie plus large qu'il qualifie d' « unifiée » (1980). Ce qualificatif se justifie par la nature de l'objet de la théorie, un triplet (désir, croyance, signification) embrassé dans un même raisonnement théorique mais aussi par la démarche adoptée pour l'analyser puisque Davidson établit entre ces entités des causalités complexes qui suggèrent que l'un ne peut être déterminé sans que les deux autres ne le soient aussi.

Après avoir montré comment Davidson fut influencé par Ramsey pour construire une théorie unifiée (2.1), et puisque chez Davidson, il n'existe aucun ouvrage consacré à cette unification, nous chercherons à reconstituer cette dernière (son objet (a), sa forme (b) et ses implications (c)) au travers de ces différents travaux (2.2). Dans la mesure où l'on ne peut étudier les désirs sans faire appel aux significations, nous montrerons que cette théorie s'inscrit en outre dans un cadre épistémologique particulier, au croisement d'une économie et d'une philosophie rendues co-dépendantes et par-là redéfinies (2.3).

Le cadre théorique et méthodologique ainsi décrit nous permettra finalement de présenter l'axiomatique du modèle pour en estimer l'apport pour la théorie de la décision économique.

2.1. Sur une idée de Ramsey

Selon Davidson, Ramsey cherchait « avant tout à fournir une assise dans le comportement à l'idée qu'une personne accorde un certain degré de créance à une proposition » (Davidson [1974, 1993a], pp. 312-313). Ce faisant, comme nous l'avons montré dans le chapitre 3 de la première partie, la théorie de Ramsey se positionne avant tout en réaction au *Traité des Probabilités* (1921) de J.M. Keynes.

Si l'idée de traiter dans une même théorie des croyances, désirs et significations (en utilisant des propositions) est empruntée par Davidson à Ramsey (1926) (2.1.2), elle prend donc toutefois sa source dans le travail de Keynes. C'est pourquoi nous revenons sur le travail originel de Keynes en matière de propositions - servant de point d'ancrage à la position de Ramsey - (2.1.1) afin de montrer comment leurs utilisations par Ramsey puis par Davidson se sont modifiées pour servir des objectifs divergents (2.1.3).

2.1.1. La conception des probabilités de Keynes

Lorsque Frank Ramsey publia son essai *Truth and Probability* en 1926, deux théories des probabilités occupaient le devant de la scène à Cambridge : celle de John Venn (1834-1923) avec l'ouvrage *The Logic of Chance* de 1866 ; et le *Traité des Probabilités* (1921) de John Maynard Keynes (1883-1946).

L'objectif de Venn était de traiter des fondements philosophiques de la probabilité. Dans *The Logic of Chance*, les probabilités étaient conçues comme une branche de la logique, sans pourtant que l'accent ne soit mis sur l'aspect mathématique de celles-ci, comme le souligne Maria Carla Galavotti [2005]. Venn voulait en effet avant tout relier les probabilités à notre connaissance empirique et plus précisément aux faits. Selon Venn, en considérant une collection d'événements, nous avions la possibilité d'observer des similarités et de détecter des répétitions (Galavotti [2005], p.76). Ces répétitions permettaient de mettre en évidence des fréquences sur lesquelles les lois de la probabilité prenaient précisément appui. C'est pourquoi la théorie de Venn fut caractérisée de théorie fréquentiste des probabilités.

Venn semble avoir une conception opposée à celle que Keynes et Ramsey adopteront. Pour ces deux derniers, la probabilité est une mesure de la croyance. Or pour Venn, il est trop difficile d'assimiler la croyance à une probabilité car la première dépend de nombreux facteurs et du contexte dans lequel elle se forme, ce qui la rend inappropriée à la construction objective de la probabilité (Galavotti [2005], p. 78).
Ce n'est, selon Venn, qu'une fois que la probabilité – comme fréquence de long court – est fondée qu'elle peut justifier une croyance.

Cette conception des probabilités sera remise en cause près de quarante ans après les premiers travaux de Venn par Keynes dans son *Traité des Probabilités*. Cet ouvrage, initialement rédigé mais non publié entre 1906 et 1911, puis remanié, après la guerre, en 1920 avant sa publication en 1921, est le fruit de multiples influences. Comme le rapporte Donald Gillies ([2000], p.26), Keynes mentionne lui-même l'influence qu'ont eue sur lui les ouvrages *The Principles of Mathematics* (1903) de Bertand Russell et les *Principia Ethica* de G.E. Moore.
L'ouvrage de Russell, par exemple, était un plaidoyer en faveur du programme logiciste appliqué aux fondements des mathématiques : il s'agissait de réduire les mathématiques à la seule logique, dans la lignée des travaux de Frege (1879). Cela consistait plus précisément à retranscrire tous les théorèmes mathématiques au sein d'un système formel déductif dont les axiomes étaient des vérités logiques. Cette priorité donnée à la logique semble avoir fortement influencé Keynes dans sa conception des probabilités[138], tout particulièrement le passage analytique, au sein du raisonnement, des preuves aux hypothèses.
L'influence de Moore est, quant à elle, relative au problème de savoir comment agir de telle sorte à ce que son action produise de bons résultats si notre vision de l'avenir est frappée par l'incertitude. Là où Moore ne voit sans doute pas le problème et s'appuie sur un concept de probabilité qui n'est pas compatible avec l'incertitude, Keynes va tenter dans son *Traité*, d'élargir le concept de probabilité de Moore a des cas plus délicats[139].

[138] Il y a en réalité plus qu'une influence mais une connexion forte entre une conception de la logique et une conception des probabilités chez Keynes.
[139] Sur ce point voir Dostaller [2005], p.42.

Enfin, une autre influence essentielle dans le *Traité des Probabilités* de Keynes est celle de David Hume (1711-1776). On pourrait voir, par exemple, dans le travail de Keynes une tentative de renouer avec la distinction de Hume entre preuves et probabilités[140] à l'aide d'une conception singulière de la causalité ; et cette référence humienne pourrait éclairer l'analyse que propose Keynes du passage d'une induction pure à la certitude, et l'effet de ce passage sur la relation de probabilité P(*a/h*).

En effet, selon Keynes, la probabilité est une relation objective entre propositions. Et la probabilité d'une proposition est relative aux données empiriques - qui sont elles mêmes des propositions - qui nous permettent ou nous donnent des raisons d'affirmer celle-ci. La probabilité de *a* si *h* est une relation dont le degré ou l'intensité est comprise entre 0 et 1 et P(*a/h*) = 1 si et seulement si *a* est la conséquence logique de *h*.

Le terme « probable » décrit plus précisément une certaine connaissance dont nous disposons, à la faveur de la vérité ou de la fausseté d'une proposition. Cette connaissance est décrite comme un ensemble de propositions qui constituent les prémisses d'un raisonnement logique menant à une conclusion. Ces prémisses sont donc l'argument qui permet d'attester la vérité d'une proposition logiquement liée à elles, une conclusion (Galavotti [2005], p. 146). Autrement dit, Keynes suppose qu'entre deux propositions quelconques, prises l'une comme prémisse et l'autre comme conclusion, il n'y a qu'une et une seule relation d'un certain type appelée relation de probabilité et que, dans n'importe quel cas donné, si la relation est de degré α de la croyance totale en la vérité de la prémisse, nous devrions, si nous étions rationnels, passer à une croyance de degré α en la vérité de la conclusion.

C'est notamment sur ce point que va porter la critique de Ramsey.

[140] Pour une analyse détaillée de ce point voir Lapidus [2000].

2.1.2. Les critiques de Ramsey

Ramsey adresse deux critiques à Keynes, critiques qui lui permettront de mettre en évidence sa propre théorie.

Ramsey considère en premier lieu qu'il n'est pas toujours possible d'avoir une idée de la relation de probabilité entre deux propositions comme le propose Keynes, par exemple entre les deux propositions « ceci est rouge » et « ceci est bleu »[141]. Dans ce cas précis, on peut considérer que ces deux propositions entretiennent des relations logiques comme la relation d'identité (de forme ou de prédicat) mais la « simple contemplation » (i*bid.*) de ces propositions ne permet pas selon Ramsey de discerner une relation de probabilité entre elles. Il serait difficile pour une personne de pouvoir lier ou faire une analogie entre ce degré de probabilité apparent et un degré de croyance actuel ou hypothétique (Ramsey [1926, 2003], p.158).

Plus précisément, Ramsey donne une ébauche du raisonnement que ferait une personne si on lui demandait quelle probabilité donne l'une de ces propositions par rapport à l'autre (*ibid.*). La réponse de l'auteur est que nous considérons généralement ce que nous connaissons – nos degrés de croyance actuels ou hypothétiques, tels que tirés de notre expérience ou tirés d'une extension de celle-ci à un cas similaire - et que lorsque l'on tente de se représenter un degré de probabilité, on a tendance à imaginer ce qu'un homme sage penserait face à ces deux propositions. Ainsi, il est plus facile, selon l'auteur, d'établir des estimations de probabilités dans notre vie quotidienne que dans un cadre logique pur comme dans le cas des deux propositions mentionnées plus haut.

La seconde critique de Ramsey, liée à la première, est que l'introspection ne nous permet pas de percevoir des relations de probabilité. On peut supposer nos croyances vraies et attribuer un degré de croyance à une hypothèse, mais nous n'avons pas pour autant accès à une mesure de ce degré. Or, Keynes suppose, selon Ramsey, que la relation de probabilité et la relation de degré de croyance peuvent être exprimées par des nombres et que le nombre exprimant la première relation est le même que celui qui exprime ou mesure le degré de croyance approprié.

---

[141] Je reprends ici la formulation de Dokic et Engel [2001]. La traduction de Ramsey dans l'ouvrage *Logique, Philosophie et Probabilités* [2003] qui reprend les différents articles publiés par R.B.Braithwaite en 1931, mentionne plutôt « *a* est rouge » et « *b* est rouge » (Ramsey [1926, 2003], p. 158).

2.1.3. Degrés de croyance et probabilités subjectives chez Ramsey

Les critiques adressées par Ramsey à la position de Keynes quant aux probabilités des propositions l'ont conduit à adopter une autre position dont l'essentiel est décrit dans son essai *Truth and Probability* (1926). Ramsey présente son enquête comme une étude de la logique de la croyance partielle. Il tente notamment dans son essai de construire une doctrine des relations entre probabilité personnelle et action. C'est cette conception qui sera reprise et axiomatisée par Savage en 1954.

Selon Ramsey, si l'on veut assigner correctement des probabilités à nos croyances, on doit être capable de mesurer celles-ci. Or certaines de ces croyances se mesurent plus aisément que d'autres. Leur mesure est un processus ambigu qui mène à une réponse variable en fonction de la manière dont la mesure est conduite – il en est de même avec la théorie physique (Ramsey [1926, 2003], p.161).

Mesurer une croyance implique donc selon lui la construction d'un système satisfaisant qui assignerait à chaque croyance une magnitude ou un degré ayant une position définie sur un ordre de magnitude. La construction de ce système de mesure revêt une importance toute particulière puisqu'il déterminera largement la signification donnée aux croyances - la mesure engendre la signification. Ce système est chez Ramsey l'objet d'une description minutieuse par étape.

- Etape 1 : la croyance n'est pas un sentiment

Il ne s'agit pas simplement de construire des séries ordonnées de degrés mais aussi d'assigner des nombres à ces degrés d'une manière intelligible : par exemple, le degré auquel je crois en telle proposition se situe aux deux tiers de la certitude.

La première étape de ce processus de mesure consiste à reconsidérer la notion de croyance.

On pourrait, sous une première acception, supposer que le degré de croyance est quelque chose de perceptible par celui ou celle qui le ressent (Ramsey [1926, 2003], p .163). Les croyances se différencieraient en fonction de l'intensité du sentiment qui les accompagne. On pourrait appeler, selon Ramsey, ce sentiment, un sentiment de

croyance ou un sentiment de conviction, et le degré de croyance renverrait à l'intensité de ce sentiment.

L'inconvénient lié à cette conception de la croyance réside dans la difficulté d'attribuer des nombres à des intensités de sentiments. D'autant que dans notre vie quotidienne, les croyances les plus sûres ne sont pratiquement pas accompagnées de sentiments. Le terme de « sentiment » renvoie chez Ramsey à un « sentiment-de-croyance » ou « sentiment de conviction ». Sous ce prisme, le degré de croyance serait une mesure de l'intensité du sentiment, ce que se refuse à considérer Ramsey.

- Etape 2 : La croyance comme attitude propositionnelle

La seconde manière de concevoir la croyance, celle qui est effectivement retenue par Ramsey, consiste à considérer le degré de croyance comme une propriété causale de la croyance, c'est à dire la disposition à agir qu'ouvre la croyance. Ramsey reprend ici la définition d'une croyance comme « attitude propositionnelle » selon l'expression de Russell [1926].

Pour Ramsey, qui suit Russell, une croyance peut ne pas mener effectivement à une action. Mais elle peut effectivement mener à l'action à la suite de certaines circonstances, tout comme un morceau d'arsenic est appelé toxique non pas parce qu'il a effectivement tué ou tuera quelqu'un mais parce qu'il tuera quelqu'un s'il est absorbé.

Pour construire sa mesure des degrés de croyance, Ramsey s'intéresse donc uniquement aux croyances dispositionnelles[142] plutôt qu'à des croyances du type « la terre est ronde », son étude porte sur les croyances auxquelles on pense rarement mais qui vont guider les actions de l'individu dans le cas où ce sera pertinent. Dans la suite du raisonnement, il sera donc uniquement question de ces croyances comme attitudes propositionnelles.

- Etape 3 : La mesure d'une attitude propositionnelle

[142] La théorie de la croyance comme disposition à l'action fut principalement proposée par Alexander Bain et reprise par Peirce. C'est en lisant ce dernier que Ramsey se familiarisa avec cette thèse, qui forme l'une des bases du pragmatisme (Dokic et Engel [2001], p.17).

Selon Ramsey, la méthode la plus habituelle, que l'on a coutume d'utiliser quand il s'agit de mesurer les croyances d'une personne, consiste à lui proposer une mise et observer la plus basse cote que la personne accepte.
Ramsey souligne toutefois l'inexactitude de cette mesure en raison de l'utilité décroissante de la monnaie mais aussi du fait que la personne considérée peut avoir une certaine « avidité » ou une « répugnance » particulière pour le fait de miser. La difficulté réside donc dans la séparation des forces qui agissent ensemble : la croyance, l'aversion ou non pour le risque, et l'utilité décroissante de la monnaie.
Devant les difficultés liées à l'isolation de ces différentes « forces », Ramsey propose une théorie de la croyance plus générale, basée sur une théorie psychologique générale : nous agissons dans une certaine voie car nous pensons qu'elle est la plus probable ou plausible pour réaliser les objectifs de nos désirs. Ainsi, pour Ramsey, les actions d'une personne seraient complètement déterminées par ses désirs et ses opinions. Les individus agissent de manière à maximiser leur utilité espérée.

Ramsey considère la croyance comme une « attitude propositionnelle » c'est-à-dire impliquant une relation à une proposition de la forme « X croit que p ». Une attitude propositionnelle implique une certaine attitude par rapport à une proposition et plus précisément par rapport à la vérité de celle-ci. Appréhender la croyance de cette manière revient, comme on l'a vu dans le paragraphe précédent, à la concevoir comme une disposition à agir. Ainsi comme le souligne Pascal Engel lorsqu'il s'interroge sur le rôle de la croyance dans l'explication de l'action, « croire que p, c'est être disposé à parier sur la vérité de p » (Engel [1998], p.330). Autrement dit, croire que p c'est être disposé à agir de telle sorte à ce que p soit effectivement vrai mais jusqu'à quel point ? Mesurer la croyance revient à évaluer soit la distance que l'individu est prêt à parcourir s'il croit que la route sur laquelle il se trouve est la bonne route lorsque le pari est d'ordre pratique, soit la quantité d'argent que l'individu est prêt à miser sur la vérité de p. Ce que cherche à mesurer Ramsey c'est le degré de croyance en la vérité d'une proposition. Comme on l'a vu dans le chapitre II de la partie 1, Ramsey a proposé une méthode opérationnelle permettant de mesurer le degré de croyance d'un agent vis-à-vis d'une proposition arbitraire *p*, une méthode qui permette de déceler ou d'obtenir l'évaluation du sujet. Mais ce que cherche à définir Ramsey, ce n'est pas la croyance « pleine »

(Ramsey [1929, 2003], p. 189) dont les valeurs seraient 0 et 1 mais la croyance partielle de degré 2/3 par exemple. Comme nous l'avions vu dans la partie 1, l les préférences de l'agent peuvent être représentées par une fonction d'utilité u(.) définie à une transformation affine positive près. A présent, il s'agit de décrire la méthode proposée par Ramsey pour mesurer la croyance partielle, mesure qui est nécessairement liée à celle des utilités.

Etape 4 : mesurer les probabilités subjectives à partir de la mesure des degrés de croyance

Pour mesurer les probabilités subjectives, Ramsey utilise un raisonnement bayésien[143] fondé sur le comportement de pari des individus, qui permettent de déterminer des mesures des degrés de croyance, et la conditionnalité des probabilités.

L'idée est d'apporter une réponse à la question suivante : comment mesurer le degré de croyance d'un individu vis-à–vis de p si p dépend de la vérité d'une autre proposition q par exemple (on pourrait imaginer, comme Sahlin [1990] le propose, que p est la proposition suivant laquelle la situation économique va s'améliorer et q la proposition suivant laquelle le gouvernement va dévaloriser au moins de 20%).

Ramsey construit une méthode qui s'appuie sur ce qu'il appelle « les lois fondamentales de la croyance probable » (Ramsey [1926, 2003], p. 173). Ces « lois » sont énoncées afin de s'assurer que les degrés de croyance calculés à partir des propositions éthiquement neutres, respectent le calcul des probabilités.

(1) Degré de croyance en p + degré de croyance en ~p = 1
(2) Degré de croyance de p étant donné q + degré de croyance en ~p étant donné q = 1
(3) Degré de croyance en (p et q) = degré de croyance en p × degré de croyance en q étant donné p
(4) Degré de croyance en (p et q) + degré de croyance en (p et ~q) = degré de croyance en p.

[143] En référence au révérend Thomas Bayes (1702-1761) qui avait énoncé un théorème relatif aux probabilités conditionnelles.

Autrement dit, comme on l'a vu dans la première partie de la thèse, une fois que l'on a fixé des propositions éthiquement neutres de degré ½ et ¼, il est possible de déterminer les intervalles d'utilité (u(x)-u(y)). Une fois ces intervalles d'utilité déterminés, on peut, si les lois définies précédemment sont vérifiées, obtenir la mesure des degrés de croyance des individus (P(.)). La mesure des probabilités subjectives sera alors obtenue comme dans l'exemple suivant.

Supposons un individu qui est face aux choix :

B1 : a si q est vraie, b si ¬q est vraie

B2 : c si (p est vraie et q est vraie), d si (¬p est vraie et q est vraie), b si ¬q est vraie,

Supposons que u(.) est une fonction d'utilité (comme nous l'avions présenté dans la partie 1 chapitre II) et P(.) une fonction représentant le degré de croyance d'un agent dans une proposition et que cette fonction satisfait les lois de la probabilité définie ci-dessus ; si l'individu est indifférent entre B1 et B2 alors il s'ensuit que :

$$P(q)\,u(a) + P(\sim q)\,u(b) = P(p \vee q)\,u(c) + P(\sim p \vee q)\,u(d) + P(\sim q)\,u(b)$$

Si l'on veut établir la mesure de la probabilité subjective de p si q, on peut supprimer les u(b), on obtient :

$$u(a) = \left(\frac{P(p \vee q)}{P(q)}\right) u(c) + \left(\frac{P(\sim p \vee q)}{(P(q)}\right) u(d)$$

Comme une probabilité conditionnelle est définie ainsi $P(p/q) = P(p \vee q) / P(q)$

Et comme $P(\sim p \vee q) = 1 - P(p \vee q)$

Alors $\frac{u(a) - u(d)}{(u(c) - u(c)} = P(p/q)$

Cette présentation de la croyance partielle sous forme de probabilités permet à Ramsey de développer simultanément une mesure de l'utilité et une mesure de la croyance partielle. Cette mesure des croyances partielles est donc liée selon Ramsey à la mesure des utilités. Seulement, Davidson, Suppes et Siegel n'avaient utilisé en 1957 que la partie relative à la mesure des utilités par des intervalles également espacées en utilisant

la proposition éthiquement neutre E* sans faire usage des probabilités conditionnelles. En revanche, Jeffrey utilisera lui la méthode relative aux croyances partielles pour construire son modèle (1965, 1983).

2.2. Le triplet désirs/ significations/ croyances

C'est au sein de la théorie de la décision que Davidson tente de démêler et délier les rôles respectifs des désirs (ou utilités) et des croyances (probabilités) (Davidson [1957], p. 9). Mais cette distinction n'est qu'artificielle car, comme on l'a vu, l'axiomatique proposée par Ramsey nous permet de les déterminer simultanément. L'objectif de cette séparation artificielle est simplement d'identifier les rôles respectifs des désirs, croyances et significations.

Comme le souligne Isaac Levi [1999], Davidson a toujours été préoccupé par la compréhension des attitudes propositionnelles de croyance et de désir comme de « facteurs motivant le contrôle de la délibération de la prise de décision » (Levi [1999], pp. 531-532). Même si ses vues sur le sujet ont changé, la problématique soulevée par ses collègues et lui-même dans les années 1950, demeure « un stimulus important pour le développement des idées modelées par son travail antérieur » (*ibid.*).

Le couple croyance-désir est, comme nous l'avons montré dans le chapitre 1, effectivement largement analysé notamment dans la théorie de la décision de Davidson écrite en 1957. Mais, ce que l'auteur a tiré des expériences des années 1950, c'est aussi l'idée que la théorie de la décision canonique négligeait un pan entier du mental : les significations (voir section IV chapitre I). C'est l'interconnexion des désirs, des croyances et des significations qui permet de mettre en évidence une « image du mental » plus complète et exhaustive. Ainsi, si l'analyse de Davidson commence dans les années 1950 avec une investigation du couple désir/croyance permettant d'expliquer les choix d'individus placés dans un univers risqué, elle se poursuit dans les années 1960 mais cette fois dans le cadre de la philosophie de l'action. Ces deux entités sont chacune corrélée aux significations et chaque corrélation ouvre un champ d'investigation singulier. Ainsi le couple désir/signification permet une analyse de la

manière dont les désirs – et les autres attitudes évaluatives – sont exprimés et compris par nos semblables. Le couple croyance/signification renvoie à des considérations relatives à la philosophie du langage

Après un rappel de l'analyse du couple désirs/croyances et de son évolution dans les années 80 (2.2.1), nous nous intéressons successivement aux deux autres couples désirs/ significations (2.2.2) et croyances/ significations (2.2.3).

2.2.1. Le couple désirs/croyances

L'analyse du couple croyances-désirs réalisée dans la théorie de la décision de 1957 de Davidson est à nouveau au cœur de sa théorie de l'action en 1963.

Davidson l'introduit plus précisément en théorie de l'action en reprenant le syllogisme pratique aristotélicien : schématiquement, le modèle initial repose sur une conception simple du raisonnement pratique d'Aristote dont les prémisses sont des désirs et des croyances donc la désirabilité de l'action correspondante est déduite par l'agent. Davidson s'engage donc dans une certaine interprétation du syllogisme pratique dont la conclusion est une action de l'agent.

Le couple croyance-désir constitue, selon Davidson, la « raison primaire » pour laquelle l'agent a accompli une action. Plus précisément, « *R* n'est une raison primaire pour laquelle un agent a accompli l'action *A* sous la description *d* que si *R* consiste en une pro-attitude de l'agent à l'égard d'actions qui ont une certaine propriété et en la croyance de l'agent que *A*, sous la description *d*, a cette propriété » (Davidson [1963, 1993a], p.18). Ainsi, la description et l'explication de l'action sont-elles réduites à l'utilisation de deux attitudes propositionnelles qui mènent logiquement et causalement à l'action : un désir (ou une autre pro-attitude comme des volontés, des envies, des incitations…) de l'agent qui le mène à vouloir, à être disposé à agir pour l'accomplissement de telle action ayant telle caractéristique ; une croyance que l'action accomplie a la caractéristique en question.

Même si l'on peut avancer l'idée qu'un agent a des raisons variées d'accomplir une certaine action, il y a néanmoins *une* raison pour laquelle il l'accomplit, c'est *la* raison primaire de l'action, qui se trouve aussi en être *la* cause. Plus précisément, « Une raison ne rationalise une action que si elle nous conduit à voir quelque chose que l'agent a vu ou cru voir dans son action – un trait, une conséquence ou un aspect quelconque de l'action que l'agent a voulu, désiré, prisé, chéri, considéré comme étant de son devoir, bénéfique, obligatoire, ou agréable » (Davidson [1963, 1993a], p.15). Donner la raison des actions, c'est donc non seulement la décrire comme le fruit d'attitudes propositionnelles mais aussi utiliser des attitudes propositionnelles causalement pour expliquer l'action. Ainsi, comme le souligne Davidson : « Justifier une action et l'expliquer vont souvent de pair ; c'est pourquoi nous indiquons souvent la raison primaire qu'on a eue pour faire une action en avançant une proposition qui, si elle était vraie, aurait aussi pour effet de vérifier, de justifier, ou de confirmer la croyance ou l'attitude pertinente de l'agent » (Davidson [1963, 1993a], p. 22).

D'autant que c'est à la lumière de la raison primaire qu'avait l'agent de faire telle action que l'agent apparaît dans son rôle d' « Animal Rationnel » (Davidson [1963, 1993a], p. 22). Autrement dit, donner la raison primaire pour décrire et expliquer une action, c'est faire usage du concept de rationalité. Plus précisément, la description d'une action sous forme d'une explication causale met en jeu un certain type de rationalité induite dans l'action.

2.2.2 Le couple désirs/significations

Le couple désir/signification relève de la connexion entre le langage et l'évaluation. Plus précisément, ce couple met en évidence les liens qui existent entre la manière dont un individu évalue, ou valorise un objet par exemple et la signification (pour le locuteur et pour l'interprète du locuteur) que cette évaluation recouvre.

Par évaluation, Davidson n'entend pas un acte verbal mais une attitude. Les attitudes évaluatives correspondent à toutes les attitudes comme désirer, vouloir, chérir, tenir pour correct ou obligatoire, ainsi que les versions comparatives et négatives de ces attitudes.

Le désir ne constitue donc qu'un type particulier d'évaluation.

Pour lui, les évaluations, tout particulièrement le désir, sont liées aux significations des phrases d'un individu. On pourrait alors simplement imaginer qu'en disposant de la signification d'une phrase prononcée par un individu, nous avons accès à une partie de l'explication de la manière dont l'individu avait l'intention, en prononçant cette phrase, de réaliser ce qu'il a fait. Mais selon Davidson, il n'y a aucune règle qui relie systématiquement, de manière stricte, les mots, et leur interprétation attendue, signifiée, et ce que le locuteur a l'intention d'asserter, demander, commander (Davidson [1984, 2004], pp. 21-22). Autrement dit, une phrase ne signifie pas nécessairement la même chose pour le locuteur et l'interprète, ou pour plusieurs interprètes. L'évaluation est dès lors elle aussi différente en fonction eu locuteur ou de l'interprète.

La solution proposée par l'auteur est de se placer dès le départ du point de vue de l'interprète, c'est-à-dire l'agent qui tente de comprendre, d'interpréter et traduire les mots et les phrases d'un individu : « La clé pour comprendre tous ces phénomènes mentaux est de les considérer du point de vue de l'interprète. Les voir de cette perspective nous conduira à apprécier les éléments inéluctablement objectifs et intersubjectifs non seulement dans le langage et les croyances mais aussi dans l'évaluation » (Davidson [1984, 2004], p.20).

L'idée de Davidson est qu'un interprète doit être capable de dire quand un locuteur tient une phrase qu'il dit comme vraie ou fausse, ou s'il veut qu'elle soit vraie, ou qu'il a l'intention de la rendre vraie. Nous verrons notamment dans le chapitre 3 comment Davidson introduit cette compétence de l'interprète.

2.2.3 Le couple croyances/significations

L'analyse du couple croyance/signification est essentiellement traitée dans les articles de Davidson relatif à la philosophie du langage.

Une question centrale relative à l'analyse de ce couple est celle de savoir si les croyances qui interviennent ici sont du même type que les croyances que nous venons d'évoquer dans la section relative au couple désirs/croyances.

Il semble à première vue que les croyances, lorsqu'elles sont analysées par Davidson dans le couple croyances/significations ne soient pas des probabilités comme dans le couple désirs/croyances. En effet, comme le souligne Pascal Engel [1984,1998], on peut

d'un côté définir la croyance comme un état psychologique comme c'est le cas en théorie de la décision standard et on peut, de l'autre, on peut la définir comme une relation à une proposition et dans ce cas, l'assimilation à une probabilité est plus délicate. Même si ces deux versants de la croyance sont interdépendants, il n'en reste pas moins qu'ils ne renvoient pas à la même image de l'action. Ces deux points de vue sont d'ailleurs présents dans les travaux de Ramsey. Dans le premier cas, comme le souligne Pascal Engel [1998], la croyance est un « état dispositionnel » passif, « délié d'un acte verbal d'assertion » (*ibid*., p.328). C'est l'exemple de l'arsenic proposé par Ramsey : « un morceau d'arsenic est appelé toxique non pas parce qu'il a effectivement tué ou tuera quelqu'un, mais parce qu'il tuerait quelqu'un qui en prendrait » (Ramsey [1926]). Autrement dit, la croyance est conçue comme une disposition à l'action : elle n'est pas effective en dehors du moment de l'action. Dans le deuxième cas, la croyance est un acte volontaire d'assertion par lequel l'individu affirme qu'une certaine proposition est vraie. C'est ce deuxième cas qui est à l'œuvre ici et cette conception de la croyance est liée à celle de signification.

En effet, la signification et la croyance « jouent des rôles extrêmement liés et complémentaires dans l'interprétation du langage » (Davidson [1974, 1993b], p.208).

La difficulté est de procéder à une analyse simultanée de ces deux éléments sans postuler au départ ni l'un ni l'autre. En effet, il semble vain de tenter d'inférer d'une énonciation quelconque la croyance sans connaître la signification et vice versa. La raison à cela est qu'un « locuteur qui tient une phrase pour vraie en telle circonstance le fait en partie en raison de ce qu'il veut dire, par, une énonciation de cette phrase, et en partie en raison de ce qu'il croit » (Davidson [1974, 1993b], p. 210).

Imaginons par exemple que je dise « Socrate est un chien », toute analyse de cette phrase qui ne tiendra pas en compte du fait que Socrate est le nom de l'animal qui se tient devant moi et dont je crois qu'il est un chien, ne permettra pas de saisir ce que je veux signifier par cette phrase.

Dès lors, puisque ces deux éléments sont liés et codéterminés, Davidson en conclut que « lorsque nous interprétons des énonciations en partant de zéro – dans une interprétation *radicale* – nous devons en quelque sorte présenter simultanément une théorie de la croyance et une théorie de la signification » (Davidson [1974, 1993b], p. 210).

La méthodologie utilisée pour analyser simultanément les croyances et les significations est calquée sur celle utilisée par Davidson en 1957. Plus précisément, ce que recherche Davidson, c'est une méthode qui produise le même effet que la méthode de Ramsey consistant à utiliser des propositions éthiquement neutres, c'est-à-dire fixer un élément pour en déterminer un autre, qui lui est corrélé.

De la même manière que l'on part des préférences ordinales dans la version de la théorie de la décision de Ramsey, on tente de se focaliser sur une donnée qui ne nécessite ni ne postule des croyances et de significations. Cette donnée c'est le « tenir pour vrai » c'est-à-dire un ensemble de phrases que le locuteur tient pour vraies. Avoir accès aux phrases qu'un locuteur tient pour vraies c'est non seulement mettre en relief l'interdépendance des croyances et des significations mais aussi - comme dans le cas de la théorie de la décision avec les désirs et les croyances – les rôles respectifs de ces deux entités : « L'interdépendance de la croyance et de la signification est ainsi évidente : un locuteur juge une phrase vraie en raison de ce que signifie la phrase (dans son langage), et en raison de ce qu'il croit » (Davidson [1974, 1993b], p. 200).

Mais comment décrire cette attitude ?

Cette attitude est elle-même une croyance. Seulement, c'est une attitude qui n'implique pas une connaissance d'une large gamme de croyances et des significations du locuteur. C'est une attitude « dont on peut supposer qu'un interprète sache l'identifier avant de pouvoir interpréter puisqu'il peut savoir qu'une personne a l'intention d'exprimer une vérité en énonçant une phrase sans avoir la moindre idée de la vérité *dont il s'agit* » (Davidson [1973, 1993b], p. 200). Tenir pour vraie une phrase serait selon Davidson une attitude globale qui en recouvre plusieurs autres comme souhaiter qu'une phrase soit vraie ou encore vouloir qu'elle soit vraie. Autrement dit, une grande panoplie d'attitudes face à des phrases peut être représentée par l'attitude de tenir pour vraie une phrase. Comme on le verra dans le chapitre 3 de la partie II, cette attitude est liée au principe de charité. Ce principe, que l'on peut concevoir en partie comme l'analogue aux canons de rationalité de la théorie de la décision revient à fixer l'ensemble des croyances grâce à l'attitude de tenir pour vraie une phrase et permet donc de déterminer les significations.

2.3. Economie et philosophie : dépendance mutuelle

Dans les chapitres précédents, nous avons insisté sur l'objectif de Davidson de mettre en évidence les connexions entre la théorie de la décision comme il l'a testé dans les années 1950 et des questions de philosophie du langage et de philosophie de l'action qu'il a analysé dans les années 1960-1970.

Dans cette section, nous allons tenter d'établir un certain nombre de connexions en marquant premièrement l'interdépendance de la théorie de la décision et de l'interprétation du langage défendue par l'auteur (2.3.1), puis le transfert de méthodologie opérée par Davidson de la première vers la seconde (2.3.2) et enfin la possibilité d'unir leurs contenus au sein d'une théorie unifiée (2.3.3).

2.3.1 La théorie de la décision et la théorie de l'interprétation du langage ont besoin l'une de l'autre

Comme on l'a vu, il y a plus qu'une analogie entre théorie de la décision et théorie de l'interprétation, il y a un lien.

Les choix entre des options comme celles présentées en 1957 sont déterminés par deux facteurs « psychologiques » (Davidson [1974, 1993b], p. 215) : « les valeurs relatives que celui qui choisit place sur les résultats » c'est-à-dire les intensités de préférences et « la probabilité qu'il assigne à ces résultats », probabilités subjectives comme on l'a vu dans le chapitre I. La méthode opérationnelle de Ramsey permet d'accéder à ces deux facteurs en partant d'une donnée minimale : les préférences ordinales.

Ce problème ressemble, selon Davidson, au problème de l'interprétation. Comme nous l'avons expliqué, l'interprétation du langage implique de déterminer simultanément les croyances et les significations. Seulement, « la solution en théorie de la décision est nette et satisfaisante, il n'y a rien d'aussi bien à notre disposition en théorie de la signification » (Davidson [1974, 1993b], p. 214). L'une des raisons avancées par Davidson pour expliquer cette moindre précision est qu'il est plus délicat d'attribuer des nombres à des phrases qu'attribuer des nombres à des résultats. L'isomorphisme de la classe des nombres et des utilités permet une telle assimilation mais les significations

ne permettent pas une telle assimilation : « Les propositions sont bien plus vagues que les nombres » (Davidson [1974, 1993b], p. 217).

L'analogie mentionnée plus haut va plus loin puisque Davidson parle de « lien ».

Là où la théorie de la décision rencontre ce que l'on avait identifié comme des « effets de présentation » tels que décrits par Edwards, la théorie de l'interprétation est face à la difficulté consistance à savoir quand un individu tient une phrase pour vraie.

Dans le cas de la théorie de la décision, il est utile, selon Davidson, pour apprendre les préférences d'un agent de « décrire les options en mots » (*ibid.*). Il est selon l'auteur, nécessaire d'avoir accès à une « information finement filtrée concernant les croyances et les intentions » (*ibid.*). Cette information n'est pas accessible, selon l'auteur, sans faire référence à une donnée mentale jusque là négligée : les significations.

Dans le cas de la théorie de l'interprétation, il est nécessaire d'avoir accès à une gamme plus large de croyances que la simple croyance en jeu dans le « tenir pour vrai ». Ces croyances sont à l'œuvre dans la structure même de la théorie de la décision puisque celles-ci, lorsqu'elles sont combinées aux utilités cardinales, donnent une rationalisation pertinente du choix d'un individu.

On comprend dès lors pourquoi Davidson milite pour une théorie qui unifierait les contenus respectifs des deux théories prises isolément et ce notamment grâce aux travaux de Jeffrey permettant de réduire l'ontologie de la théorie de la décision à une ontologie de propositions uniquement (Davidson [1974, 1993b], p. 218).

Cette opération peut se réalisée selon Davidson si les méthodologies respectives des deux théories peuvent être rapprochées et comparées.

2.3.2 Une même méthodologie

Le lien mentionné entre théorie de la décision et théorie de l'interprétation se reflète par un une analogie des méthodes.

Ainsi, selon Davidson, « les choix réels en théorie de la décision correspondent à des énonciations réelles dans l'interprétation » (Davidson [1974, 1993b], p.215). Les données de base sont donc dans les deux cas des éléments observables.

L'analogie se précise : « Aux degrés de croyance et aux désirs postulés par la théorie de la décision correspondent les croyances et les significations en théorie de

l'interprétation. La partie observable est constituée dans un cas par les préférences ou les choix, et dans l'autre par le comportement verbal » (*ibid.*). Un parallélisme est donc construit entre préférences ordinales et énonciations.

Les méthodes semblent similaires pour Davidson car « Dans chacun des cas, on ne peut déterminer un élément sans déterminer les deux autres. Si l'on poursuit l'analogie, ce que nous avons besoin en théorie de l'interprétation est une donnée qui soit l'homologue de la notion de préférence ordinale entre des paris qui sert de levier pour déterminer les degrés de croyance et les différences de valeur ». Autrement dit, dans chacun des cas, on part d'une donnée minimale qui ne suppose pas ce que l'on tente d'expliquer, ceci permettant d'éviter la pétition de principe.

Le véritable élément qui corresponde aux préférences ordinales et l'attitude de « tenir pour vrai » que nous avions évoqué plus haut.

2.3.3 Vers une théorie unifiée : l'expérimentateur devient interprète.

En introduisant le rôle des significations au cœur de la théorie de la décision dans la version de Bolker-Jeffrey, Davidson élargit le rôle initial de l'expérimentateur. En fait, son rôle va s'étendre à celui d'interprète. En effet, comme le mentionne Davidson, « la théorie de l'interprétation est l'affaire, conjointement, du linguiste, du psychologue et du philosophe » (Davidson [1974, 1993b], p. 208). Autrement dit, en introduisant les significations et en leur accordant un rôle dans le comportement de décision des individus, l'expérimentateur doit à la fois mesurer les utilités cardinales et les probabilités subjectives mais aussi interpréter les énoncés des sujets et cette dernière procédure constitue selon Davidson une mesure du mental au même titre que la mesure proposée par Ramsey même si la première est moins précise que la seconde.

Comme nous allons le montrer dans le chapitre 3, l'extension du rôle de l'expérimentateur vers celui d'interprète nécessite une interaction profonde avec le sujet. Plus précisément, l'interprétation dont va parler Davidson implique une comparaison des désirs, des croyances et des significations du sujet avec ceux de l'expérimentateur.

CONCLUSION

L'objectif de ce chapitre était de décrire et d'analyser les fondements théoriques et méthodologiques de la théorie unifiée que propose Davidson dans les années 1980. Pour cela, nous sommes revenus sur l'influence de Ramsey qui, une fois de plus, est décisive pour le projet de l'auteur. En effet, la suggestion de Davidson d'unifier la théorie de la décision et la théorie de l'interprétation du langage au sein d'une même théorie, repose sur un usage des deux sens de la croyance à l'œuvre dans les travaux de Ramsey. Ainsi, la croyance vue à la fois comme facteur psychologique assimilable à une probabilité et comme une relation à une proposition, sert de clé de voûte permettant d'unir les contenus spécifiques de la théorie de la décision et de la théorie de l'interprétation du langage.

Cette analyse est renforcée par la présentation des trois couples (désirs/croyances, désirs/significations et croyance/significations) qui forment la théorie unifiée.

L'intégration d'une théorie de l'interprétation du langage au sein de la théorie de la décision revient à faire usage de deux mesures du mental simultanément. La mesure de la théorie de la décision s'appuie sur la méthode opérationnelle de Ramsey alors que celle de la théorie de l'interprétation du langage est issue des travaux de Tarski et de Quine.

L'usage de ces deux mesures n'est pas sans effet : l'expérimentateur devient interprète et la théorie de la décision est dotée d'une dimension supplémentaire qui dépasse le cadre strictement individuel et met en jeu l'interaction du sujet locuteur avec l'expérimentateur-interprète.

Chapitre 3 : Le modèle de l980 comme expression de la théorie unifiée

L'objectif de l'article de Donald Davidson [1980], « A unified theory of thought, action and meaning », est le suivant : il s'agit de construire une théorie permettant d'interpréter les mots d'un locuteur en se basant sur une analyse et une détermination simultanée des désirs, des croyances et des significations d'un agent. L'idée étant de surmonter les défaillances de la théorie de la décision (en particulier celles du modèle de 1957), que nous avons identifiées ci-avant (à la fin de la première partie de la thèse, et dans le chapitre 1 de cette seconde partie).

En effet, une théorie de l'interprétation est l'affaire, selon l'auteur, « conjointement, du linguiste, du psychologue et du philosophe » (Davidson [1974, 1993b], p.209).

L'intérêt pour l'économiste d'une telle théorie est double :

- Cette théorie, que Davidson qualifie d'« unifiée » constitue une tentative de dépassement et d'enrichissement du modèle de 1957. Elle peut en ce sens être considérée comme la construction d'une nouvelle théorie de la décision. (3.2)
- Pour ce faire, Davidson utilise en outre le modèle de l'économiste Richard Jeffrey (1965, 1983) dont il transpose presque à l'identique la structure et la méthodologie. (3.1.). Il s'agit donc pour Davidson - à une époque où il s'est pourtant tourné davantage tourné vers la philosophie de l'action et du langage - d'étudier un objet au cœur de la théorie économique avec les outils mêmes de cette dernière. Outre l'apport analytique d'une telle démarche, c'est donc aussi un nouvel éclairage épistémologique à la question des rapports entre l'économie et la philosophie que nous cherchons ici à offrir.

3.1. L'emprunt à Jeffrey

Comme nous l'avons signalé, pour Davidson, la défaillance la plus importante de la théorie de la décision telle qu'elle est établie dans les années 1950 repose sur l'impasse faite sur une analyse des significations. Ce défaut est le plus important car il explique et engendre beaucoup des critiques que l'on peut adresser à la théorie de la décision (le traitement des conflits de désirs, la manière dont les choix évoluent dans le temps et ce que Tversky (1975) appelle l'effet de certitude, les problèmes relatifs à l'effet de présentation…).

C'est pourquoi, selon Davidson, « ce que l'on doit ajouter à la théorie de la décision, ou incorporer à l'intérieur d'elle, est une théorie de l'interprétation pour un agent, une manière de dire ce que ses mots signifient » (Davidson [1980, 2004], p.155).

Il ne s'agit pas, néanmoins, pour ce faire d'introduire au sein du modèle de décision de 1957 des significations ad hoc mais de chercher à décrypter les significations que les acteurs attribuent aux propositions en même temps que l'on a accès à leurs croyances et désirs.

Autrement dit, il s'agit de construire une théorie, basée sur les modèles standards de théorie de la décision, qui incorpore l'analyse des significations sans pour autant postuler à l'avance l'existence des entités à expliquer, comme Davidson le mentionne : « cette addition doit être faite en l'absence d'information détaillée sur les croyances, les désirs, ou les intentions » (*ibid.*, p. 155).

La raison pour laquelle une analyse des significations en théorie de la décision est indispensable (comme cela a été expliqué dans le chapitre 1) est notamment que les expérimentateurs supposent généralement que les mots utilisés par le sujet et ceux utilisés par l'expérimentateur peuvent être interprétés de la même manière. En effet, lorsque des choix sont proposés aux sujets des expérimentations, le langage et ses significations, sont en quelque sorte imposés à l'individu[144] (Davidson [1974, 1993a], pp.315-316). Or, selon Davidson, imposer le langage de l'expérimentateur escamote les significations que les sujets attribueraient eux-mêmes aux propositions et par là conduit

[144] De manière plus subtile, on pourrait dire que les significations sont supposées connues car elles sont celles de l'expérimentateur.

à négliger toutes les informations sur le comportement des sujets qui découleraient de l'analyse des ces significations : « établir qu'une attribution de croyance ou de désir est correcte pose à peu près les mêmes problèmes que montrer que nous avons compris les paroles de quelqu'un d'autre » (*ibid.*).

Mais l'interaction nécessaire entre théorie de la décision et interprétation du langage va plus loin selon Davidson. Les deux théories se complètent puisqu'elles constituent des mesures du mental et traitent d'éléments combinés : « La théorie de la signification et la théorie bayésienne de la décision sont faites l'une pour l'autre. La théorie de la décision doit être libérée de l'hypothèse d'une connaissance de la signification déterminée indépendamment ; la théorie de la signification en appelle à une théorie du degré de croyance afin de faire un usage sérieux des relations du support de la preuve » (Davidson [1980, 2004], p.158).

La solution proposée par Davidson fait usage des outils proposés par la théorie de la décision puisque la structure proposée par l'auteur s'inspire largement des travaux de l'économiste Richard Jeffrey [1983] : « nous devons à Richard Jeffrey une version de la théorie bayésienne de la décision qui ne fait pas directement usage des paris mais considère les objets de la préférence, c'est-à-dire les objets sur lesquels les probabilités subjectives et les valeurs relatives sont assignées, comme des propositions » (Davidson [1980, 2004], p.160).

Le modèle de Jeffrey permet de mêler théorie de la décision - version Ramsey-Savage – et théorie du langage - étant donnée la place centrale accordée aux propositions - , et ce grâce aux outils mathématiques proposés par Ethan Bolker [1966, 1967] – c'est pourquoi nous parlerons plus volontiers du modèle de Bolker-Jeffrey.

3.1.1 Jeffrey et les préférences sur des propositions

Jeffrey propose de bâtir « une nouvelle théorie des préférences entre (la vérité) des propositions » (Jeffrey [1983], p. xi). Comme nous l'avons montré dans le chapitre 1,

chez Ramsey, les désirabilités sont attribuées aux conséquences et les probabilités à des conditions possibles ou à des propositions. Les paris permettent de lier les désirabilités et les probabilités subjectives.

Jeffrey propose un système qu'il qualifie d' « alternatif » [1983], p.59) à celui de Ramsey dans lequel les paris n'occupent plus cette place centrale, un système qui attribue des probabilités et des désirabilités aux mêmes objets, des propositions. Les paris sont en effet remplacés par des « opérations élémentaires sur des propositions » (Jeffrey [1983], p. xi).

En effet, selon Jeffrey, croire, c'est avoir une certaine attitude vis-à-vis d'une proposition. Désirer, c'est en avoir une autre vis-à-vis de la même proposition.

Désirer *x*, c'est désirer avoir *x*. Désirer avoir *x* c'est désirer qu'il soit le cas que *x* ou que *x* survienne, ou encore que *x* soit vrai, donc ce qui est désiré c'est une proposition[145]. Ce que propose Jeffrey, c'est une interprétation du langage ordinaire qui soit cohérente avec les besoins de sa théorie. Tout en proposant un modèle différent de celui de Ramsey en ce qui concerne les notions de paris et de conséquences, l'objectif central de la théorie de Jeffrey est de rendre opérationnel, c'est-à-dire utilisable, la suggestion de Ramsey selon laquelle les objets des croyances des individus sont des propositions.

Pour arriver à ce résultat, Jeffrey part de la théorie de Ramsey (1926). Ce dernier montre que si la désirabilité du jeu $\{B \text{ si } A \text{ ; } C \text{ si } \bar{A}\}$ où A, B et C sont des propositions est :

$$p(A)\, d(AB) + p(\bar{A})\, d(\bar{A}C),$$

alors la fonction $p(\cdot)$ est définie de façon unique, la fonction $d(\cdot)$ étant définie, comme chez vNM, à une transformation affine près ($D(X) = a\, d(X) + b$) (Jeffrey [1983], p.95). Il suffit donc d'attribuer une désirabilité pour obtenir la classe des désirabilités qui représentent la même relation de préférence. En effet, chez Ramsey - selon Jeffrey - une fois que les désirabilités sont assignées à une paire de conséquences classées de manière

[145] Si l'on considère par exemple que les propositions sont relatives aux conditions de vérité d'un énoncé quelconque.

différente, les désirabilités des autres conséquences sont déterminées (*a* et *b* sont déterminés) par déduction.

Jeffrey reprend cette méthodologie mais s'en démarque de deux manières. D'une part, il propose une théorie de la préférence qui est unifiée dans le sens où elle attribue, comme on l'a dit, des probabilités et des désirabilités aux mêmes objets. Il ne fait pas de différence entre ce sur quoi porte la désirabilité (les actes et les conséquences chez Savage et Ramsey) et ce sur quoi porte la probabilité (un évènement comme élément d'état du monde selon Savage et Ramsey). Les préférences, utilités et probabilités étant toutes définies sur le même ensemble de possibilités ou perspectives (*prospects*). Ces perspectives sont des propositions auxquelles Jeffrey applique le calcul propositionnel. Ainsi les désirs sont considérés comme portant sur la vérité des propositions (Jeffrey [1983], p. xi) et les probabilités représentent le degré de croyance d'une proposition.

Cette théorie, d'autre part, est « non causale », ce qui signifie pour lui qu'il n'existe plus de fonction reliant les actes et les conséquences. La non causalité provient donc du raisonnement exclusif sur les propositions, et sur leurs probabilités.

Avant de présenter les axiomes de Jeffrey sur les probabilités et les désirabilités (3.1.1.2), et les conditions de leur existence (3.1.1.3), commençons par mettre en lumière les éléments du calcul propositionnel utilisé par Jeffrey (3.1.1.1). Nous pourrons alors recombiner les axiomes, leurs conditions d'existence et le calcul propositionnel de Jeffrey pour les présenter pour permettre une comparaison avec les axiomatiques de vNM et Savage (3.1.1.4).

3.1.1.1 Calcul propositionnel et algèbre booléenne

Le cadre mathématique de la théorie de Jeffrey est particulier.

Il considère d'abord qu'une proposition peut être considérée comme un sous-ensemble de l'ensemble de tous les mondes possibles, le sous-ensemble consistant dans les mondes où la proposition est vraie.

Il définit ensuite une série d'opérations élémentaires que l'on peut effectuer dans cet ensemble. Ainsi, on considère que :
A est une proposition et ~A est son complément ;
$A \vee B$ est l'union de A et de B ;
$A \wedge B$ est l'intersection de A et de B
Si A et B sont contraires[146] (*contraries*) (deux propositions qui ne peuvent pas être vraies toutes les deux), ce sont des ensembles disjoints.

Imaginons en outre un exemple où trois propositions A, B et C telles que : A « Je suis à un rendez-vous galant », B « Je m'endors », C « Je tombe par terre ».
A chaque fois que A est la disjonction de deux propositions contraires A_1 et A_2 (par exemple $A_1 = A \wedge B$ et $A_2 = A \wedge \sim B$ c'est-à-dire A_1 correspond à « Je suis à un rendez-vous galant et je m'endors » et A_2 « Je suis à un rendez-vous galant et je ne m'endors pas » ; A_1 et A_2 sont donc des propositions incompatibles), on peut dire d'une part que la mesure de la probabilité de A, $p(A)$[147] est telle que $p(A) = p(A_1) + p(A_2)$ et que l'utilité de A est telle que $U(A) = \frac{p(A_1)\ U(A_1) + p(A_2) U(A_2)}{(p(A_1) + p(A_2)}$.

Ainsi on peut en déduire que $U(A) = \frac{p(A_1)}{p(A)} U(A_1) + \frac{p(A_2)}{p(A)} U(A_2)$ sachant que $\frac{p(A_1)}{p(A)}$ et $\frac{p(A_2)}{p(A)}$ correspondent respectivement aux probabilités de A_1 et A_2 conditionnées à A[148].
Selon cette formule, l'utilité de A est l'espérance d'utilité lorsque A est vraie.

[146] A et B ne sont pas opposées car A n'est pas le complément de B mais A et B sont incompatibles.
[147] Broome la note μ (A)
[148] Nous reprenons ici la présentation de Broome [1990].

Comme on l'a dit, la particularité du modèle de Bolker-Jeffrey est que les préférences portent sur des objets spécifiques, des propositions.

Ainsi, si le sujet préfère la proposition A à la proposition B, c'est qu'il préfère que la proposition A soit vraie plutôt que B.

Pour être plus précis, les préférences portent sur des éléments non nuls d'une algèbre booléenne sans atome (*atom free boolean algebra*).

Il est d'usage d'exprimer cette algèbre comme suit :

Soit $< \ell, \vee, \wedge , \sim >$ une algèbre booléenne sans atome. Cette expression permet de dire que toute perspective (prospect) dans l'ensemble considéré peut toujours être décomposée en une disjonction de la manière dont A se décompose en A_1 et A_2 – car l'ensemble ne comporte pas d'atome libre c'est-à-dire d'événement unique.

On peut décomposer A_1 en $A_{11} = A_1 \wedge C$ et $A_{12} = A_1 \wedge \sim C$ de telle sorte que $A_1 = A_{11} + A_{12}$. (en reprenant notre exemple : A_{11} correspond à « Je suis à un rendez-vous galant et je tombe par terre » et A_{12} « Je suis à un rendez-vous galant et je ne tombe pas par terre)

Et

$$U(A_1) = \frac{p(A_{11})}{p(A_1)} U(A_{11}) + \frac{p(A_{12})}{p(A_1)} U(A_{12})$$

Cette hypothèse signifie que toute perspective se décompose en une disjonction et que son utilité est l'espérance des utilités des éléments disjoints.

On définit par ailleurs en général les « extrema » comme suit : ℓ contient T, la proposition nécessairement vraie et F la proposition nécessairement fausse. Ainsi,

$T = A \vee \sim A$ pour tout A dans ℓ

$F = \sim T$

T est l'ensemble de tous les mondes possibles et F l'ensemble vide.

Le modèle de Bolker-Jeffrey exclut toutefois l'ensemble vide : $\ell' = \ell - \{F\}$.

Une algèbre booléenne sans atome est une algèbre booléenne dont chaque élément a un sous-élément strict non nul :

Pour chaque $A \in \ell$ autre que F, il existe un $B \in \ell$ tel que $B \rightarrow A$ et $B \neq A$ et $B \neq F$.

On suppose aussi que cette algèbre est complète. Cela signifie qu'elle contient toutes les disjonctions d'ensembles arbitraires contenant des membres opposés ou contraires.

Avant de décrire les conditions que doit remplir la relation de préférence, il nous faut revenir sur les axiomes relatifs aux désirabilités et aux probabilités.

3.1.1.2 Axiomes de désirabilités et de probabilités

Comme cela a été mentionné, les préférences, utilités et probabilités sont toutes définies sur le même ensemble de perspectives.
Soient deux propositions A et B (par exemple A « il pleut » et B « il vente »). Selon Jeffrey, il est possible de disposer de d'établir un tableau de tous les cas, toutes les situations de vérité de ces propositions. Dans le cas de deux propositions il existe 4 cas pour lesquels on peut avoir accès aux probabilités et désirabilités que l'agent leur assigne. D'une manière plus générale, le chercheur peut avoir accès aux probabilités et désirabilités notées p(.) et d(.) que l'agent attribue aux 2^n combinaisons de « vrai-faux » pour les propositions (conjonctions) de la forme *AB…V* (à *n* propositions élémentaires).

Par exemple, dans le cas $n = 2$, on a le tableau suivant, où les 4 valeurs des fonctions $p(\cdot)$ et $d(\cdot)$ sont les données de base, attribuées par l'agent aux 4 cas envisagés.

A	B	p (·)	d (·)
V	V	0,2	2
V	F	0,3	- 1
F	V	0,4	1
F	F	0,1	5

Tableau 11

Ces données sont alors combinées (*computed*) pour en déduire la probabilité des deux propositions A et B et leurs désirabilités. On opère aussi sur ces données des opérations élémentaires qui permettent d'obtenir tous les cas envisageables de propositions ($A \vee B$, $A \wedge \sim B$ … union, intersection, complémentaire, combinaison des uns et des autres – dans le cas présent les possibilités sont limitées parce qu'on a pris $n = 2$).
Dans notre exemple, cela donne :

p(A) = 0,2 + 0,3 = 0,5 (tous les cas où A est vrai)

p(B) = 0,2 + 0,4 = 0,6 (tous les cas où *B* est vrai)

p(A∨B) = 0,2 + 0,3 + 0,4 = 0,9[149] (tous les cas où au moins *A* ou *B* sont vrais).

A partir de notre exemple, cela signifie qu'à partir des probabilités que l'individu assigne au fait qu'il va, par exemple, pleuvoir et venter, nous sommes en mesure de déterminer la probabilité pour l'individu, qu'il pleuve seulement par exemple.

L'idée de Jeffrey, c'est d'attribuer des désirabilités (censées refléter les préférences de l'agent) aux 2^n propositions conjointes[150]. Pour cela, il adopte la règle consistant à définir la désirabilité d'une proposition comme la moyenne pondérée des désirabilités des cas pour lesquels la proposition est vraie, les pondérations étant les probabilités proportionnelles des cas (Jeffrey [1983], p. 78).

Par exemple, dans l'exemple du tableau ci-dessus, on obtient en appliquant cette règle :

d(A) = (0,2·2 + 0,3·(-1))/(0,2 + 0,3) = 0,2.

d(B) = (0,2·2 + 0,4·1)/(0,2 + 0,4) = 0,6.

On pourrait même calculer la désirabilité A∨B :

d (A∨B) = ((0.5/0.5+0.6)0.2)+ ((0.6/0.5+0.6)0.6) = 0.418

On voit donc, comme on l'expliquera plus loin que sur l'échelle de désirabilité de l'individu considéré, d(A) < d (A∨B) < d(B), ce qui correspond à la condition de moyenne présentée dans le paragraphe 4 de cette section.

Il vient que B est préférée à *A∨B* qui est préférée à A

On peut selon lui déduire du tableau initial, de la cartographie des cas, « un classement de toutes les désirabilités » (*ibid.*, p. 80). Notons que comme dans ce que Jeffrey appelle la théorie « classique » (celle de Ramsey) pour le classement des préférences et des utilités, on peut inférer des données initiales un classement des désirabilités des propositions mais on ne peut pas faire l'inverse.

[149] On vérifie que p(A∨B) = p(A) + p(B) - p(AB) = 0,5 + 0,6 – 0,2 = 0,9

[150] Cette idée est d'ailleurs considérée par Richard Bradley [2007] comme une avancée par rapport au modèle de Savage.

Cet exemple est généralisé et formulé sous forme d'axiomes portant sur les désirabilités et les probabilités.

La traduction formelle de cette propriété de la fonction $d(\cdot)$ est faite dans l'axiome suivant :

Axiome sur les désirabilités (Jeffrey [1983], p. 80):

La présentation des axiomes est ici réalisée à l'aide des notations de l'auteur, A et B sont les propositions.

Si $p(A \wedge B) = 0$ et $p(A \vee B) \neq 0$, alors

$$d(A \vee B) = \frac{p(A)}{p(A) + p(B)} d(A) + \frac{p(B)}{p(A) + p(B)} d(B).$$

Ainsi comme dans la théorie classique de Ramsey et vNM, le critère de l'utilité espérée permet de faire un classement des désirabilités et de déterminer la prise de décision.

A cet axiome, Jeffrey ajoute un axiome sur les probabilités, plus classique.

Axiome sur les probabilités

La fonction $p(\cdot)$ vérifie les 3 propriétés qui caractérisent toute loi de probabilité (valeurs positives et somme égale à 1, probabilité de l'union – conjonction - égale à la somme des probabilités, si l'intersection – disjonction - est vide).

Si on applique l'axiome sur les désirabilités à la proposition nécessaire (toujours vérifiée) : $A \vee \sim A$, notée T et donc telle que $p(T) = 1$, alors :

$$d(A \vee \sim A) = d(T) = p(A)\, d(A) + p(\sim A)\, d(\sim A)$$

Comme $p(\sim A) = 1 - p(A)$,

$$d(T) = p(A)\, d(A) + (1 - p(A))\, d(\sim A)$$

Et par conséquent :

$$p(A) = \frac{d(T) - d(\sim A)}{d(A) - d(\sim A)}$$

Ce qui montre comment les probabilités peuvent être déduites des désirabilités.

A partir de ces axiomes Jeffrey énonce la « condition d'existence » (Jeffrey [1983], p.96) suivante :

Le couple $(p(\cdot), d(\cdot))$ vérifie la condition d'existence, par rapport à une relation de préférence, s'il vérifie les axiomes 1 et 2 et

A préféré à B si et seulement si $d(A) \geq d(B)$.

Si un couple vérifie la condition d'existence, alors la fonction $d(\cdot)$ associe un nombre à chaque proposition de façon à ce que le classement des nombres corresponde à celui de la relation de préférence (à la façon des fonctions d'utilité).

3.1.1.3 Théorèmes d'existence et d'unicité

Le but de Jeffrey est alors de montrer qu'il existe au moins un couple $(p(\cdot), d(\cdot))$ qui vérifie la condition d'existence – et qui, par conséquent, représente (*mirrors*) les préférences de l'agent.

Pour cela, il commence par montrer que tous les couples qui ont cette propriété, s'ils existent, se déduisent les uns des autres selon les relations :

$$P(A) = p(A)(cd(A) + d)$$

$$D(A) = \frac{ad(A) + b}{cd(A) + d}$$

Ces transformations sont l'équivalent de la transformation affine de Ramsey et vNM, sauf que dans le cas présent il y a 4 paramètres qui interviennent, au lieu de 2 et qu'il s'agit d'une transformation fractionnelle. Ainsi comme dans la théorie classique de Ramsey et vNM, outre le critère de l'utilité espérée, il existe une échelle, un index des désirabilités espérées censé représenté le même classement que les désirabilités $d(\cdot)$.

En fait, pour que l'équivalence ait lieu entre les couples $(p(\cdot), d(\cdot))$ et $(P(\cdot), D(\cdot))$, les paramètres de la transformation doivent vérifier 3 conditions :

1. $ad - bc$ strictement positif (équivalent au $a > 0$ dans le cas de l'utilité espérée de la théorie de Ramsey Savage)
2. $c\,d(T) + d = 1$, ce qui établit un lien entre les paramètres si on se donne $d(T)$ et limite les degrés de liberté à 3.
3. $c\,d(A) + d > 0$ pour tout A.

Ces conditions sont nécessaires pour démonter que si $(p(\cdot), d(\cdot))$ vérifient les conditions d'existence, alors $(P(\cdot), D(\cdot))$ en font autant.

Pour déterminer les paramètres qui caractérisent un classement « cardinal » des préférences, il suffit donc de donner une valeur à la fonction $d(\cdot)$ pour trois propositions non équivalentes (Jeffrey propose d(T) = 0, d(G) = 1 (autrement dit il fixe en zéro et une unité) et discute assez longuement du cas restant).

3.1.1.4 Reformulation de l'axiomatique de Jeffrey

Pour établir une comparaison plus précise avec les théories de Ramsey, vNM, et Davidson (1957), et montrer en quoi le modèle de Davidson (1980) s'en inspire ou s'en démarque, nous allons présenter l'axiomatique de Jeffrey en détaillant cette fois les axiomes à l'aide des travaux mathématiques de Bolker (1967), travaux qui donnent la preuve mathématique des théorèmes présentés par Jeffrey. Nous nous appuierons à la fois sur les articles de Bolker mais aussi sur la présentation qui en est faite par Broome en 1990 car elle permet une comparaison plus directe avec Savage notamment.

Comme dans les théories de vNM, Savage et Davidson (1957), si les préférences satisfont certains axiomes, alors elles peuvent être représentées par une fonction d'utilité espérée.

La relation de préférence doit d'abord être transitive et trichotomique sur ℓ'c'est à dire :
H 1 : $\succsim$ est une relation de préférence transitive telle que si $A \succ B$ et $B \succ C$ alors $A \succ C$.
H2 : $\succsim$ est trichotomique c'est-à-dire que pour tous A et B dans ℓ', une seule de ces relations est valide ; $A \succ B$, $A \approx B$, $B \succ A$.

Elle doit aussi satisfaire deux autres conditions :

(i) Condition de moyenne
Si A, B dans ℓ' sont des contraires alors
$A \succ B$ implique $A \succ (A \vee B) \succ B$
Et $A \approx B$ implique $A \approx (A \vee B) \approx B$

(ii) Condition d'impartialité

Si A, B et C dans ℓ' sont des paires de contraires et,

A ≈ B mais ~A ≈ C, et (A ∨ C) ≈ (B ∨ C), alors pour tout D dans ℓ', qui est le contraire de A et B, (A ∨ D) ≈ (B ∨ D).

Enfin, la relation de préférence doit être continue c'est-à-dire que si $\{A_n\}$ est une suite monotone croissante ou décroissante dans ℓ' et qu'elle converge vers A, c'est-à-dire que $A = \bigcup A_n$ et $A = \bigcap An$, et que $B \succ A \succ C$ alors il existe un $n \in \mathbb{N}$(l'ensemble des entiers naturels) tel que $B \succ A_n \succ C$.

Les deux conditions de moyenne et d'impartialité ressemblent respectivement aux axiomes d'indépendances faible et forte. On détaillera ce parallélisme dans la section suivante. Pour le moment, il s'agit de présenter les théorèmes d'existence et d'unicité du modèle de Bolker-Jeffrey.

Théorème d'existence : si la relation de préférence est continue, transitive et trichotomique sur ℓ' et qu'elle vérifie la double condition de moyenne et d'impartialité, alors il existe une fonction $U : \ell' \rightarrow \mathbb{R}$ et une mesure de probabilité $\mu : \ell' \rightarrow [0,1]$ tels que :

1) $U(A) \geq U(B)$ si et seulement si $A \gtrsim B$
2) Pour tous A et B tels que A se décompose en deux propositions contraires A_1 et A_2 définies par $A_1 = A \wedge B$ et $A_2 = A \wedge \sim B$ (d'où $A = A_1 \vee A_2$), $U(A) = \frac{p(A_1)}{p(A)} U(A_1) + \frac{p(A_2)}{p(A)} U(A_2)$ sachant que $\frac{p(A_1)}{p(A)}$ et $\frac{p(A_2)}{p(A)}$ correspondent respectivement aux probabilités de A_1 et A_2 conditionnées à A.

La condition de linéarité en probabilité indique que ce théorème est une version particulière de l'utilité espérée.

Théorème d'unicité :

Soient μ , μ ’ des mesures de probabilités et ν , ν ’ des mesures signées sur une algèbre booléenne complète sans atome. μ, ν représentent les mêmes préférences que μ', ν ’ si et seulement si

$$\nu \text{ ’} = a\nu + b\mu$$

$$\mu \text{ ’} = a\nu + d\mu$$

Où $ad - bc > 0$

$c\nu(T) + d = 1$

Et $c\nu(A) + d\mu(A) > 0$ pour tout A dans ℓ’.

La transformation de μ, ν en μ', ν ’ transforme l’utilité U en

$$U' = \frac{\nu'}{\mu'} = \frac{a\nu + b\mu}{c\nu + d\mu} = \frac{aU+b}{cU+d}$$

Comme on l’a dit, ces conditions sont nécessaires pour démonter qu’il existe au moins un coupe ($p(\cdot)$, $d(\cdot)$) qui vérifie les conditions d’existence. Les transformations de U permises par la théorie de Bolker-Jeffrey sont des transformations linéaires fractionnelles. La classe des transformations est donc plus petite que dans les autres versions de l’utilité espérée de vNM, Ramsey et Savage. Ceci s’explique directement par le type d’objets sur lesquels portent les préférences.

3.1.2. Jeffrey, von Neumann et Morgenstern et Savage

Il existe des points de ressemblance entre le modèle de Jeffrey et ceux de vNM et Savage (3.1.2.1) comme les conditions de moyenne et d’impartialité. Toutefois en introduisant les propositions dans l’analyse, Jeffrey propose un modèle original (3.1.2.2).

3.1.2.1 Les conditions de moyenne et d’impartialité et les axiomes d’indépendance

La condition de moyenne évoquée plus haut est l’analogue de l’axiome d’indépendance présenté par Savage [1954].

En effet, comme nous l'avions vu dans la première partie de la thèse (chapitre 2), l'axiome d'indépendance, dans la version moderne de Luce et Raiffa [1957] par exemple, impose que si deux loteries sont jugées indifférentes par un individu, alors le fait de combiner chacune de ces deux loteries avec une troisième ne changera pas la relation d'indifférence entre les deux premières loteries. De même, dans l'exemple de l'homme d'affaires évoqué dans la section 2.2 de la partie I, Savage présente l'axiome d'indépendance comme le fait de ne pas considérer que l'issue de la prochaine élection présidentielle va influencer le choix de l'homme d'affaires entre investir et ne pas investir.

L'autre idée sous-jacente à la condition de moyenne est que si une proposition A est préférée à une proposition B, l'union de ces deux propositions va se situer à sur un point de l'échelle de désirabilité l'individu entre la désirabilité de A et la désirabilité de B, comme chez vNM et Ramsey. Seulement, comme le mentionne Broome [1990], dans la mesure où les propositions conservent leurs probabilités respectives lorsqu'elles sont combinées (contrairement aux loteries chez vNM par exemple), il n'est pas possible de proposer à l'individu une situation dans laquelle les deux propositions A et B sont combinées avec des probabilités choisies arbitrairement. Pour cette raison, Broome considère que la condition de moyenne est plus restrictive que l'axiome d'indépendance (Broome [1990], p.483).

La condition d'impartialité vise essentiellement à vérifier si deux propositions (A et B) sont jugées également probables par un individu. L'idée est ici de former des disjonctions avec une troisième proposition (C) qui n'est pas jugée indifférente aux deux premières. Si l'individu pense que les deux disjonctions ($A \vee C$ et $B \vee C$) sont aussi probables l'une que l'autre, alors c'est qu'il estime que A et B sont également probables[151].

[151] Cependant, comme le souline Broome [1990], cette condition, telle qu'elle vient d'être présentée, peut être considérée comme une pétition de principe puisqu'elle utilise l'utilité espérée comme explication et comme résultat. En effet, en postulant que l'utilité d'une disjonction est la moyenne de l'utilité des éléments disjoints, pondérée par leurs probabilités, la condition présuppose la linéarité en probabilités, c'est-à-dire l'utilité espérée, et constitue dans le même temps, une condition permettant de dériver la fonction d'utilité que l'individu va maximiser sous le critère de l'utilité espérée.

3.1.2.2 Différences entre le modèle de Bolker-Jeffrey et les modèles de von Neumann et Morgenstern et de Savage

Dans la théorie de vNM, comme on l'a vu, les perspectives auxquelles sont confrontés les individus - et donc les objets sur lesquels portent les préférences - sont des loteries. Chez Savage, ces perspectives sont des actions alors que chez Jeffrey ce sont des propositions.

La différence principale entre le modèle de Bolker-Jeffrey et les modèles de vNM et de Savage tient en une idée : lorsque les possibilités ou perspectives (*prospects*) sont combinées par l'opération de disjonction elles conservent leurs propres probabilités, rendues conditionnelles par la disjonction.

Dans la théorie de vNM, comme l'explique Bolker en 1967 lorsque le sujet a déjà une idée de ses probabilités subjectives, il ne sert à rien de lui proposer un pari. Celui-ci entrainerait en effet une modification de probabilités des perspectives. Pour Bolker et Jeffrey, la forme du pari influence les probabilités subjectives associées aux propositions. C'est pourquoi les auteurs ne proposent pas d'analyser les loteries ou paris auxquels doivent faire face des individus mais uniquement les probabilités et désirabilités associées à des combinaisons frai-faux de propositions.

Pour mieux comprendre cette idée, on peut reprendre à profit l'exemple proposé par Bolker [1967] :

Imaginons un individu qui a la possibilité de choisir pour son emploi du temps de demain trois activité : soit de nager (activité notée s), soit de jouer dans un quatuor à cordes (q), soit de jouer de l'alto seul (v) demain.

Bolker propose de considérer que l'individu pense qu'il y a 60% de chance qu'il y ait un temps clément, de telle sorte à ce qu'il puisse nager, et que s'il pleut , il y a 50% de chance que son quatuor ne se réunisse pas. On peut déduire de ces estimations faites par le sujet que la mesure des probabilités subjectives de ce dernier est -respectivement pour s, q et v – 0.6 ; 0.2 et 0.2. Dès lors, selon Bolker, si l'on demande à l'individu la désirabilité du pari (s, v ; ½, ½), il se pourrait que l'individu n'est pas d'opinion puisque les probabilités en jeu dans ce pari ne sont pas cohérentes avec sa propre estimations des probabilités respectives de chaque activité.

L'objection est la suivante : si l'on considère que les probabilités ½, ½ du pari proposé sont liées au jet d'une pièce de monnaie telle que si elle tombe sur face l'individu pourra nager et si elle tombe sur pile l'individu pourra jouer de l'alto, l'individu peut ne pas considérer qu'il y a un lien entre le jet de la pièce et le type de temps qu'il fera demain (beau temps, pluie…). Dès lors, l'individu n'aura pas d'opinion sur le pari et il ne sera pas possible de construire une mesure de la désirabilité comme vNM et Ramsey le proposent.

C'est pourquoi le modèle proposé par Bolker-Jeffrey appréhende la préférence comme un désir pondéré de manière à prendre en compte le monde tel qu'il est. Le problème du pari proposé est que celui-ci constitue une loterie purement arbitraire et non pas un questionnement relatif à la préférence d'un sujet par rapport à des événements.

Comme le souligne Broome [1990], les théories de vNM et Ramsey notamment combinent des résultats ou conséquences en formant des paris. Cela implique une attribution artificielle de probabilité ou un état de la nature qui a sa propre probabilité pour chaque résultat. Un pari constitue donc une modification artificielle des relations causales du monde. Dès lors, proposer un pari de ce type reviendrait à proposer un pari impossible à l'individu, ce que Jeffrey tente d'éviter. En posant dès le départ que les perspectives conservent leurs probabilités, Jeffrey dépasse cette difficulté et cette avancée théorique constitue pour lui l'avantage central de sa théorie (Jeffrey [1983], p. 157).

3.2. Le modèle de Davidson (1980)

Si Davidson choisit de reprendre l'essentiel du modèle de Jeffrey, il s'en démarque toutefois puisqu'il fonde sa théorie non pas sur des propositions mais plutôt sur des phrases non interprétées.

Pour lui, en effet, les propositions peuvent être assimilées aux significations des phrases. Or, comme nous l'avons mentionné, Davidson refuse de considérer que les significations sont données à l'avance. Partir des significations pour arriver aux significations relèverait de la pétition de principe, on postulerait ce que l'on veut

expliquer. D'autant que l'interprétation radicale[152] implique de partir de données fondamentales, les moins « contaminées » (Davidson [1974,1993b], p.273) possibles par la théorie : « si nous savons quelles propositions un agent choisit parmi d'autres, notre problème original d'interprétation du langage est une fois de plus supposé résolu. Ce dont nous avons besoin est d'obtenir les résultats de Jeffrey en ne disposant que de préférences parmi des phrases non interprétées » (Davidson [1980, 2004], p. 160).
Le fait de considérer des phrases non interprétées plutôt que des propositions ne change rien aux étapes du raisonnement – permettant de découvrir les désirabilités et probabilités subjectives – proposées par Jeffrey. Il s'agira toutefois de déterminer une variable de plus, les significations, variable permettant de construire une théorie pour l'interprétation des phrases.
Ainsi, l'élément primitif qui sert de point d'appui à la méthode est la préférence (faible) qu'une phrase plutôt qu'une autre soit vraie.

Nous revenons donc sur la manière donc Davidson cherche à déterminer les utilités et les probabilités (3.2.1) avant d'analyser la manière dont les significations, à leur tour, sont codéterminées en nous appuyant sur la philosophie du langage de Davidson (3.2.2).

3.2.1. La détermination des utilités et des probabilités

Le modèle de Davidson décalqué de celui de Jeffrey se présente alors comme suit.
Davidson présente un axiome analogue à l'axiome de désirabilité proposé par Jeffrey mais cette fois-ci en substituant aux propositions des phrases non interprétées.

Axiome de désirabilité :

Si prob $(s_1 \wedge s_2) = 0$ et prob $(s_1 \vee s_2) \neq 0$, alors

$$\text{des}(s_1 \vee s_2) = \frac{(\text{prob}(s_1)\text{des}(s_1) + \text{prob}(s_2)\text{des}(s_2))}{(\text{prob}(s_1) + \text{prob}(s_2))}$$

[152] Nous reviendrons en détails sur cette expression sans la section 3.2. Pour le moment considérons que l'interprétation radicale est la situation où l'interprète ne connait pas le langage de l'individu qu'il cherche à comprendre et interpréter, et qu'en ceci le processus d'interprétation débute avec des données minimales, les énonciations du locuteur dans une langue qui n'est pas connue de prime abord.

Ici, s_1 et s_2 sont des phrases et « prob (s_1) » correspond à la probabilité subjective de s_1 et « des (s_1) » à la désirabilité ou utilité de s_1. Comme cela avait été mentionné avec Jeffrey, le pari n'est pas directement présent dans cet axiome. Cependant, on peut intuitivement considérer – comme le propose Davidson – qu'un agent qui choisit de rendre vraie plutôt que fausse la phrase relative à une action « *L'agent parie un dollar* », prend un pari sur le résultat – par exemple, « *L'agent gagne cinq dollars* » – qui peut, par exemple, être considéré lié à un événement du type « *La prochaine carte est un cœur* ».

Si, par exemple, l'agent croit qu'il gagnera cinq dollars si la prochaine carte est un cœur et qu'il ne gagnera rien si ce n'est pas un cœur, il aura alors un intérêt spécial au fait que la vérité de « *L'agent parie un dollar* » aille de pair avec la vérité ou la fausseté de « *La prochaine carte est un cœur* ». On peut noter ces deux phrases par les lettres « s_1 » et « s_2 » et obtenir une formulation de la désirabilité qui renvoie à un pari :

$$\text{des } (s_1) = \frac{(\text{prob } (s_1 \wedge s_2) \text{ des } (s_1 \wedge s_2) + \text{prob } (s_1 \wedge \sim s_2) \text{des } (s_1 \wedge \sim s_2))}{\text{prob } (s_1)}$$

Cette formulation renvoie à un pari.

Imaginons le cas spécial où $s_2 = \sim s_1$. Il s'ensuit que :

(1) $\text{des } (s_1 \vee \sim s_1) = \text{des } (s_1) \text{ prob } (s_1) + \text{des } (\sim s_1) \text{ prob } (\sim s_1)$

Puisque $\text{prob } (s_1) + \text{prob } (\sim s_1) = 1$, (1) peut s'écrire :

(2) $$\text{prob } (s_1) = \frac{\text{des } (s_1 \vee \sim s_1) - \text{des } (\sim s_1)}{\text{des } (s_1) - \text{des } (\sim s_1)}$$

Si, comme on l'a vu avec Jeffrey, on assigne le chiffre 0 à n'importe quelle vérité logique T, on peut réécrire (2) :

(3) $$\text{prob } (s_1) = \frac{1}{1 - \frac{(\text{des}(s_1)}{\text{des } (\sim s_1)}}$$

Cette écriture permet de mettre en évidence ce que Jeffrey appelle une « bonne » option. En effet, on a vu que la désirabilité de toute proposition A pouvait s'écrire (à l'aide de notations simplifiées) :

des (T) = prob(A) des (A) + (1 – prob(A)) des (~A)

Si on a prob A = 1, la proposition A correspondra à la proposition nécessairement vraie T. Si on a si prob A = 0, la proposition A correspondra à la proposition nécessairement fausse F. Si prob A ≠0 et prob A ≠ 1, alors la proposition nécessairement vraie se situera à mi chemin entre A et ~A. En effet, on a vu que T = A∨~A, on peut donc imaginer que la proposition nécessaire correspond à un pari sur A où le gain est des A – desT et la perte des T- des ~A. On peut donc en déduire, selon Jeffrey, que A est bon si des A > des T et A est une mauvaise option si des A < des T.

Cela implique, d'après (3) que toute phrase et sa négation ne peuvent pas être toutes deux bonnes ou toutes deux mauvaises.

Puis Davidson introduit ce qu'on appelle traditionnellement en logique l'incompatibilité.

L'incompatibilité est la négation de la conjonction. On utilise la notation $s_1|s_2$ ce qui se lit « s_1 incompatibilité s_2 ». $s_1|s_2$ prend la valeur faux dans un seul cas et un seul : lorsque s_1 et s_2 prennent l'un et l'autre la valeur vrai. On traduit généralement l'incompatibilité par la phrase : « on n'a pas à la fois s_1 et s_2 ».

Dans le cas spécial où s_2= ~ s_1 on peut écrire l'incompatibilité comme « ~ (s_1 et ~ s_1) ».

Appliqué aux préférences, on obtient :

Si des (s_1) > des (~(s_1 et ~ s_1)) alors
des (~(s_1 et ~ s_1)) ≥ des (~s_1) et
Si des (~(s_1 et ~ s_1)) > des (s_1)
des (~s_1) ≥ des (~(s_1 et ~ s_1))

Davidson propose d'utiliser le connecteur barre de Scheffer « | » pour décrire la négation de la conjonction. La raison pour ce choix est claire. La particularité de la barre de Scheffer est qu'elle est un connecteur universel (Sheffer [1913]), c'est-à-dire un

connecteur à partir duquel on peut déterminer tous les autres connecteurs comme la conjonction, la disjonction et la négation. Dès lors, s'il est possible de déterminer en premier lieu la barre de Scheffer, il sera possible, tel que Scheffer la prouver, de déduire tous les autres connecteurs.
L'incompatibilité peut se reformuler comme suit :

Si des (s_1) > des ((t|u) | ((t|u) | (t|u)))
des ((t|u) | ((t|u) | (t|u))) ≥ des (s_1| s_1), et
si des ((t|u) | ((t|u) | (t|u))) > des (s_1) alors
des (s_1| s_1) ≥ des ((t|u) | ((t|u) | (t|u)))[153].

La spécificité de cette écriture est visible lorsqu'on présente les tables de vérité (initialement introduite par Wittgenstein [1922]) des connecteurs.
En logique, on considère qu'un énoncé ou une phrase est une expression susceptible d'être vraie ou fausse. Tout énoncé a une valeur de vérité et une seule : soit vrai, soit faux. Cela revient à dire que l'on adopte le principe de bivalence.
Par exemple, pour définir l'opérateur logique appelé négation, nous posons que son effet est d'inverser la valeur de vérité de la proposition à laquelle on l'applique. Si l'on note ~ s_1 pour désigner la négation de s_1, on peut représenter la négation dans la table de vérité suivante :

s_1	~ s_1
V	F
F	V

Plus généralement, on peut représenter les connecteurs (respectivement l'intersection ∧, l'union ∨, l'implication → et l'incompatibilité |) par le tableau suivant :

[153] Les lettres t et u correspondent à des phrases.

P	Q	(p ∧ q)	(p ∨ q)	(p→q)	p|q
V	V	V	V	V	F
V	F	F	V	F	V
F	V	F	V	V	V
F	F	F	F	V	V

Tableau 12

La dernière colonne nous intéresse tout particulièrement puisqu'elle correspond à l'incompatibilité c'est-à-dire au connecteur universel que cherche à déterminer Davidson. Comme le tableau l'indique, (p|q) prend la valeur faux dans un seul cas et un seul : lorsque p et q prennent l'un et l'autre la valeur vrai.

A partir de la reformulation des axiomes de Jeffrey et de l'introduction de ces connecteurs, l'idée de Davidson est de procéder de la même manière que Jeffrey quant au calcul des désirabilités et des probabilités. Comme nous l'avons vu dans la section 3.1.2, le chercheur ou peut avoir accès aux probabilités et désirabilités notées p(.) et d(.) que l'agent attribue aux 2^n combinaisons de « vrai-faux » pour les propositions (conjonctions) de la forme *AB…V* (à *n* propositions élémentaires) par exemple. A partir de ces données, nous avons montré comment, selon Jeffrey, il est possible de calculer les désirabilités et les probabilités qu'attribue l'individu à chaque proposition isolément mais aussi à toutes les combinaisons de propositions. Davidson reprend cette méthode : « Il est maintenant possible de mesurer la désirabilité et la probabilité subjective des toutes les phrases, car l'application des formules comme (2) et (3) nécessite seulement l'identification des connecteurs de phrases vérifonctionnels » (Davidson [1980, 2004], p. 164).

Seulement, « à ce point, les probabilités et désirabilités de toutes les phrases ont été en théorie déterminées […] Il reste à esquisser les méthodes qui mènent à une interprétation complète de toutes les phrases, c'est-à-dire la construction d'une théorie de la vérité pour le langage d'un agent. L'approche est l'une de celles que j'ai discutées dans de nombreux articles et est inspirée du travail de Quine sur la traduction radicale » (*ibid.*).

La prochaine étape consiste donc à déterminer les significations à partir d'une théorie de l'interprétation inspirée des travaux de Quine et de Tarski sur lesquels nous revenons maintenant.

3.2.2 Philosophie du langage de Davidson

Une fois que les probabilités et désirabilités de toutes les phrases ont été déterminées, il reste à déterminer, selon Davidson, les significations des phrases.

En effet, même si les connecteurs vérifonctionnels ainsi que les phrases logiquement vraies ou fausses ont été identifiés, aucune phrase complète n'a été interprétée.

Davidson utilise ici ses recherches en philosophie du langage et plus précisément ce que l'on appelle habituellement le « programme de Davidson » ainsi que « l'interprétation radicale » – inspirée des travaux de Quine [1960] qui s'y rattache : « Il reste à esquisser les méthodes qui mènent à une interprétation complète de toutes les phrases, c'est-à-dire la construction d'une théorie de la vérité pour le langage d'un agent. L'approche est l'une de celles que j'ai discutées dans de nombreux articles et est inspirée du travail de Quine sur la traduction radicale » (Davidson [1980, 2004], p. 164).

Le programme de Davidson renvoie à toutes les questions relatives à la forme que devrait prendre une théorie satisfaisante de la signification des propos d'un locuteur.

Cette question de la forme est notamment discutée dans l'article de Davidson publié en 1967, « Vérité et Signification ». Dans cet article, Davidson tente d'apporter une réponse à la question de savoir quelle théorie pourrait fournir une analyse « de la manière dont les significations des phrases dépendent de la signification des mots » (Davidson [1967, 1993b], p. 41).

Parmi les théories satisfaisantes qui pourraient proposer une analyse comme celle-ci, Davidson discute initialement une théorie qui produirait des phrases de la forme « *s* signifie *m* » où *s* est la description structurale de la phrase - c'est-à-dire la description d'une expression comme une « concaténation d'éléments pris dans une liste finie fixe (par exemple de mots ou de lettres » (Davidson [1967, 1993b], p. 43) - et *m* un « terme singulier désignant la signification de cette phrase » (Davidson [1967, 1993b], p. 45). Le problème de cette théorie dont les conséquences seraient des énoncés de la forme « *s*

signifie *m* » est non seulement qu'elle fait usage de significations (pour *m* par exemple) alors qu'elle est plutôt sensée les déduire mais aussi que paradoxalement, l'usage de significations n'est d'aucune aide pour construire une théorie qui détermine le sens d'une phrase à partir du sens de ses parties composantes, les mots. Plus précisément, comme le mentionne Davidson : « la seule chose que les significations ne fassent pas est huiler les roues d'une théorie de la signification – tout au moins tant que nous requérons d'une telle théorie qu'elle nous fournisse de manière non triviale la signification de chaque phrase dans le langage » (Davidson [1967, 1993b], p. 46). Comme on le verra, l'une des solutions proposées par Davidson pourrait consister à remplacer le « *m* » de « *s* signifie *m* ». Il s'agirait d'écrire par exemple « *s* signifie que *p* » où *p* est une phrase. Là encore, Davidson ne semble pas satisfait, du fait notamment de l'expression « signifie que ». L'auteur considère que la « théorie aura fait son travail si elle fournit, pour chaque phrase *s* du langage étudié, une phrase correspondante (remplaçant « p ») qui, d'une certaine manière encore à clarifier, « donne la signification » de « s » » (Davidson [1967, 1993b], p. 49).

Le problème de cette théorie est qu'elle postule en quelque sorte ce qu'elle tente d'expliquer en utilisant l'expression « signifie que », ce qui réintroduit subrepticement des significations. Nous sommes face ici à une pétition de principe. Davidson va analyser plusieurs de ces théories pour enfin accéder à celle qui lui semble la plus à même de remplir l'objectif qu'il s'est fixé : analyser la manière dont les significations des phrases dépendent de la signification des mots sans postuler à l'avance les significations que l'on tente de déterminer.

Comme on va le voir, cette théorie que choisit finalement Davidson, est une théorie inspirée de celle de Tarski pour les langages formels.
Ce « programme de Davidson » est lié conceptuellement et méthodologiquement à ce que l'on appelle l' « interprétation radicale », terme emprunté à Quine.

L'interprétation radicale correspond aux conditions empiriques que doit remplir une théorie de l'interprétation : « la théorie est vraie si ses implications empiriques sont vraies ; nous pouvons tester la théorie en faisant l'échantillonnage des implications

qu'elle a pour la vérité. Cela veut dire remarquer si oui ou non les interprétations typiques que fournit une théorie des énonciations d'un locuteur sont correctes » (Davidson [1974, 1993b], p.209).

Comme à tous les niveaux de la théorie unifiée, l'objectif est d'éviter la pétition de principe en postulant dès le départ les concepts que l'on tente de déterminer. Or le concept de signification qui fonde toute compréhension du langage présuppose a un contenu trop riche de la même manière que les utilités cardinales constituent une information plus détaillée que les utilités ordinales. Ainsi, il convient de partir d'un concept plus primitif, à partir duquel la signification pourra être déterminée. En donnant à la vérité un rôle central dans la construction d'une théorie de l'interprétation, Davidson renoue avec la conception véri conditionnelle de la signification (la signification est donnée par ses conditions de vérité) de Frege.

La particularité du concept de vérité selon Davidson est « qu'on ne peut espérer l'étayer avec quelque chose de plus transparent ou de plus simple à comprendre » (Davidson [1996], p. 265).

Précisément, Davidson ne cherche pas à définir la vérité mais à la « traiter […] comme primitive et d'extraire une analyse de la traduction ou de l'interprétation» (Davidson [1973,1993b], p. 199).

Davidson va prendre comme modèle la théorie de Tarski [1956] car il considère que sous certaines conditions, les méthodes utilisées par ce dernier pour définir la vérité pour les langues formelles peuvent être transposées - au moins en partie et moyennant une redéfinition conceptuelle de la théorie - aux langues naturelles. L'objectif de Davidson est de renverser, comme on va le voir, le schéma tarskien : « Ce que je propose est de renverser l'ordre de l'explication : en supposant la traduction donnée, Tarski était capable de définir la vérité ; l'idée est de traiter la vérité comme primitive et d'extraire une analyse de la traduction ou de l'interprétation » (Davidson [1973,1993b], p. 199).

L'autre emprunt majeur de Davidson pour construire cette théorie de l'interprétation est celui fait à Quine [1960].

De la théorie de Quine, Davidson tire trois éléments par rapport auxquels il se démarque[154] :

- Pour Quine comme pour Davidson, l'objectif d'une théorie de l'interprétation radicale[155] est de déterminer les significations d'un locuteur : « Nous interprétons un fragment de comportement linguistique quand nous disons ce que veulent dire les mots d'un locuteur en telle circonstance où il les emploie. On peut estimer qu'il s'agit là d'un travail de redescription » (Davidson [1974, 1993b], p. 208).

  Cependant là où « Quine décrit les événements ou les situations en termes de modèles de stimulation », Davidson « préfère une description qui se fasse dans les termes qui ressemblent plus à ceux de la phrase que l'on étudie » (Davidson [1979, 1993b], p 332). Autrement dit, là où Quine fonde la traduction sur des prémisses behavioristes comme des stimulations sensorielles, Davidson a recourt à des attitudes intentionnelles minimales comme celle de tenir pour vrai pour résoudre le problème de l'interprétation
- Il emprunte aussi à Quine sa méthode indirecte qui suggère de noter les « conditions dans lesquelles le locuteur étranger donne son assentiment ou son dissentiment à toute une série de phrases » (Davidson [1970, 1993b], p.103) pour analyser la notion de signification car « étant donné qu'une définition de la vérité détermine la valeur de vérité de chaque phrase du langage objet (relativement à une phrase du métalangage), elle détermine la signification de chaque mot et de chaque phrase » (Davidson [1967, 1993b], p.51). Seulement, comme cela a été mentionné dans le premier point, ce repérage de l'assentiment ou du dissentiment ne se fait pas sur la base des mêmes données pour Quine et Davidson : Quine les fonde sur des éléments behavioristes alors que Davidson tente de repérer une certaine attitude propositionnelle.
- Tous deux s'accordent aussi sur l'indétermination conséquente de l'interprétation : « Pourtant Quine a raison, me semble-t-il, de soutenir qu'il restera un important degré d'indétermination une fois introduites toutes les données empiriques » (Davidson [1970, 1993b], p.103). En effet, même si Davidson impose un certain

[154] Pour plus de détails sur la manière dont Davidson se démarque de Quine, voir Engel [1994].
[155] Quine parle plutôt de traduction radicale et cette différence de termes révèle une différence de projet.

nombre de restrictions sur les états mentaux comme par exemple les postulats de rationalité utilisés en théorie de la décision (transitivité, cohérence) ou en théorie de l'interprétation le principe de charité - que nous évoquerons plus loin – ces restrictions ne suffisent pas à déterminer les interprétations de manière unique et en cela elles sont indéterminées. Plus précisément, cette indétermination des interprétations renvoie à une idée similaire en théorie de la décision. Nous avions vu notamment que les fonctions d'utilité n'étaient pas définies de manière unique mais par une certain classe de transformations (affines ou linéaires) et ce malgré les différents axiomes de cohérence et de rationalité imposés. En théorie de l'interprétation, l'interprétation n'est pas déterminée de manière unique et ce malgré les conditions de cohérence et de compréhension nécessaires à l'interprétation. Parmi celles-ci on peut mentionner par exemple le principe de charité, qui peut être comme un principe qui maximise l'accord entre le locuteur et l'interprète.

Nous commencerons par envisager la forme que doit prendre une telle théorie en présentant à grands traits les composantes du programme de Davidson (3.2.2.1). A partir des conditions formelles énoncées, il sera possible de mettre en relief un certain nombre de conditions empiriques qui constituent l'interprétation radicale de l'auteur (3.2.2.2). Parmi ces conditions empiriques, nous insistons ensuite sur le principe de charité qui revêt une importance particulière dans l'œuvre de Davidson à plus d'un titre (3.2.2.3).

3.2.2.1. Le programme de Davidson

La question de la forme que doit prendre une théorie de l'interprétation du langage et plus précisément une théorie de la signification renvoie à une distinction fondamentale entre une théorie du concept de signification et une théorie de la signification au sens d'une analyse de la signification de « toute expression douée de sens » (Davidson [1970, 1993b], p.93).

Davidson propose d'unir les expertises du philosophe et du linguiste. En effet, alors que Davidson écrit, une division du travail s'opérait entre les travaux des philosophes portant sur le concept de signification et les travaux des linguistes portant notamment

sur une sémantique linguistique. Plus précisément, Davidson considère que l'analyse seule des significations manque l'objectif visé, l'interprétation : « Paradoxalement, la seule chose que les significations ne fassent pas est huiler les roues d'une théorie de la signification ». Or, comme le souligne Engel [1994], Davidson cherche à donner une réponse à la question de savoir ce qu'est le concept de signification en apportant une réponse à la question de savoir ce qu'est une théorie de la signification. La raison pour cela est donnée par Davidson : « une théorie de la vérité apporte une réponse précise, profonde et testable à la question de savoir comment des ressources finies suffisent à expliquer les capacités sémantiques infinies du langage » (Davidson [1970, 1993b], p.93).

Ainsi, une théorie de la signification doit prendre la forme d'une théorie de la vérité inspirée de celle proposée par Tarski afin d'éclairer en retour le concept de signification. Cette posture s'explique, comme on l'a dit, par le fait que Davidson se refuse de partir du concept même de signification.

Même avant de s'intéresser à la forme que doit prendre la théorie, il s'agit pour Davidson de définir un certain nombre d'objectifs que la théorie doit remplir, ces objectifs donnant corps au projet d'ensemble.

Les conditions formelles sont dépendantes de conditions relatives[156] au but recherché : une théorie acceptable de la signification. Une telle théorie doit, selon Davidson « rendre compte de la signification (ou des conditions de vérité) de toute phrase en l'analysant comme étant composée, sous diverses formes permettant la détermination des conditions de vérité, d'éléments pris dans un stock fini » (Davidson [1970, 1993b], pp.94-95). Autrement dit, les phrases sont composées de parties dont la signification détermine celle de la phrase toute entière. Cette structure des phrases par parties permet de mettre en relief leurs conditions de vérité (déterminées comme on va le voir par une théorie de vérité comme celle de Tarski).

Une seconde exigence est que la théorie « fournisse une méthode pour décider, étant donné une phrase quelconque, ce qu'est sa signification » (*ibid.*).

A partir de ces deux conditions, Davidson indique que la théorie « montre que le langage qu'elle décrit est *susceptible d'être appris et scrutable* ».

[156] Pour une présentation détaillée des conditions « constitutives », voir Engel [1994], pp. 8-12.

Enfin, une troisième condition : « les énoncés des conditions de vérité pour les phrases individuelles impliquées par la théorie devraient, d'une manière qui reste encore à préciser, faire appel aux mêmes concepts que les phrases dont ils énoncent les conditions de vérité » (Davidson [1970, 1993b], p.95). Autrement dit, il ne s'agit pas de réintroduire subrepticement des significations dans l'énoncé des conditions de vérité.

A partir de ces conditions, Davidson est en mesure de décrire la forme que devrait prendre une théorie de la signification pour les langues naturelles.
Premièrement, il s'agit de formuler ce qu'il est d'usage d'appeler un métalangage qui théorise ou traduit des énoncés d'un langage-objet.
Les expressions du langage-objet sont des descriptions structurales c'est-à-dire qu'elles décrivent l'expression comme une « concaténation d'éléments pris dans une liste finie fixe par exemple de mots ou de lettres) » (Davidson [1967, 1993b], p. 43).
La théorie recherchée s'appuie sur la théorie de la vérité pour les langages formels proposée par Tarski dans les années 1950.
La raison de cet emprunt est la théorie de Tarski[157] rempli les conditions évoquées plus haut même si elle est relative à un langage formel. La difficulté principale d'un tel transfert de méthode est qu'il se heurte à l'objection qu'un langage formel n'est pas un langage naturel.
La particularité d'un langage naturel est que les nombreuses phrases qui le composent « varient en valeur de vérité selon le temps où elles sont émises, le locuteur et même, peut-être, le public auquel elles s'adressent » (Davidson [1970, 1993b], p. 97). Il sera nécessaire dès lors, pour tout interprète voulant comprendre quelles phrases un individu tient pour vraies, de spécifier l'environnement dans lequel prend place ces phrases (temps, lieu, et sujet qui énonce ces assertions).

Dans son article « The Semantic Conception of Truth and the Foundations of Semantics» publié en 1944, Tarski tente de donner une définition satisfaisante de la notion de vérité c'est-à-dire une définition à la fois correcte formellement (une définition qui stipule un certain nombre de règles et de notions corrélées à la notion de

[157] Pour plus de développements sur l'emprunt de Davidson à Tarski voir Engel [1989] et [1994], Rivenc [1998].

vérité) et matériellement adéquate (l'idée est d'appréhender la vérité non pas comme une notion nouvelle mais comme une notion dont on fait usage quotidiennement, tout spécialement lorsqu'on se réfère aux phénomènes psychologiques) (Tarski [1944], pp. 341-342).

Tarski propose une définition de la vérité qui, selon lui, remplit les exigences mentionnées.

L'exemple qu'il donne est le suivant :

Considérons la phrase suivante : « la neige est blanche ». Une théorie de la vérité doit pouvoir dire dans quelles conditions cette phrase est vraie ou fausse.

Ainsi pour l'exemple mentionné, l'équivalence suivante coïnciderait avec la définition recherchée :

La phrase « la neige est blanche » est vraie si et seulement si la neige est vraie.

Lorsque l'on étend cet exemple au cas général, la définition devient :

(T) *X* est vrai si et seulement si *p*

X peut être remplacé par un nom d'un énoncé du langage-objet et *p* un énoncé auquel le prédicat vrai se réfère. Ceci correspond à la condition d'adéquation matérielle évoquée plus haut : c'est la Convention T. Cette convention enjoint de transcrire tout énoncé du langage-objet sous la forme d'équivalence-T (comme *X* est vrai si et seulement si *p*). Chaque équivalence est une définition partielle de la vérité. Toutes les équivalences-T forment une définition adéquate de la vérité (Tarski [1944], pp. 344) car ces caractérisations sont appliquées récursivement.

La difficulté que rencontre Davidson est que l'énoncé qui est à droite du biconditionnel – si et seulement si – doit être la traduction de l'énoncé de droite (dans le cas où le métalangage et le langage-objet sont distincts. Or, lorsqu'il est question de rapporter des attitudes propositionnelles comme des croyances, il est trompeur de considérer la notion de traduction. En effet, dans la phrase « B croit que *p* » le fait de savoir qu'elle est vraie ne donne pas d'information sur la signification qu'elle contient car on peut tenir pour vraie la phrase « L'administration simultanée de diclofénac et méthotrexate nécessite une surveillance rigoureuse », c'est-à-dire croire qu'elle est vraie si je l'ai lu sur la posologie d'un médicament, sans savoir pourtant ce qu'elle signifie.

L'objectif est alors de préserver la fonction que remplit la traduction dans la théorie de Tarski mais en supposant le moins possible une telle notion : « La théorie aura fait son

travail si elle fournit, pour chaque phrase du langage étudié, une phrase correspondante (remplaçant *p*) qui, d'une certaine manière qui doit être clarifiée, « donne la signification » de *s* » (Davidson [1967, 1993b], pp. 49-50). Comme on l'a dit, l'usage de la notion même de signification est proscrit. Il faudra se contenter de données minimales en inversant le mécanisme de la théorie : « Ce que je propose est de renverser l'ordre de l'explication : en supposant la traduction donnée, Tarski était capable de définir la vérité ; l'idée est de traiter la vérité comme primitive et d'extraire une analyse de la traduction ou de l'interprétation » (Davidson [1973, 1993b], p. 199).
Ainsi, là où Tarski s'intéresse à une définition de la vérité, Davidson ne recherche qu'une théorie de celle-ci (Davidson [1974, 1993b]. Cette théorie doit être « relativiser à des temps et à des locuteurs (et peut-être à bien d'autres choses) » (*ibid.*).

Après avoir mentionné les conditions constitutives et formelles d'une théorie de la signification, il reste à préciser quelles sont les conditions empiriques.

3.2.2.2. L'interprétation radicale

Le terme « interprétation radicale » est une manière pour Davidson de signifier son emprunt à la « traduction radicale » de Quine. L'ouvrage de Davidson, *Enquêtes sur la vérité et l'interprétation*, commence d'ailleurs par cette phrase : « à W.V. Quine sans qui ce livre ne serait pas ». Mais les deux expressions - « interprétation radicale » et « traduction radicale » - ne sont pourtant pas identiques.
Pour Quine, nous sommes amenés à utiliser un procédé de traduction dans au moins trois types de situations :

- La première est celle où les langues sont apparentées c'est-à-dire lorsque les langues, ne serait-ce que par leurs mots, se ressemblent (Quine [1960, 1977], p. 60)[158].
- La seconde situation est celle où les langues ne sont pas apparentées mais lesquelles il existe « certaines tables de concordance »[159] (*ibid.*).

[158] Quine donne l'exemple du frison et de l'anglais *(ibid.*).
[159] Quine donne l'exemple du hongrois et du français *(ibid.*).

- Enfin, troisième situation, la « traduction radicale proprement dite », c'est-à-dire lorsqu'il s'agit de traduire « la langue d'un peuple resté jusqu'ici sans contact avec notre civilisation » (*ibid.*).

Dans ce dernier cas, la traduction est dite radicale car nous ne disposons au départ d'aucun élément nous permettant de construire un manuel de traduction, mis à part les énoncés verbaux de l'indigène ainsi que les mouvements de son corps. Comme Quine l'explique, « les secours apportés par les interprètes sont plus pauvres » (*ibid.*).

Cette traduction radicale est illustrée par Quine par un exemple célèbre où un linguiste tente de construire un manuel de traduction pour un langage étranger qu'il ne comprend pas. Quine imagine la situation où un lapin détale non loin d'un l'indigène - dont le langage est l'objet de recherche du linguiste - et qui, en voyant cette scène, dit « Gavagai ». Le linguiste pourrait indiquer dans son manuel de traduction que « Gavagai » signifie « Lapin » ou « Tiens, un lapin » en s'appuyant sur le fait que face à stimulus extérieur – le lapin qui détale -, l'indigène répond verbalement par « Gavagai » (Quine [1960, 1977], p. 61). Pour vérifier la validité de son manuel de traduction composé dans les premiers temps de bribes de discours, la linguiste n'a pas d'autres choix que de « soumettre à l'approbation de son informateur des phrases de cette langue » (*ibid.*). Autrement dit, le linguiste doit recueillir les approbations et les désapprobations de l'indigène face à des phrases. Il lui faudra encore s'assurer de la manière dont l'indigène exprime une approbation sans pour autant avoir la possibilité de s'appuyer sur des codes ou autres habitudes de son propre langage car il ne pourrait être sûr que les gestes et autres mouvements corporels sont les mêmes pour lui et pour l'indigène. L'idée est donc de faire l'expérience des « élocutions spontanées » de l'indigène qui conduisent ce dernier à approuver ou désapprouver une phrase ou une question aussi brève que « Gavagai ? » lorsqu'un lapin détale non loin.

L'interprète tentera plus précisément, de construire une chaîne causal entre, par exemple, le lapin qui détale comme stimulus c'est-à-dire comme cause d'une réaction chez l'indigène et l'assentiment ou le dissentiment à la question « Gavagai ? » comme conséquence du stimulus (Quine [1960,1977], p. 62).

L'interprétation radicale de Davidson, même si elle constitue un emprunt à Quine, ne revêt pas une position théorique similaire comme Davidson le souligne lui-même : « Le terme 'd'interprétation radicale' est destiné à suggérer la parenté étroite qu'il a avec le terme quinien 'traduction radicale'. Toutefois, parenté ne veut pas dire identité, et 'interprétation' mis à la place de 'traduction' implique des différences : un plus grand accent mis sur ce qui est explicitement sémantique dans le premier cas » (Davidson [1973, 1993b], p.188). Autrement dit, le changement de terme correspond comme on l'a vu avec le renversement de Tarski, à un changement de priorité : l'objectif de Davidson n'est pas la traduction mais la compréhension.

Cependant, Davidson va reprendre la distinction utilisée par Quine entre les langues apparentées et langues inconnues par rapport à celle de l'interprète.

L'interprétation radicale, comme chez Quine, correspond à la situation où la langue de l'indigène est restée sans contact jusqu'alors avec la langue de l'interprète.

L'objectif est cette fois - en reprenant la méthodologie inspirée de Tarski – de construire une « procédure consistant à concevoir une théorie de la vérité pour une langue indigène » (Davidson [1973, 1993b], p. 202). Cette procédure a pour but d' « ajuster notre logique au nouveau langage, dans la mesure requise pour obtenir une théorie satisfaisante à la convention T » *(ibid.*). Autrement dit, l'idée est la suivante : si nous ne disposons pas au départ de données permettant directement de relier le langage de l'indigène et le langage de l'interprète (s'il n'est par exemple pas possible de dire qu'ils appartiennent à la même communauté linguistique), Davidson propose que l'interprète suppose, par exemple, que « ce qui pousse quelqu'un à avoir telle croyance, provoque la même croyance chez moi si la même cause [de celle-ci] survient » (Davidson [1986, 2004], p. 69). Cette méthode repose sur un principe essentiel : nous n'avons pas la possibilité d'assigner une attitude à quelqu'un si l'on pense que son rôle dans les pensées d'un autre est différent du rôle qu'elle joue dans les nôtres (*ibid.*). La procédure initiale peut donc consister à ajuster la logique du langage de l'interprète à celle du langage de l'indigène. Ce principe est d'ailleurs un principe récurrent de Davidson : de même qu'on ne peut donner sens aux attitudes propositionnelles d'un individu que si l'on considère que celui-ci est rationnel[160], on ne peut imaginer qu'un

[160] Sans que cela nous empêche de considérer des cas d'irrationalité.

langage est doté de mots et de phrases doués de sens que si nous postulons que ce langage est caractérisé par une logique qui a des points communs avec la logique de notre propre langage : « La présomption méthodologique de rationalité n'interdit pas d'attribuer des pensées et des actions irrationnelles à un agent, mais elle fait peser une contrainte sur ces attributions » (Davidson [1975, 1993b], p. 234).

La première étape de l'interprétation radicale consiste, selon l'auteur, à « introduire la logique de la théorie de la quantification du premier ordre (plus identité » c'est-à-dire d'identifier « les prédicats, les termes singuliers, les quantificateurs, les connecteurs et l'identité ; en théorie, elle règle les questions de la forme logique » (Davidson [1973, 1993b], p. 202).

La deuxième étape « est particulièrement consacrée aux phrases comportant des indexicaux ; ces phrases jugées tantôt vraies tantôt fausses, en fonction des changements que l'on peut découvrir dans le monde » (*ibid.*).

La dernière étape, selon Davidson, « s'occupe des phrases restantes, celles sur lesquelles ne se fait pas un accord unanime, ou celles dont la valeur de vérité estimée ne dépend pas systématiquement des changements qui peuvent se produire dans l'environnement » (Davidson [1973, 1993b], p. 203)

Mais l'interprétation radicale est aussi le moyen utilisé par Davidson pour décrire la nécessité d'interpréter en même temps plusieurs données mentales : les désirs, les croyances et les significations ; c'est-à-dire d'unir les analyses singulières de chacune de ces notions en une seule et même analyse qui ne se contente pas seulement de proposer une interprétation des mots d'un locuteur mais aussi une interprétation de ces états mentaux : « une théorie radicale de la décision doit inclure une théorie de l'interprétation et ne peut la présupposer » (Davidson 1974, 1993b], p. 217). Elle est « radicale » dans la mesure où une théorie « unifiée » du triplet désirs-croyances-significations se fonde sur des données observables minimales qui ne présupposent pas acquis les éléments qu'elle tente d'expliquer. Ainsi, par exemple, en théorie de la décision, l'observation des choix parmi des issues et plus précisément l'observation des préférences ordinales n'offre aucune information sur les intensités des préférences, c'est-à-dire sur les préférences cardinales. De même, en théorie de l'interprétation,

« tenir une phrase pour vraie » n'implique pas que la signification de cette phrase soit connue.

L'interconnexion entre les désirs, les croyances et les significations pose la question relative au point de départ de toute interprétation. En effet, s'il n'est pas possible de postuler, par exemple, des entités comme des significations pour produire une théorie de la signification c'est qu'il faut se contenter de données minimales qui ne présupposent pas ce que l'on cherche à définir et expliquer : « Puisque nous ne pouvons espérer interpréter l'activité linguistique sans savoir ce que croit un locuteur, et ne pouvons fonder une théorie de ce qu'il veut dire sur une découverte préalable de ses croyances et intentions, j'en conclus que lorsque nous interprétons des énonciations en partant de zéro – dans une interprétation *radicale* – nous devons en quelque sorte présenter simultanément une théorie de la croyance et une théorie de la signification » (Davidson [1974, 1993b], p.212). Cette posture théorique est la même lorsque l'on s'atèle à la théorie unifiée. Nous devons présenter dans ce cadre une théorie de la croyance, une théorie de la signification et une théorie du désir.

En effet, la théorie de Tarski et celle de Davidson sont des théories empiriquement testables et ce parce que la théorie de Tarski – et donc celle de Davidson – constitue une mesure du mental au même titre que la méthode opérationnelle proposée par Ramsey.

La convention T est une théorie testable (voir 1993b p. 102).Plus précisément, tous les théorèmes qui suivent la structure imposée par la convention T sont testables. La question est de savoir ce qu'entend Davidson par « testable » lorsque ce terme est relatif à une théorie de la signification pour une langue naturelle. L'idée centrale de Davidson est que l'interprète doit être « en mesure de reconnaître quand les biconditionnels requis sont vrais » (Davidson [1970, 1993b], p.102). Plus précisément, pour Davidson, ce test va de soi puisque toute tentative pour interpréter un langage implique une compréhension de celui-ci : « en principe, il n'est pas difficile de tester l'adéquation empirique d'une théorie de la vérité qu'il ne l'est pour un locuteur français compétent de décider si des phrases telles que « 'la neige est blanche' est vraie si et seulement si la neige est blanche » le sont » (*ibid.*). Autrement dit, les conditions de vérité d'une phrase

ne semblent pas poser de difficulté, du moins lorsque le langage du locuteur et celui de l'interprète sont identiques[161].

Ainsi par exemple, un schéma du type[162] :

La phrase *s* est vraie (en français) pour un locuteur *u* au temps *t* si et seulement si *p*

A pour rôle de « fournir un test d'adéquation d'une théorie de la vérité : une théorie acceptable doit impliquer une phrase vraie qui est la forme [du schéma ci-dessus] quelle que soit la phrase du français qui serait décrite par l'expression canonique remplaçant 's' » (Davidson [1969, 1993b], p. 81). Ainsi, un biconditionnel du type *X* est vrai si et seulement si *p* comme, par exemple la neige est blanche est vraie si et seulement si la neige est vraie constitue un d'adéquation d'une théorie de la vérité si par exemple elle est énoncé par le locuteur en *t* et si et seulement si la neige est blanche en *t*.

3.2.2.3. Principe de charité et normes d'interprétation

Parmi les conditions empiriques qui sous-tendent la théorie unifiée, il en est une qui revêt une importance considérable dans l'œuvre de Davidson, le principe de charité. Celui-ci est ainsi une norme de cohérence ou de rationalité qui tire son origine des travaux des philosophes Neil Wilson (1959) et Willard von Orman Quine (1960) (a). Toutefois Davidson lui faut jouer d'autres rôles dans sa théorie. Ainsi joue t-il pour l'interprétation celui des les propositions éthiquement neutres de la théorie de la décision. Il permet de fixer les croyances pour l'interprétation (b). Bien qu'il soit à la fois norme de cohérence et fixateur de croyances, il ne peut néanmoins conduire à lever l'indétermination qui caractérise toute forme de traduction (c).

(a) Le principe de charité comme norme de cohérence

[161] On pourrait dire lorsque le langage-objet, est contenu dans le métalangage, c'est-à-dire la théorie construire par l'interprète.

[162] Exemple de Davidson [1969, 1993b], p. 81.

Comme le souligne Isabelle Delpha [2001], dans sa forme la plus générale, le principe de charité est un principe de correction qui nous enjoint de « faire crédit aux autres » et pour cela de « chercher l'interprétation la plus favorable à leurs propos » (Delpha [2001], p.7). Créé dans les années 1950 par Neil Wilson, il sera surtout utilisé par Quine pour sa traduction radicale avant d'être repris par Davidson.

- L'héritage de Wilson et de Quine

On peut retracer l'origine de ce principe dans les travaux de Neil Wilson [1959] à qui l'on doit le terme. Wilson utilisait le principe de charité comme contrainte de traduction pour les textes historiques de manière à rendre vraies et cohérentes le plus grand nombre de phrases. Plus précisément, le principe de charité peut être compris chez Wilson comme un principe de maximisation de la vérité des phrases. Ceci peut amener l'interprète, c'est-à-dire la personne qui tente de comprendre un texte ou le discours d'un locuteur, à remettre les mots à leurs places voire même à les remplacer lorsque telle ou telle phrase est de toute évidence fausse. Ainsi, on pourrait donner l'exemple – que ne donne pas Wilson – d'un locuteur assertant qu'il a plongé sa voiture dans le pot de confiture et qu'il a garé sa cuillère dans le garage, ne pourrait être intelligible si on le prenait aux mots. Mais il suffirait de remplacer voiture par cuillère dans la première partie de la phrase et cuillère par voiture dans la seconde pour que le sens nous paraisse clair et la vérité de cette phrase maximisée, autrement dit d'utiliser le principe de charité.
La difficulté pointée par Wilson est que cette maximisation ne va pas de soi. Au sujet des textes historiques, cette maximisation pourrait se heurter à une « réécriture de l'histoire » (Delpha [2001], p.15).

Quine reprend ce principe de charité dans son ouvrage *Le mot et la chose*. Le principe de charité prend place cette fois dans une entreprise de traduction radicale d'un langage étranger que nous ne comprenons pas. Voyons ce à quoi correspond la traduction radicale pour Quine. Le principe de charité est pour lui une maxime de traduction indiquant que « la stupidité de notre interlocuteur, au-delà d'un certain point, est moins

probable qu'une mauvais traduction » (Quine [1960, 1977], p. 101). Autrement dit, Quine suggère de privilégier l'entente et la correspondance entre les énoncés du locuteur et ceux de l'interprète.

- Le principe de charité chez Davidson

Davidson reprend l'usage que fait Quine du principe de charité. Selon lui, « déchirés entre la nécessité de donner un sens aux mots du locuteur et la nécessité de donner un sens à la trame de ses croyances, le mieux que nous puissions faire est de choisir une théorie de la traduction qui maximise l'accord »[163] (Davidson [1968, 1993b], p. 155).
En ce sens, le principe de charité constitue, comme chez Quine, une maxime de rationalité ainsi qu'une règle de méthodologie.
Plus généralement, le principe de charité est un principe de cohérence qui joue le même rôle en théorie de l'interprétation que les conditions de rationalité en théorie de la décision. Comme nous l'avons mentionné, plusieurs axiomes de la théorie de la décision comme l'axiome de transitivité ou de complétude, reflètent la dimension rationnelle de la prise de décision et donc aussi la rationalité des attitudes qui ont conduits l'individu à choisir telle issue ou telle suite d'issues. Le principe de charité est de manière analogue une condition de rationalité pour l'interprétation.

(b) Le principe de charité comme levier pour déterminer simultanément les croyances et les significations

Mais Davidson va plus loin. Il utilise le principe de charité comme levier pour déterminer à la fois les croyances et les significations.
En effet, ce principe est appliqué de manière « systématique » et en cela, Davidson instaure la maximisation de l'accord. Mais comme il le souligne « la minimisation du désaccord, ou la maximisation de l'accord est un idéal confus. Le but de l'interprétation n'est pas l'accord mais la compréhension […] il n'est pas plus facile de spécifier « la forme d'accord correcte » que de dire en quoi consiste une bonne raison de soutenir une

[163] Notons que Davidson utilise ici le terme de traduction alors que dans d'autres articles il se refuse à utiliser ce concept car il l'estime trop proche de celui de signification.

croyance donnée » (Davidson [1993b], p.15). Cette croyance est, comme on l'a vu, l'attitude qui consiste à tenir pour vraie une phrase en un temps et en un lieu. Cette attitude est le point de départ permettant d'avoir accès, comme nous l'évoquions au début de ce chapitre, à la fois aux croyances et aux significations. Elle permet de mettre en évidence l'interdépendance de ces deux groupes d'états mentaux. Lorsque cette croyance est généralisée à l'ensemble des phrases-T pour lesquelles la convention T est appliquée, il s'ensuit que c'est l'ensemble des croyances que l'interprète considère comme correcte. Plus précisément, le principe de charité, en maximisant l'accord, et lorsqu'il est appliqué à l'attitude de tenir pour vraie une phrase, revient à « garder constante la croyance » (Davidson [1973, 1993b], p. 203). Ainsi, le principe de charité a la même fonction que les propositions éthiquement neutres de Ramsey à savoir maintenir fixe l'un des facteurs pour en déterminer un autre.

(c) Le principe de charité et l'indétermination

Puisque Davidson s'inspire de Quine et de la traduction radicale pour décrire sa théorie de l'interprétation (et ce, comme on l'a vu, en reprenant Tarski et la convention T, Davidson est conduit à renverser la chaîne causale allant de la traduction à la vérité et ceci en plaçant au premier plan l'interprétation) nous allons présenter dans les grandes lignes l'analyse de l'indétermination que propose Quine.

Comme ce dernier le souligne lui-même, les manuels de traduction réalisés à partir du comportement observable de l'indigène « peuvent être élaborés selon des principes divergents, tous compatibles avec la totalité des dispositions à parler et cependant incompatibles entre eux » (Quine [1960], p.58). Plus précisément, les manuels peuvent diverger car leurs liens entre eux peuvent être distendus.

Le cœur du problème est, semble-t-il, l'impossibilité à vérifier la traduction du langage de l'indigène dans notre propre langage. Aucune procédure de confirmation ou d'infirmation n'est à la disposition de l'interprète (Quine [1960], p. 117). Le problème est d'autant plus délicat lorsqu'il s'agit de construire ce que Quine appelle une « hypothèse analytique » c'est-à-dire des « tables de concordance » entre les mots et phrases de l'indigène et les mots et phrases de l'interprète (Quine [1960], pp. 111-112). Autrement dit, l'interprète, en vu de comprendre le langage de l'indigène, doit utiliser

des méthodes le menant à établir un certains nombres d'hypothèses permettant de classer les données verbales de l'indigène. L'interprète établit des conjectures sur le langage de l'indigène. L'idée est donc initialement de construire une structure (logique) qui représenterait à grands traits le langage de l'indigène. Seulement, comme le souligne Quine ([1960], p. 118), il n'existe pas « de manière objective au sujet de laquelle on puisse être dans le vrai ou dans l'erreur ». Quine évoque plus précisément au moins sept causes qui « empêchent d'apprécier » cette vérité (*ibid.*). Ces causes « masquent » selon Quine l'indétermination. Elles empêchent de douter, voire de remettre en cause les hypothèses initiales concernant la traduction du langage de l'indigène. Parmi ces causes, la plus importante selon Quine est sans doute « l'impression persistante qu'un vrai bilingue, à coup sûr, est en mesure de faire généralement des corrélations qui soient univoquement correctes entre des phrases appartenant respectivement à chacun de ces deux langues » (Quine [1960], p. 119). En fait, ce sont tous les éléments que le linguiste introduit nécessairement - comme par exemple ses propres canons de rationalité ou encore « les contraintes pratiques » de la traduction - qui l'empêchent de prendre en compte l'indétermination de la traduction.

La raison de l'indétermination est donc double. D'une part, rien n'assure que les liens entre les stimuli et la traduction soient assez fermes pour que la construction de manuels de traduction divergents soit impossible. D'autre part, puisque de tels manuels sont possibles, il est probable que certaines phrases dans l'un des manuels n'aient pas leurs équivalents dans un autre et ceci laisse finalement la traduction indéterminée.

CONCLUSION

Le chapitre 3 est relatif, comme le chapitre précédent, à la théorie unifiée. Cette fois, il est question de présenter formellement la théorie unifiée défendue par Davidson. Cette présentation met en évidence l'emprunt de Davidson à Richard Jeffrey. Ce dernier avait proposé dans les années 1960 une théorie de la décision dont l'ontologie était réduite aux propositions uniquement. Le philosophe modifie légèrement la théorie de Jeffrey en proposant d'utiliser des phrases non interprétées plutôt que des propositions. Ceci permet à Davidson de présenter son propre modèle en 1980.

Une fois les utilités et les probabilités déterminées, l'auteur propose une série d'étapes permettant d'avoir accès aux significations du sujet. Ces étapes s'inspirent des travaux de Tarski et Quine.

Chapitre 4.

Apports et limites de la théorie de Davidson (1980)

Après avoir présenté les influences de Davidson, et les modèles qu'il construit pour analyser les décisions dans les années 1980, nous cherchons ici à estimer l'apport de sa théorie. Cet apport se situe pour nous à plusieurs niveaux. Nous cherchons d'abord à déterminer si la seconde théorie construite dans les années 1980 que nous venons de présenter permet de surmonter les défauts de la première, analysée dans la première partie de la thèse, et par là, si l'évolution du travail de Davidson a permis d'enrichir sa théorie de la décision depuis les années 1950 (4.1).

Puis, que les limites de la théorie de 1957 soient dépassées ou non, nous nous demandons, si la théorie enrichie des années 80 n'en possèdent pas elle-même d'autres en termes de cohérence interne, ou au regard des objectifs de départ fixés par Davidson (4.2).

4.1 La théorie de 1980 permet-elle de surmonter les limites de celle de 1957 ?

Dans le premier chapitre de cette deuxième partie, nous avons présenté un ensemble de critiques formulées par Davidson près de vingt ans après ses premières expériences en théorie de la décision. La théorie construite en 1980 a-t-elle permis d'y répondre ?

Nous allons revenir successivement sur ces critiques en commençant par le problème des conflits des désirs (4.1.1). Puis nous tenterons de voir si la théorie unifiée est en mesure de répondre à la critique de Davidson relative au caractère statique de la théorie de la décision (4.1.2). Nous nous demanderons enfin s'il n'y a qu'en répondant à ces critiques que la théorie de la décision de 1980 constitue un enrichissement de celle de 1957. La comparaison entre les deux théories doit être selon nous être resituée en fonction de l'évolution des objectifs de Davidson (4.1.3).

4.1.1. Les conflits de désirs dans la théorie de 1980

Dans le premier chapitre de la partie II, nous avions expliqué pourquoi, selon Davidson, la théorie de la décision telle qu'il l'avait testée dans les années 1950 à Stanford n'était pas apte à régler des problèmes de conflits de désirs. Nous avions établi que les conflits de désirs renvoient à ce que Davidson désigne par l'incontinence. L'agent incontinent considère qu'une action *b* est meilleure qu'une action *a* mais pourtant il accomplit *a* avec une raison qui n'est pas la cause de son action. Cette déconnexion entre raison et cause constitue selon Davidson une certaine forme d'irrationalité.

Il s'agit à présent de voir si la théorie de 1980 permet de surmonter cette critique qu'adresse Davidson à sa propre théorie de 1957.

Nous avions vu que selon Davidson, l'agent incontinent n'arrive pas à se comprendre lui-même et qu'en cela il était « sourd » à lui-même (Davidson [1970, 1993a], p. 65).

Le modèle de 1980 peut schématiquement être décrit pas deux avancées majeures par rapport au modèle de 1957.

La première avancée est sans doute l'introduction des phrases non interprétées (sur le modèle de l'introduction des propositions par Jeffrey) comme données de base de la théorie de la décision et l'élément sur lequel portent les préférences (désirabilités) et les probabilités. A ce niveau, une analyse des conflits de désirs semble toujours non opérationnelle puisque l'interprète-expérimentateur se contente uniquement de recueillir les préférences et les probabilités qu'attribuent les individus aux divers états du monde pertinents pour son choix. Or comme on l'a vu, l'incontinence a directement trait aux choix. Autrement dit, l'analyse de la faiblesse de la volonté est dirigée vers l'agent qui a accomplit une action contre son meilleur jugement.

On peut considérer que la deuxième avancée majeure est l'analyse des significations faisant partie intégrante de la théorie d'ensemble qui proposée par Davidson.

Mais le modèle dans lequel s'articulent les désirs, les croyances et les significations est toujours un modèle d'utilité espérée où les préférences doivent être rationnelles pour être représentées par une fonction d'utilité et pour postuler que l'agent maximise son utilité espérée. Or dans le cas du conflit de désirs, l'agent n'entretient pas des préférences rationnelles. L'agent incontinent manifeste des préférences qui sont

déconnectées de son choix à un moment donné mais sans que cette déconnexion ne soit constante puisqu'il se peut très bien que les préférences et les choix soient à nouveau connectés le lendemain ou même la minute d'après.
Ainsi, les modèles d'utilité espérée n'ont pas la possibilité d'intégrer les conflits de désirs dans l'axiomatique, c'est pourquoi le modèle de 1980 qui reste dans ce cadre ne peut y faire face.

4.1.2. La théorie de 1980 est-elle statique ?

Une autre critique avait été formulée à l'égard de la théorie de 1957, son caractère statique : elle traite des préférences et des croyances qui ne changent pas, du moins au cours de la session où elles sont estimées (Davidson, Suppes, Siegel [1957], p. 80). Pourtant en 1957, lorsque les auteurs de *Decision Making* analysèrent la portée et la possibilité d'extension de leur modèle, ils suggéraient en particulier de tenter de la rendre dynamique en intégrant des éléments de théorie de l'apprentissage notamment.
La question qui se pose est donc de savoir si ces éléments ont effectivement été effectivement intégrés dans le modèle de 1980.
Comme on l'a vu, ce dernier reste un modèle d'utilité espérée même si ce sont essentiellement les valeurs de vérité et les attitudes face à des propositions (Jeffrey) ou face à des phrases non interprétées (Davidson) qui permettent de construire à la fois une échelle cardinale des utilités et une échelle subjective des probabilités. Tel que nous l'avons analysé, rien n'indique que des éléments d'une théorie de l'apprentissage aient été intégrés dans ce nouveau modèle. Il s'agit toujours de dresser une coupe instantanée des préférences et des estimations de probabilités en un temps donné et dans un contexte précis. En ce sens, on pourrait dire que la théorie de la décision de 1980 demeure statique.

En revanche, la théorie de l'interprétation qu'introduit Davidson pour construire sa théorie unifiée en 1980 nous semble pouvoir quant à elle être qualifiée de dynamique. En effet, le processus d'interprétation se construit comme on l'a vu selon un processus d'essais et d'erreurs. Il se déroule en plusieurs étapes sans que les attributions de croyances et de significations ne soient définitives. Le principe de charité, comme on

vient de le voir, enjoint de maximiser l'accord entre l'interprète et le locuteur. Ainsi, si, dans un premier temps, il se trouve que les interprétations et les phrases-T qui leur sont associées ne soient pas vérifiées empiriquement, c'est-à-dire qu'elles ne donnent pas les valeurs de vérité des phrases du locuteur en un temps et en un lieu précis, l'interprète va être amené à corriger ses attributions et donc tous les théorèmes (représentés par les phrases-T) qu'il avait de prime abord construits. Autrement dit, un interprète fait évoluer les interprétations qu'il assigne aux phrases d'un locuteur.

La théorie unifiée, dans son ensemble, gagne donc un élément dynamique visant à corriger d'éventuelles erreurs d'interprétations. Cependant rien n'indique que la théorie dans son ensemble en devienne dynamique d'autant qu'il se peut qu'il y ait une évolution des interprétations que l'interprète assigne au locuteur sans que cela signifie que les états de ce locuteur soient eux-mêmes dynamiques.

4.1.3. Un enrichissement spécifique de la théorie de la décision, de son statut épistémologique, au sein de l'œuvre de Davidson

Deux théories sont construites, l'une en 1957, l'autre en 1980. La deuxième enrichit sans nul doute la première en introduisant une analyse des significations.

Mais cette introduction ne modifie-t-elle pas aussi le statut épistémologique de la théorie ?

Il nous semble que si et ce, pour au moins deux raisons.

Davidson, après s'être rendu compte que Ramsey avait trouvé la solution de la fixation des probabilités subjectives, s'est intéressé au versant expérimental de la théorie de la décision comme en témoigne le modèle de 1957. Il propose avec Suppes et Siegel une approche expérimentale liée à une axiomatisation complexe toutes deux très détaillées. Or, le modèle de 1980 ne présente guère de procédure expérimentale ou d'expériences, pourtant si chères à Davidson dans les années 1950. Ce que ce dernier cherche à mettre en lumière en 1980, c'est la possibilité de déterminer simultanément désirs, croyances, significations. L'expérimentation passe au second plan en termes d'objectifs. Et la théorie de la décision ne peut se concevoir qu'insérée dans une théorie plus vaste, la théorie unifiée. L'absence d'expérimentations dans le modèle de 1980 ne constitue donc pas un oubli de Davidson mais bien un choix théorique et méthodologique. Pour mettre

en lumière ce choix, nous allons le resituer au regard des objectifs mêmes de l'auteur. L'introduction des significations modifie le statut épistémologique de la théorie pour une seconde raison. Nous montrerons que la théorie unifiée n'est plus présentée dans un cadre strictement individuel, comme la théorie de la décision de 1957, du fait notamment de l'élargissement des compétences de l'expérimentateur (4.1.3.2).

4.1.3.1. D'une approche expérimentale (1957) à une absence d'expérimentations (1980)

Pour expliquer la modification de la place des expérimentations entre 1957 et 1980, nous nous appuyons sur les trois travaux successifs dans lesquels Davidson introduit la théorie unifiée : en 1980 comme une théorie unifiée de la pensée, de la signification et de l'action (a), en 1985 comme une nouvelle base pour la théorie de la décision (b) et en 1990 au sein d'une discussion sur les théories de la vérité (c). Ces trois dates correspondent à trois formulations identiques de l'axiomatique de la théorie unifiée telle que présentée dans le chapitre précédent. Toutefois, le contexte dans lequel cette théorie est présentée est à chaque fois différent. La présentation de ces trois contextes nous permettra de comprendre pourquoi la théorie unifiée ne dispose pas d'un volet expérimental.

(a) La théorie unifiée en 1980

La théorie unifiée est présentée par Davidson en 1980 comme une « stratégie pour relier le discours à son cadre humain » (Davidson [1980, 2004], p.151). Autrement dit, l'idée est ici de relier l'analyse des significations aux modèles permettant de décrire l'action et les décisions. Le caractère intensionnel de la théorie de la décision telle que nous l'avons présenté dans le chapitre I de cette partie, permet précisément d'unir les contenus de la théorie de la décision et de la théorie de l'interprétation du langage. L'idée de l'auteur est de traiter les significations, les désirs et les croyances comme « des éléments totalement coordonnés dans une compréhension de l'action » (Davidson [1980, 2004], p.153).

Le point central discuté par l'auteur en 1980 est la proximité des enjeux et des méthodes entre théorie de la décision et théorie de l'interprétation : « Ce que l'on doit ajouter à la théorie de la décision, ou incorporer à l'intérieur d'elle, est une théorie de l'interprétation pour un agent, une manière de dire ce que ses mots signifient » (Davidson [1980, 2004], p.154). Cette insistance sur l'interdépendance des deux théories est liée au rôle central des significations dans les deux cas : « La théorie de la signification et la théorie bayésienne de la décision sont faites l'une pour l'autre. La théorie de la décision doit être libérée de l'hypothèse d'une connaissance de la signification déterminée indépendamment ; la théorie de la signification en appelle à une théorie du degré de croyance afin de faire un usage sérieux des relations du support de la preuve » ((Davidson [1980, 2004], p.158). Cette interdépendance permet d'envisager une union des deux théories. Mieux, elle incite à concevoir leurs enjeux respectifs au sein d'une seule et même théorie : « Mais établir ces dépendances mutuelles ne suffit pas, pour que chaque théorie puisse être développée comme une base pour l'autre. Il n'y a pas de voie simple pour ajouter l'une à l'autre puisque pour débuter chacune d'elles requiert un élément que l'autre possède » (*ibid.*). L'avantage de l'union de ces deux théories est, selon Davidson, non seulement l'élargissement de la base empirique de la théorie – autrement dit la théorie unifiée nous permet de traiter un plus grand nombre de données mentales instanciées par le comportement observable et verbal – mais aussi la possibilité de présenter simultanément deux mesures du mental permettant de décrire un contenu plus riche et précis.

L'objectif n'est donc plus comme en 1957, de montrer que la théorie de l'utilité espérée peut être testée empiriquement mais de mettre en lumière les possibilités qu'offre une théorie unifiée.

Si en 1980, Davidson insiste sur la nécessité d'unir la théorie de la décision et la théorie de l'interprétation, il n'hésitera pas à parler de « base » pour la théorie de la décision en 1985.

(b) Une nouvelle base pour la théorie de la décision : Davidson [1985]

En 1985, dans la revue *Theory and Decision*, Davidson revient sur la théorie unifiée mais sous un autre angle cette fois. Il présente son projet comme une approche « fondationnelle » (Davidson [1985], p. 87). L'idée n'est plus de supposer que les événements, les états de la nature, et les objets des préférences sont donnés à l'instar des théories de Ramsey, vNM et Savage mais de considérer uniquement les événements, les états de la nature et les objets qui sont pertinents pour l'individu. Ainsi, Davidson utilise t-il le même type d'argument que celui de Bolker lorsque celui-ci évoque la déformation des probabilités dans les modèles standards.

Les objectifs de Davidson sont donc présentés différemment. Il propose de montrer quelles sont les conditions nécessaires pour disposer d'une théorie unifiées, des conditions formelles et empiriques. La méthodologie suivie par l'auteur se décompose en deux étapes :

- Premièrement, Davidson cherche à déterminer la forme d'une théorie permettant à la fois de déterminer les utilités et les probabilités subjectives mais aussi d'identifier les propositions sur lesquelles portent ces utilités et probabilités. L'objectif est de montrer à quoi pourrait ressembler une théorie satisfaisante des désirs, croyances et significations. Ce type de question relative à la forme de la théorie est essentiel pour comprendre le projet de Davidson. Cette question avait déjà été posée par Davidson lorsqu'il s'intéressait à la forme que devrait prendre une théorie de la signification. Comme Michael Dummett (cité par Pascal Engel, [1994], pp.3-4) l'explique concernant la théorie de la signification, « ce n'est pas que l'on considère que la construction d'une théorie de la signification pour un langage […] soit envisagé comme un projet réalisable mais on pense qu'une fois énoncer les principes généraux en accord aves lesquels une telle construction peut être accomplie, nous serons arrivés à une solution des problèmes concernant la signification ». C'est cette même idée qui anime Davidson en 1985 relativement à la théorie unifiée.

- La deuxième partie consiste à fixer les utilités et les probabilités – ce qui revient à maintenir le statut de théorie statique de la théorie de la décision – pour appliquer la théorie de l'interprétation du langage permettant de découvrir de ce que les phrases du sujet signifient. Cette étape est empirique – à raison d'être expérimentale – selon Davidson car comme on l'a vu, toutes les phrases-T doivent être soumises à la vérification empirique par l'interprète.

Intégrer une théorie de l'interprétation du langage à la théorie de la décision dans une version remaniée de celle de Jeffrey permet de mettre en évidence une nouvelle base pour la théorie de la décision dans son ensemble, une théorie fondée non seulement sur des normes de rationalité et de cohérence mais aussi sur une interprétation radicale des contenus mentaux.

Ces deux étapes ont des implications quant au statut de la théorie. Selon Davidson, il est d'usage de faire une différence entre théories descriptives et normatives de la décision. Pourtant pour lui, cette distinction n'a pas de sens en ce qui concerne la théorie unifiée : « Comme image de la manière dont un agent parfaitement rationnel devrait agir, [les théories bayésiennes] on s'accorde sur le fait qu'elles sont essentiellement correctes même si parfois elles sont trop simplifiées […] Comme théories descriptives pourtant, elles sont considérées au mieux comme limitées en application et idéalisées de manière absurde […] Je doute qu'il y ait une voie intéressante pour comprendre cette distinction » (*ibid.,* p. 89). Le problème soulevé ici par Davidson est l'absence de référence à une interprétation empirique de la théorie. En cela, l'auteur renoue avec les considérations théoriques et épistémologiques discutées dans le modèle de 1957. Autrement dit, la distinction descriptif/normatif n'éclaire pas le contenu empirique de la théorie. La théorie unifiée a bien, selon lui, des applications empiriques, toutefois Davidson la présente comme une « théorie de la rationalité » (*ibid*. p.88) à propos de laquelle il se dit « profondément sceptique sur la possibilité d'en faire un test significatif » (*ibid*.). C'est pourquoi, continue Davidson, « je cesserai de parler des expérimentateurs et de leurs sujets et parleraient plutôt d'agents, acteurs, êtres humains et leurs interprètes » (*ibid,* p. 89). Davidson suggère, en effet, que la théorie unifiée « n'aspire pas à être expérimentée » mais qu'elle constitue « une description idéalisée

des modèles de comportement qui nous permet de pénétrer les pensées et significations des autres » (Davidson [1985], p. 89).

(c) Théorie de la vérité et théorie unifiée (1990) : un plaidoyer en faveur de la méthode de Jeffrey.

Après avoir été présentée comme l'union de la théorie de la décision et de l'interprétation en 1980, ou sous l'aspect d'une théorie de la rationalité en 1985, la théorie unifiée se trouve présentée dans ce troisième article en privilégiant une troisième dimension.

En 1990, dans l'article « The Structure and Content of Truth », Davidson revient en effet sur les aspects méthodologiques de la théorie unifiée. Si l'essentiel de l'article est relatif à la place de la théorie de la vérité de Tarski par rapport aux autres théories de la vérité notamment celles des pragmatistes, l'auteur insiste sur les enjeux pour lui de l'utilisation d'une théorie de la vérité dans le style de celle de Tarski.

La théorie de la signification proposée par Davidson doit prendre la forme d'une théorie de la vérité inspirée de celle proposée par Tarski afin d'éclairer en retour le concept de signification. Comme on l'a vu, la signification n'est pas première dans la théorie mais elle constitue toutefois l'objectif ultime.

L'une des idées de l'article de 1990 est d'insister sur « l'environnement psychologique immédiat des aptitudes linguistiques » (Davidson [1990], p. 315). Il s'agit donc pour Davidson de présenter une fois de plus l'intérêt d'unir à la théorie de la signification qu'il propose une analyse fine des désirs et des croyances.

L'intérêt est cette fois lié à la valeur de vérité d'une phrase que l'on peut déterminer par l'adoption d'une démarche proche de celles de Ramsey et de Quine vis-à-vis, respectivement de la théorie de la décision et à la théorie de la signification. Davidson l'explique ainsi : « une personne tient une phrase pour vraie selon les résultats de deux types de considérations : ce que cette phrase signifie et ce qu'il croit » (1990, p. 41). Comment alors démêler le rôle des croyances et celui des significations ? Exactement comme Ramsey démêle celui des croyances et des désirs en fixant l'un (les probabilités) pour obtenir l'autre (les utilités). De même, Quine parvient à fixer les croyances – grâce

au principe de charité – afin de déterminer les significations à l'aide d'un manuel de traduction.

La théorie unifiée telle qu'elle est présentée en 1990 vise à proposer une démarche qui puisse déterminer simultanément trois entités (désirs, croyances, significations) pour donner un classement des préférences sur des phrases non interprétées. L'objectif revendiqué par l'auteur est n'est toutefois pas d' « éclairer directement la manière dont on en vient à comprendre les autres dans la vie quotidienne (…). Il s'agit d'un exercice conceptuel destiné à révéler la dépendance vis-à-vis d'attitudes propositionnelles à un niveau assez fondamental pour éviter de devoir saisir ces attitudes une à une » (Davidson, 1990, p. 325). Davidson insiste cette fois sur l'intérêt secondaire des implications empiriques au regard de la méthode trouvée.

Cette méthode est bien sûr celle de Jeffrey [1965] : « Nous devons à Richard Jeffrey une version de la théorie de la décision qui ne fait pas un usage direct des paris, mais traite comme objets de préférence, des objets auxquels les probabilités subjectives sont attribuées, et les objets auxquels les valeurs relatives sont attribués sont considérés uniformément comme des propositions. Jeffrey a montré en détail comme extraire des probabilités subjectives et des valeurs à partir de préférences que les propositions soient vraies » (Davidson [1990], pp.323-324).

4.1.3.2. L'interaction sujet-expérimentateur

Si la théorie de la décision de 1957 est une théorie de l'utilité espérée qui est présentée dans un cadre strictement individuel, tel n'est plus le cas de la théorie unifiée. L'interprétation engage au moins deux acteurs : un locuteur et un expérimentateur-interprète. La théorie de la décision n'est plus individuelle, elle n'existe que par l'interaction des individus.

En effet, l'idée centrale défendue pas Davidson est que le processus d'interprétation met en scène trois variétés de connaissance[164] : ce que je sais de moi, « ce que je sais du monde », « ce que je sais des autres ». Ce qui diffère entre ces trois variétés, c'est le mode d'accès à la réalité (Davidson [1991, 2001], p.205).

[164] Ce que Davidson appelle la triangulation.

Davidson traite des différentes variétés de connaissance lorsqu'il tente de répondre non seulement aux doutes sceptiques qui ont parcouru toute la philosophie mais aussi lorsqu'il tente de trouver un fondement à la connaissance en général. C'est pourquoi lorsque celui-ci évoque la connaissance de soi, il englobe toutes les attitudes propositionnelles qui peuvent être présentes dans l'esprit, et lorsqu'il évoque la connaissance des autres, il se demande comment nous pouvons être sûrs de l'existence des autres esprits mais surtout sur quelle base empirique nous pouvons certifier de l'existence de tels esprits. Or une grande partie de nos attitudes sont à la fois privées et sociales. Elles sont le fruit de notre histoire et de notre apprentissage du langage. Un trait constitutif de cet apprentissage consiste dans la possibilité de communiquer et, par ce biais, d'appliquer aux autres le même type d'attitudes que celles que nous nous attribuons à nous-mêmes.

Même si la connaissance que j'ai de moi est caractérisée par ce que Davidson appelle « l'autorité à la première personne » – chaque personne sait, mieux que quiconque, ce qu'elle a dans la tête sans faire appel à la moindre preuve pour cela –, il n'en reste pas moins que nous avons besoin de faire appel à des éléments extérieurs pour expliquer nos propres pensées. Notons que ceci ne constitue en rien une atteinte à l'autorité à la première personne, au contraire. Cette autorité n'est pas amoindrie du fait qu'elle dépende et soit expliquée par les facteurs publics et sociaux. Si l'accès aux comparaisons intrapersonnelles est donné par cette autorité, leur fondement est donc à la fois public et interne.

En fait, il existe un lien causal fort entre les utilisateurs du langage, les événements qui surviennent et les objets du monde et tout ceci détermine la manière dont nous apprenons et communiquons avec les autres. Mieux, cela rend un esprit accessible à un autre en principe (Davidson [1988, 2001], p.52). Davidson refuse de faire appel aux conceptions qu'il qualifie de « standards » de la subjectivité (Davidson [1988, 2001], p 50). Ces conceptions considèrent uniquement les états mentaux qui occupent l'esprit sans faire référence au monde extérieur. Pour l'auteur, il n'existe pas de tels états. Il n'y a aucun mot ou concept qui ne soit pas compris ou interprété directement ou indirectement en termes de relations causales entre les gens et le monde (Davidson [1988, 2001], p. 51). De manière plus générale, il n'est d'aucune utilité de vouloir séparer ce qui provient de moi et ce qui provient du monde.

Pour l'auteur, l'existence de la pensée et de la communication – et donc de toutes les attitudes évaluatives – est le fait que deux créatures ou plus réagissent au monde extérieur et se répondent mutuellement (Davidson [1997, 2001], p. 83). Mieux encore, la source même de l'objectivité réside précisément dans cette intersubjectivité (*ibid.*).

Selon l'auteur, il n'existe pas de langage privé car à moins que celui-ci soit partagé, il n'y a aucune manière de distinguer entre le fait d'utiliser le langage correctement et l'utiliser de manière incorrecte (Davidson [1991, 2001], p.209). Le langage est l'élément qui à la fois nous met en contact avec autrui, et qui nous permet d'attribuer et de se voir attribuer des attitudes évaluatives. Dès lors, nous sommes à la fois un locuteur et un interprète. Nous devons comprendre pour être compris. En tant qu'interprète, comme le souligne Davidson : « je ne peux pas faire mieux au départ que de supposer que ce qui pousse quelqu'un à avoir telle croyance, provoque la même croyance chez moi si la même cause [de celle-ci] survient » (Davidson [1986, 2004], p. 69). Cette méthode repose sur un principe essentiel : nous n'avons pas la possibilité d'assigner une attitude à quelqu'un si l'on pense que son rôle dans les pensées d'un autre est différent du rôle qu'elle joue dans les nôtres (*ibid.*). L'idée fondamentale qui rend la triangulation possible et qui assure la connexion entre « moi », « autrui » et « le monde » est le fait que les systèmes de pensée (le mien et celui d'autrui) doivent concorder.

Comme le souligne Davidson : « Pour comprendre le discours de quelqu'un, je dois être capable de dire les même choses que cette personne ; je dois partager son monde » (Davidson [1982, 2001], p. 105).

Or, rendre les attitudes propositionnelles intelligibles pour nous implique nécessairement de les rendre appropriées à notre propre schème, c'est-à-dire à notre propre représentation, et ceci jusqu'à un certain degré, sans jamais que l'objectivité ne soit menacée, puisqu'elle trouve sa source dans l'intersubjectivité[165] (Davidson [1986, 2004], p.69). Une condition *sine qua non* à cette opération de triangulation consiste à supposer que les autres sont largement cohérents dans leurs croyances. On retrouve ici le principe de charité.

[165] C'est cette même idée qui sera d'ailleurs défendue par Davidson dans sa conception des comparaisons interpersonnelles d'utilité.

Ainsi, l'expérimentateur-interprète, en devenant un acteur de la compréhension des actions et décisions du sujet, fait basculer la théorie de la décision dans une perspective immédiatement duale puisque l'acteur est replacé dans un contexte et dans une interaction constante avec ses interprètes.

4.2. Cohérence et limite internes de la théorie de 1980 : le problème de l'indétermination

Si la théorie unifiée de 1980 constitue un enrichissement de celle de 1957, elle n'est toutefois sur le plan de la cohérence interne pas sans limites. Le défaut souvent relevé est celui de l'indétermination à laquelle elle conduit. Car si la théorie de l'utilité espérée des années 1950 fournit un critère de décisions définis à une classe de fonctions affines près, tel n'est pas le cas de la théorie unifiée. Il est en effet très difficile d'identifier un ensemble cohérent de critères de décision auxquelles elle conduirait.

Cette indétermination constitue selon les philosophes Collins et Rawling, un problème que Davidson n'aurait pas su résoudre. En explicitant les critiques respectives liées à l'indétermination formulées par Rawling (2001) (4.2.1) et Collins (1999) (4.2.2), nous cherchons à déterminer si celle-ci constitue une limite de la théorie unifiée de Davidson ou si elle peut être décrite d'une autre manière (4.2.3).

4.2.1. L'indétermination dans la théorie unifiée : Rawling [2001]

Piers Rawling, ancien étudiant de Davidson, proposa en 2001 une évaluation de la théorie unifiée de Davidson.

L'article écrit par Rawling à ce propos prend pour point de départ l'analogie développée par Davidson « entre la manière dont les phrases fonctionnent dans l'attribution d'attitudes propositionnelles et la manière dont les nombres fonctionnent dans la théorie de la mesure » (Rawling [2001], p.238).

Chez Davidson l'analogie est largement déclinée et justifie le fait que l'indétermination en théorie de l'interprétation ne soit pas plus grave qu'une indétermination de mesure de longueur.

Rawling adopte un point de vue tout à fait différent. Il commence à rappeler que Davidson construit, de manière analogue à sa théorie de 1957, un modèle « normatif », « idéalisé » (*ibid.*) de l'interprétation qui repose sur des conditions que Rawling qualifie de « structurales » et qui sont en fait les conditions formelles et empiriques dont nous avons parlé dans la dernière section et en particulier le principe de charité.

Une myriade d'interprétations est toutefois possible à partir de ces conditions comme une classe de fonctions d'utilité remplit les conditions de rationalité. Pour Rawling, le principe de charité montre ses limites, même « face à toutes les preuves pertinentes » puisqu'il laisse plusieurs interprétations ouvertes : « les données dont dépendent tous ces sujets ne nous offrent aucune manière de séparer les contributions de la pensée, de l'action, du désir et de la signification respectivement. Ce que nous devons construire ce sont des théories globales, et bon nombre de théories feront tout aussi bien (Davidson [1979, 1993b], p.346) ».

Si dans la théorie de 1957 et dans la théorie de 1980, certaines conditions nécessaires à la rationalisation du comportement n'empêchent pas qu'il existe une classe de solutions possibles, Davidson demeure ambigu sur la différence de nature entre la classe définie en 1957 et celle de 1980. Il semble en effet que Davidson n'accorde pas le même statut à la classe des interprétations possibles et à la classe des utilités. En effet, dans le premier cas il parle d'indétermination et non dans l'autre, et il cherche perpétuellement à réduire le champ de l'indétermination de la traduction en utilisant l'analogie avec la mesure physique.

Dans son article de 1989, « What is Present to the Mind », Davidson propose par exemple de comparer la mesure du poids et la mesure du mental (Davidson [1989, 2001], p.59-60). Concernant le poids, les comparaisons sont possibles grâce à des nombres. Les nombres mettent en relation des poids et donnent un sens à des différences de valeurs par exemple mais cela ne signifie pas que les nombres sont une propriété intrinsèque des poids. L'intérêt d'utiliser des nombres est qu'ils permettent de préserver les rapports entre les objets de poids différents. De la même manière, selon Davidson, ce dont nous avons besoin pour la mesure du mental est une « collection d'entités » (*ibid.*) nous permettant de représenter les propriétés pertinentes ainsi que les relations entre les états mentaux. Mais de la même manière que les nombres n'ont aucun rôle en physique c'est-à-dire que les attribuer ou non aux objets ne change rien par

exemple à leurs poids, il peut suffire, pour ce qui est des états mentaux, de les représenter dans une relation causale qui, par exemple, donne les raisons d'une action sans pour autant dire que ces états ont une existence singulière dans l'esprit de l'agent. Lorsqu'il s'agit de décrire des états mentaux sous la forme d'attitudes propositionnelles c'est-à-dire sous la forme d'attitudes vis-à-vis de propositions (désirer que *p* est vraie, croire que *p* etc.), la situation reste inchangée : même si nos attributions d'attitudes ont un sens puisqu'elles décrivent un état mental cohérent avec une énonciation verbale, il n'en reste pas moins qu'elles peuvent ne pas être présentes dans l'esprit ou ne pas correspondre à un point du cerveau sans que cela bouleverse le sens de ces attributions.

Rawling souligne ainsi que « Davidson tente de rendre cette indétermination bénigne en invoquant l'analogie avec la manière dont les phrases fonctionnent dans l'attribution des attitudes propositionnelles, et la manière dont les nombres fonctionnent dans la théorie de la mesure : le fait qu'il y ait différentes manières d'interpréter nos semblables n'est pas plus alarmant que le fait que la longueur peut être représentée par différents schèmes de mesure comme les pieds et les mètres. Si nous activons cette analogie pourtant, il en ressort que l'indétermination est plus alarmante que Davidson ne le concède » (Rawling, 2001, p. 244).

Les efforts de Davidson pour réduire l'indétermination dans la théorie unifiée de 1980 sont pourtant insuffisants selon Rawling ; l'analogie invoquée par Davidson ne peut selon lui être un argument valable. Elle pose en effet problème selon Rawling.

S'il existe, selon lui, un algorithme qui permet de passer d'une échelle de nombres à une autre par exemple dans la mesure des longueurs, il n'existe rien de tel dans le cas de l'interprétation. Or selon Rawling, cet algorithme aurait permis de prolonger significativement l'analogie. Dans la mesure en physique, comme l'explique Rawling, on représente par des nombres à partir d'une structure sous-jacente qui est invariante entre les échelles. Dans le cas où l'on utilise des phrases pour attribuer des attitudes, « nous avons besoin d'identifier la structure sous-jacente qui est invariante entre les différents schèmes d'interprétation » (Rawling, 2001, p. 246). Cette structure invariante doit consister en des « liens et des relations entre eux » (*ibid.*). Dans le cas de la mesure de la longueur, « nous avons, selon Rawling, les objets physiques qui se tiennent dans la

relation ‘est aussi long que’. La relation ne change pas quelque soit les représentations. » (*ibid.*).

Autrement dit, pour Rawling, on ne peut pas trouver dans la théorie de l’interprétation une structure invariante. Le système de coordonnées entre toutes les interprétations n’est pas aussi strict que celui d’un système de transformations de fonctions d’utilité par exemple. Mieux, en cherchant un dénominateur commun entre les interprétations, c’est la singularité des attitudes propositionnelles dans leur ensemble qui semble être remise en cause car cette recherche d’un élément commun revient à miner l’existence même des attitudes propositionnelles. Il faudrait de toute évidence, trouver une preuve invariante des significations et des attitudes propositionnelles dans les phrases mêmes, et cela irait contre l’antiréductionnisme défendu par Davidson. Les significations en seraient réduites, dès lors que l’on cherche un point commun entre elles, à une sorte d’élément invariant et c’est précisément ce que Davidson voulait éviter.

L’explication repose sur le fait qu’il existe un homologue au monisme anomal de la théorie de l’action en théorie de l’interprétation. Dans le cadre de notre présentation de la théorie de l’action de Davidson, nous avions vu dans le premier chapitre de la partie II, que les états mentaux étaient des états physiques dans la mesure où ils se logeaient dans le cerveau mais le non réductionnisme de Davidson revenait à dire que les états mentaux ne se réduisaient pas aux états physiques dans la mesure où l’on ne peut pas fournir, selon l’auteur, de lois en psychologie qui aient la même valeur scientifique que les lois de la physique et que par ailleurs, le mental survient sur le physique.

Il en est de même pour la théorie de l’interprétation : « les discours susceptibles d’être interprétés ne sont rien d’autre que […] des actions associées à des intentions non linguistiques assorties […] ; et ces actions ne sont en retour rien d’autre que (identiques à) des mouvements intentionnels des lèvres et du larynx » (Davidson [1973, 1993b], p. 189). Mais les énonciations ne se réduisent pourtant pas à des états purement physiques. Elles ne « prennent sens » (Engel [1994], p. 238) que lorsqu’elles sont décrites intensionnellement c'est-à-dire dans le registre de la compréhension, comme lorsque l’on évoque des attitudes propositionnelles. Dès lors, en cherchant un élément commun en théorie de l’interprétation, un algorithme permettant de lier toutes les interprétations, on défait nécessairement le monisme anomal, pour la théorie de l’action tout autant que pour la théorie de l’interprétation. Et pourtant, cet algorithme aurait pu faire office de

preuve permettant d'unir les interprétations tout comme les fonctions d'utilité sont reliées par un système de transformations.
Rawling en conclut qu'il existe une indétermination très importante au sein de la théorie unifiée qui en limite considérablement la portée.

4.2.2. L'indétermination des désirabilités et des probabilités: l'analyse de Collins [1999]

Si Rawling souligne la trop grande nonchalance de Davidson par rapport au problème de l'indétermination, le philosophe John Collins a proposé en 1999 une analyse intéressante de cette question en posant le problème à partir des équations de Bolker-Jeffrey.
Son idée consiste à fouiller la différence entre l'indétermination relative à la classe des utilités dans le modèle canonique de l'utilité espérée et celle de la théorie unifiée. Autrement dit, il s'agit pour lui de montrer ce que devient la détermination de l'utilité à une fonction affine près lorsque celle-ci prend place dans une théorie plus large où il s'agit de déterminer simultanément les désirs, les croyances et les significations.
Avant de procéder à un calcul numérique qui servira d'exemple, Collins rappelle l'ambigüité maintenue dans l'œuvre de Davidson à l'égard des résultats de sa théorie unifiée. Il identifie plus précisément deux types d'attitudes adoptées par Davidson :

- Dans une perspective optimiste telle que présentée par Davidson dans son article « The Structure and the content of truth » en 1990 (issu des Dewey Lectures (1989)), la théorie unifiée permet de réunir dans une seule théorie les méthodes de Ramsey et de Quine. Mieux, la théorie unifiée est la preuve qu'il est possible de représenter, au sein d'une même théorie, l'interdépendance des désirs, des croyances et des significations. Davidson considère cette théorie comme un « exercice conceptuel » (Davidson [1990], p.325) permettant d'accéder à ces trois données mentales en même temps.
- Mais l'auteur avait, quelques années auparavant, pointé une difficulté quasi structurelle de cette théorie unifiée et c'est la perspective pessimiste qu'identifie Collins. En effet, expliquait lui-même, « Je voudrais une théorie de la

> signification comme la version de Ramsey de la théorie de la décision, mais la meilleure que je connaisse d'y parvenir manque cet objectif (Davidson [1980, 2004], p.155). Ce qui est en cause ici c'est l'indétermination sous-jacente à la théorie de l'interprétation.

Ce que propose Collins, c'est d'évaluer la place de cette indétermination lorsque les contenus de la théorie de la décision et de la théorie de l'interprétation sont unifiés en reprenant les équations de Bolker Jeffrey.

Comme nous l'avions montré dans le chapitre 3 de cette seconde partie, le but de Jeffrey était de montrer qu'il existe au moins un couple (p(·), d(·)) qui vérifie la condition d'existence – et qui, par conséquent, représente les préférences de l'agent.

En fait, pour que l'équivalence ait lieu entre les couples (p(·), d(·)) et (P(·), D(·)), les paramètres de la transformation doivent vérifier 3 conditions :

1. ad – bc strictement positif (équivalent au $a > 0$ dans le cas de l'utilité espérée de la théorie de Ramsey Savage)
2. c d(T) + d = 1, ce qui établit un lien entre les paramètres si on se donne $d(T)$ et limite les degrés de liberté à 3.
3. c d(A) + d > 0 pour tout A.

Ces conditions sont nécessaires pour démonter que si (p(·), d(·)) vérifie les conditions d'existence, alors (P(·), D(·)) en fait autant.

Comme on l'a vu, le modèle proposé par Bolker-Jeffrey appréhende la préférence comme un désir pondéré de manière à prendre en compte le monde tel qu'il est, sans modifier arbitrairement les probabilités qu'attribuent les individus aux états du monde. Le problème du pari proposé dans les théories de vNM ou de Ramsey est que celui-ci constitue une loterie purement arbitraire et non pas un questionnement relatif à la préférence d'un sujet par rapport à des événements.

Il s'ensuit que les utilités et les probabilités déterminées dans le modèle de Bolker-Jeffrey le sont de manière plus restrictive que pour les modèles de vNM et Savage par exemple.

Ce que propose Collins c'est justement une application numérique mettant en évidence cette faille dans la détermination des utilités et des probabilités.

Il propose d'une part, de fixer l'échelle de désirabilités en posant a = 1 et b = 0.

Concernant l'échelle des probabilités, Collins propose, d'autre part, de partir de la désirabilité de la tautologie T auquel, comme Jeffrey et Davidson le suggèrent, on peut assigner la valeur 0. Ainsi, d(T) = 0. Il en résulte, à partir de la deuxième condition évoquée plus haut que d = 1 (c .0 + d = 1).

En introduisant ces valeurs dans les équations des couples p(·), d(·) et P(·), D(·) proposées par Jeffrey, Collins en déduit pour une proposition A quelconque :

$$D(A) = \frac{d(A)}{c.d(A)+ 1} \text{ et } P(A) = (c.d(A) + d)\, p(A)$$

Comme p(A) est positif, il s'ensuit que c.d(A) > -1. Etant donné qu'il a été supposé que l'échelle de désirabilité est fixe (en posant a = 1 et b = 0), il s'ensuit que *c* va osciller entre une valeur minimale et une valeur maximale : $-\frac{1}{\max} < c < -\frac{1}{\min}$.

Imaginons, avec Collins, que le couple p(·), d(·) est donné par les matrices suivantes :

	B	~B
A	5/1-8	1/3
~A	1/3	1/18

Matrice de probabilités p(·)

	B	~B
A	9/10	9/4
~A	-9/4	-9/2

Matrice de désirabilités d(·)

Il doit être possible, en utilisant les équations de Bolker-Jeffrey, de déterminer un autre couple P(·), D(·) qui représenterait les préférences et les probabilités de ce même individu.

Comme la valeur la plus faible de la matrice de désirabilités est -9/4 et la valeur maximale est 9/4, Collins considère que *c* est encadré comme suit : $-\frac{4}{9} < c < \frac{2}{9}$. Le cas

où $c = 0$ est le cas particulier où d(·) = D (.) et p(.) = P(.). Dès lors, Collins propose de prendre le point $c = -\frac{2}{9}$ ce qui correspond au point médian sur l'intervalle.

En utilisant les équations de Bolker-Jeffrey, on trouve deux autres matrices représentant le couple P(·), D(·) :

	B	~B
A	2/9	1/6
~A	½	1/9

Matrice de probabilité P(.)

	B	~B
A	9/8	9/2
~A	-3/2	-9/4

Matrice de désirabilité D(.)

Or, la transformation ainsi réalisée ne préserve pas la probabilité de « ~A ∧ B ». Elle est de 1/3 dans la matrice initiale et est de 1/2 après transformation. On peut s'apercevoir aussi que la transformation a renversé l'ordre de certaines comparaisons de probabilités.
Comme le souligne Jeffrey lui-même (Jeffrey [1965, 1983], pp. 95-96), le cas limite où les probabilités ne peuvent pas être transformées et les désirabilités ne sont définies qu'à une transformation linéaire positive près est le cas où l'échelle de désirabilités n'est pas bornée, ce qui est possible si l'on fixe $c = 0$.
Autrement dit, contrairement à ce qui est à l'œuvre dans la théorie de l'utilité espérée où une classe de fonctions d'utilité représentent toutes les mêmes préférences, il n'apparaît que très rarement possible dans la théorie unifiée de trouver des fonctions qui représentent les mêmes préférences. Le champ des possibles ne prend plus la forme d'une classe de fonctions définies à une fonction affines près, il n'est plus réductible à un ensemble. C'est pourquoi cette indétermination pose problème.

4.2.3 L'indétermination mine-t-elle l'ensemble de la théorie unifiée ?

D'une certaine manière, l'indétermination, que Rawling et Collins mettent en avant, pourrait amoindrir le processus d'interprétation. De manière plus générale, si cette indétermination était poussée à son paroxysme, elle reviendrait à empêcher l'élaboration d'une mesure intelligible du mental du locuteur tant par l'outil de la théorie de la décision que par l'outil de la théorie de l'interprétation. Seulement Davidson n'adhère pas à l'idée d'un échec total de l'interprétation. Les arguments qu'il mobilise contre cette possibilité sont relatifs à ce qu'il appelle le « relativisme conceptuel ».

Le relativisme revient à dire que ce qui a de la valeur est relatif à un contexte et sous le prédicat « conceptuel » il exprime l'idée que les personnes, communautés ou cultures conceptualisent ou organisent le monde de manière différente (Davidson [1986, 2004], p.40). Le langage est perçu comme « un pouvoir de mise en ordre, qui ne se distingue pas clairement de la science » (Davidson [1974b,1993b], p. 277) et qui fait face à l'expérience : « ce qui est mis en ordre, désigné tantôt comme "expérience" tantôt comme "flux de l'expérience sensorielle" et comme "données physiques" » (*ibid*.).

Or, l'une des caractéristiques de ce relativisme conceptuel est que les différents schèmes conceptuels peuvent diverger largement de telle sorte à ce qu'il n'existe pas de « système de coordonnés commun » (Davidson [1974b, 1993b], p. 268). Autrement dit, c'est l'idée de l'échec de l'intertraduisibilité comme « condition nécessaire à la différence entre schèmes conceptuels » (*ibid*.).

L'autre caractéristique est qu'il postule l'existence de quelque chose de « neutre » – un point de vue sur une montagne – et de « commun » en dehors de tous les schèmes (ici les schèmes correspondent aux différents langages c'est-à-dire les différentes manières d'organiser les pensées et les significations) (Davidson [1974b,1993b], p. 277). Autrement dit, nous pouvons trouver un point à partir duquel nous pouvons envisager les différentes représentations ; l'agent qui observe pouvant être délié de son propre schème.

Davidson se positionne en opposition au relativisme conceptuel ainsi décrit. Voici comment se décompose son raisonnement.

Tout d'abord il présente un trait représentatif du relativisme conceptuel : « Bergson nous dit où nous pouvons aller pour parvenir à un point de vue sur une montagne qui ne soit pas faussé par telle ou telle perspective provinciale » (Davidson [1974b, 1993b], p. 268). Or, pour Davidson « Il n'y a aucune chance pour que quelqu'un puisse s'élever à un certain point de vue qui lui permette de comparer les langages en se débarrassant temporairement du sien » ((Davidson [1974b, 1993b], p. 270).

L'analogie est claire. De même que « la signification, en un sens relâché de ce mot, est contaminée par la théorie », l'attribution d'attitudes comme des désirs et des croyances et l'opération de comparaison (entre les schèmes) est nécessairement à l'œuvre ((Davidson [1974b, 1993b], p. 273). Nous n'avons pas la possibilité d'assigner une attitude à quelqu'un si l'on pense que son rôle dans les pensées de l'autre est différent du rôle qu'elle joue dans les nôtres (Davidson [1986, 2004], p. 69). Rendre les attitudes propositionnelles intelligibles pour nous implique nécessairement de les rendre appropriées à notre schème jusqu'à un certain degré et cela ne les rend pas moins objectives ((Davidson [1986, 2004], p. 69).

En réponse au relativisme conceptuel qui supposait l'échec total de la traduction, Davidson propose une approche plus « modeste » où c'est l'échec partiel de la traduction qui est retenu. Le passage suivant de Davidson synthétise cette idée : « Ni un stock fini de significations, ni une réalité neutre par rapport à toute théorie ne peuvent fournir une base de comparaison entre schèmes conceptuels. Ce serait une erreur de continuer à chercher une telle base si par là nous entendons quelque chose qui serait conçu comme étant commun à des schèmes incommensurables » (Davidson [1974b, 1993b], p. 284).

Autrement dit, à travers la critique faite au relativisme conceptuel, Davidson propose une réponse à l'idée d'indétermination.

Pour faire l'analogie avec la théorie de 1957, Davidson propose ici ce que l'on pourrait appeler une borne supérieure à la théorie unifiée là où l'indétermination constituerait une borne inférieure. Cette borne supérieure est, selon nous, le versant optimiste et positif de la théorie unifiée. Elle assure en quelque sorte qu'il sera toujours possible de mesurer le mental grâce aux outils combinés de la théorie de la décision et de la théorie

de l'interprétation du langage. La borne inférieure rappelle, toutefois, que si la mesure est toujours possible, elle est, par nature, indéterminée en en cela la théorie unifiée peut sembler confinée dans son statut de théorie unifiée du comportement.

CONCLUSION

La théorie de Davidson construite en 1980 est considérée comme unifiée car elle a pour objet un triangle composé des désirs, des croyances et des significations et parce qu'elle repose sur leur détermination simultanée. Puisant à nouveau dans la théorie économique, Davidson choisit de combiner les théories de Ramsey et Jeffrey avec des arguments issus de la logique et la théorie du langage. Il crée ainsi une théorie de la décision enrichie qui fonctionne en tandem avec une théorie de l'interprétation fondée sur les propositions de Tarsky et Quine. Il ne s'agit plus de déterminer les désirs et croyances à partir de données comportementales mais de déterminer les désirs, les croyances et les significations à partir de préférences sur des phrases non interprétées.

La nouvelle théorie répond ainsi à certaines critiques adressées à celle de 1957. Elle ne peut toutefois répondre à toutes et elle suppose elle-même une certaine indétermination dans la réponse proposée, indétermination qui n'empêche toutefois pas les significations de jouer leurs rôles.

CONCLUSION GENERALE

Donald Davidson peut être considéré comme l'un des pionniers de l'expérimentation économique sur des problèmes de décision. Ses recherches à Stanford dans les années 1950, regroupées dans l'ouvrage *Decision Making* attestent à la fois d'une volonté de tester expérimentalement la théorie de l'utilité espérée de Ramsey, vNM et Savage mais aussi de mesurer l'utilité par intervalles également espacés.

Toutefois, un certain nombre de défaillances empiriques sont identifiées par Davidson lui-même dès les années 1950. La théorie de la décision standard dans sa version expérimentale est statique par nature et ne permet pas d'intégrer des conflits de désirs ou des inversions de préférences. Mais le principal problème soulevé par Davidson est qu'elle n'intègre pas une théorie de l'interprétation du langage. Plus précisément, l'expérimentateur considère comme acquises les significations qu'accordent les sujets aux options parmi lesquelles il doit choisir. Or, selon Davidson, ceci ne va pas de soi. Rien n'indique que les croyances et les significations du sujet sont identiques à celles de l'expérimentateur.

Pour tenter de répondre à cette objection, Davidson propose un nouveau modèle dans les années 1980, modèle inspiré des travaux de Jeffrey [1965]. Le modèle de Jeffrey fait usage des propositions comme support de la théorie. Ainsi s'agit-il de déterminer les désirabilités et les probabilités subjectives que les individus attribuent à des propositions, c'est-à-dire à des phrases dotées de sens. Le modèle de Davidson aménage la théorie de Jeffrey de telle sorte que les données fondamentales ne soient pas des propositions mais des phrases non interprétées. Ce changement est motivé par une tentative d'intégrer une théorie de l'interprétation du langage au cœur de la théorie. Les raisons de cette intégration découlent d'une analyse croisée de la théorie de la décision et de la théorie de l'interprétation du langage : l'auteur souligne leur interdépendance théorique et méthodologique ainsi que les enjeux d'un tel rapprochement. Ces enjeux sont présentés au sein d'une théorie unifiée des désirs, des croyances et des significations.

Cette théorie unifiée est énoncée sous la forme de la théorie de l'utilité espérée afin de répondre, dans une certaine mesure aux objections formulées au modèle de 1957 par Davidson lui-même dans les années 1970 notamment. Pourtant, cette théorie unifiée aurait pu être présentée sous la forme d'un modèle d'utilité non espérée dans

la mesure où Davidson se rapproche à plusieurs reprises d'auteurs comme Edwards [1953, 1954] ou Tversky [1975] pour critiquer non seulement son modèle de 1957 mais aussi la théorie de la décision dans son ensemble. Nous avions vu, par exemple, que Davidson s'appuie sur les « effets de présentation » mentionnés par Edwards et sur le « problème de l'interprétation » qu'évoque Tversky.

On considère aujourd'hui qu'Edwards et Tversky sont les représentants d'un courant alternatif à la théorie de l'utilité espérée : la théorie de l'utilité non espérée (Machina [2008]). Cette dernière prend appui sur les remises en cause empirique de la théorie de l'utilité espérée – comme le paradoxe d'Allais par exemple – pour produire une nouvelle théorie capable en principe de surmonter les défaillances du modèle standard. La théorie du prospect de Kahneman et Tversky [1979] – dont on peut retracer l'origine dans les travaux d'Edwards – propose une axiomatisation de l'utilité qui n'est pas linéaire en probabilités (hypothèse centrale de la théorie de l'utilité espérée). Comme le mentionne Kahneman dans un entretien avec Christian Schmidt, la véritable rupture consiste à ne pas prendre comme point de départ une logique de la décision pour construire une théorie des choix risqués. Ainsi, là où la théorie de l'utilité espérée – comme la version de Jeffrey par exemple – s'attache à mettre en évidence la logique à l'œuvre dans la prise de décision, la théorie du prospect de Kahneman et Tversky [1979] s'attache à dégager le processus mental sous-jacent à celle-ci. Cette nouvelle théorie tente de décrire par exemple l'opération mentale permettant à l'agent de « sélectionner les informations pertinentes en vue de préparer le choix »[166]. Une autre étape, liée à la première, consiste à décrire la « mise en perspective de ces informations au moyen d'un « cadrage » qui leur donne sens pour le décideur ».

Le modèle de 1980 de Davidson, alors même qu'il pourrait se nourrir des travaux fondateurs d'Edwards et de la théorie du prospect de Kahneman et Tversky, reste un modèle d'utilité espérée. La question que l'on est en droit de se poser est celle de savoir pourquoi Davidson n'a pas opté pour un modèle d'utilité non espérée.

[166] Voir entretien de Kahneman avec Christian Schmidt disponible sur le site : http://www.ffsa.fr/webffsa/risques.nsf/b724c3eb326a8defc12572290050915b/e56e36ac8c5388c3c125723d004c73f0/$FILE/Risques_067_0005.htm

Pour y répondre, nous proposons plusieurs hypothèses qui expliqueraient cette non-conversion de Davidson.

Selon une première hypothèse, on pourrait relever qu'en 1980, Davidson n'appartient plus aux cercles des économistes et des psychologues qui étaient les siens dans les années 1950. En effet, comme nous l'avions mentionné, l'environnement intellectuel de Davidson à Stanford était propice à des recherches en théorie de la décision : financement de l'armée pour des recherches dans ce domaine, nouvel idéal d'explication scientifique de la décision avec l'ouvrage de vNM (1947) et influence de McKinsey et Suppes. En 1980, Davidson n'est plus immergé dans un tel environnement, même si son intérêt pour les travaux de Jeffrey témoigne d'une curiosité constante.

Selon une deuxième hypothèse, on pourrait imaginer que la théorie de l'utilité non espérée n'avait pas encore connu, dans les années 1980 un essor suffisamment important pour que Davidson en soit informé. Mais même s'il le fut, il n'est pas sûr que Davidson ait accepté de développer un modèle d'utilité non-espérée tant la déception du modèle de 1957 fut grande ; l'intention d'en surmonter les défaillances l'aurait sans doute emporté sur une quelconque attirance théorique pour un modèle comme celui que propose la théorie du prospect.

Enfin, selon une troisième hypothèse, strictement théorique et qui nous semble la plus vraisemblable, on peut imaginer que la théorie du mental de Davidson constitue un obstacle à sa conversion à l'utilité non espérée. En effet, non seulement cette dernière considère que les problèmes relatifs au langage ne sont pas déterminants comme en attestent les travaux d'Edwards que nous avions évoqué dans le chapitre 1 de la partie II mais en plus, la théorie du mental en jeu dans la théorie du prospect par exemple, est incompatible avec celle de Davidson.

L'originalité de Davidson réside non seulement dans l'introduction des significations au cœur de la théorie de la décision mais aussi dans la proposition d'un modèle qui rend compatibles la théorie de l'utilité espérée et la théorie de l'interprétation du langage.

L'idée centrale qui sous-tend cette initiative est que l'on ne peut analyser et comprendre la décision sans faire appel au langage puisqu'il constitue à la fois un

outil de communication mais aussi un révélateur de contenus mentaux. En supposant cette connexion, Davidson prône un rapprochement théorique et méthodologique de disciplines jusque là séparées, en insistant sur les enjeux d'un tel rapprochement. L'enjeu majeur est une théorie de la mesure du mental ancrée dans l'interaction entre un sujet-locuteur et un expérimentateur-interprète.

ANNEXE : Entretien avec Marcia Cavell (extrait, mai 2008)

PVH : Tout d'abord merci d'avoir accepter de me rencontrer. C'est un honneur pour moi de discuter de Davidson avec vous. Ma thèse porte sur les écrits économiques de Davidson, essentiellement ceux en collaboration avec Patrick Suppes ainsi que sur la théorie unifiée des années 1980. Il y a deux axes dans mon travail. Le premier est relatif à son apport à la théorie de la décision et plus précisément à l'économie expérimentale. Le second est relatif à la théorie unifiée, ce qui le pousse à faire intervenir des significations au sein de la théorie de la décision et quel type de théorie de la décision il tente de construire dans les années 1980.

MC : Vous pensez que c'est pertinent ?

PVH : Oui, il y a un grand nombre de preuves textuelles qui attestent de l'importance de la théorie de la décision pour Davidson. Par exemple, Piers Rawling, l'un des étudiants de Davidson a écrit dans ce sens

MC : Je ne suis pas très au fait de ses travaux dans ce domaine. Mais de manière générale, je suis à la fois surprise et triste qu'il y ait peu de reviews sur le travail de Don. Lorsque l'on lit une review de la philosophie américaine de ces dernières années, Donald est relativement peu cité.

PVH : Quels sont les philosophes qui sont sur le devant de la scène selon vous ? Si on ne parle pas de Davidson, de qui parle-t-on ?

MC : Bernard Williams, Tom Nagel…Ils écrivent des choses à la mode et plus accessibles que les articles de Donald. Bernard Williams a par exemple écrit « Shame and assessivity » Nagel parle par exemple du sens de la vie et c'est ce que les gens lisent. Ce sont bien sûr des philosophes très intéressants mais Donald est dur à lire ! La première fois que je l'ai lu et que je suis entrée dans ses concepts, il était surpris que je lui dise que ça me semblait difficile !

Il est dur mais il est clair.

PVH : Mais Davidson évoque lui aussi des problèmes très concrets, par exemple lorsqu'il parle du principe de Charité…

MC : Les exemples de Davidson sont merveilleux. Un artiste visionnaire nommé Robert Morris, qui a découvert la philosophie de Davidson par lui-même. Il est même devenu ami lui. Ils ont même collaboré ensemble à travers un tableau où il s'est aveuglé pour le faire. Ce tableau est d'ailleurs chez moi, il s'appelle « *Blind Time Drawings with Davidson* ». L'idée était de décrire par une création artistique, un processus similaire aux liens que Donald avait présenté dans sa théorie de l'action entre l'intention et l'acte en lui-même. Morris illustrait chacun de ses dessins par un extrait de texte de Donald.

PVH : Il y a des périodes où Davidson ne semble pas avoir publié. Savez-vous ce que Davidson a fait entre 1959 et 1963 ?

MC : Entre 1959 et 1963, il a mené une psychanalyse. Il ne savait plus sur quoi travailler. Il m'a raconté qu'il se sentait un peu perdu.

PVH : A-t-il continué à lire Ramsey jusqu'à la fin de sa vie ?

MC : Oui, toujours, il n'a jamais abandonné Ramsey.

BIBLIOGRAPHIE

ALLAIS M., [1953], Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine, *Econometrica* vol.21 n°4, pp.503-546.

AMSTRONG W.E. [1950], A note on the theory of consumer's behavior**,** *Oxford Economic Papers* (2), pp. 119-122.

ARROW K., [1951], Alternative Approaches to the Theory of Choice in Risk-Taking Situations, *Econometrica*, 19(4), octobre, pp. 404-437.

BAUMOL W. [1951], The Neumann-Morgenstern Utility Index – An Ordinalist View, *The journal of Political Economy*, 59(1), pp. 61-66.

BERNOUILLI D. [1738, 1985 (tr.fr.)], Esquisse d'une nouvelle mesure du sort, *Cahiers du séminaire d'histoire des mathématiques*, tome 6, pp.61-77.

BOLKER E., [1967], A simultaneous Axiomatization of Utility and Subjective Probability, *Philosophy of Science*, 34(4), décembre, pp.333-340.

BOREL E. [1937], *Traité du calcul des probabilités et de ses applications, Tome I, Les principes de la théorie des probabilités*, Gauthier-Villars, Paris.

BRADLEY R. [2007], A Unified Bayesian Decision Theory, *Theory and Decision*, 63(3), pp.263-263.

BROOME J., [1990], Bolker-Jeffrey expected utility theory and axiomatic utilitarianism, *Review of Economic Studies*, 57, pp.477-502.

CAMERER C. [1995], Individual decision making in Kagel J.H., Roth A. (eds), *The handbook of experimental economics*, Princeton University Press, Princeton, pp.587-703.

COLLINS J., [1999], Indeterminacy and Intention in Hahn L.E.[1999], pp.501-528.

DELPHA I., [2001], *Quine Davidson : le principe de charité*, Presses Universitaires de France, Paris.

DAVIDSON D. [1963,1993a], Actions, raisons et causes in DAVIDON [1980, 1993a].

DAVIDSON D. [1967, 1993b], « Vérité et Signification » in Davidson [1993b], pp.41-68.

DAVIDSON D. [1969,1993a], « L'individuation des événements » in Davidson [1993a], pp.219-244.

DAVIDSON D. [1970a, 1993a], « Comment la faiblesse de la volonté est-elle possible ? » in Davidson [1993a], pp.37-66.

DAVIDSON D. [1970b, 1993a], « Les évènements mentaux » in Davidson [1993a] pp. 277-303.

DAVIDSON D. [1970, 1993b], « Sémantique pour les langues naturelles » in Davidson [1993b], pp.93-106.

DAVIDSON D. [1971,1993a], « L'agir » in Davidson [1993a], pp.67-92.

DAVIDSON D. [1973,1993b], « L'interprétation radicale » in Davidson [1993b], pp.187-207.

DAVIDSON D. [1974, 1993a], « La psychologie comme philosophie » in Davidson [1993a], pp.305-324.

DAVIDSON D. [1974, 1993b] « La croyance et le fondement de la signification » in Davidson [1993b], pp.208-227.

DAVIDSON D. [1974b, 1993b] « Sur l'idée même de schème conceptuel » in Davidson [1993b], pp.267-289.

DAVIDSON D. [1975, 1993b], « Pensée et discours » in Davidson [1993b], pp.228-251.

DAVIDSON D. [1978,1993a], « Avoir une intention » in Davidson [1993a], pp.119-148.

DAVIDSON D. [1979, 1993b], « L'inscrutabilité de la référence » in Davidson [1993b], pp.327-346.

DAVIDSON D. [1980], A Unified Theory of Thought, Meaning and Action, in Davidson [2004], pp.151-166.

DAVIDSON D., [1984], « Expressing Evaluations » in Davidson [2004], pp.19-38.

DAVIDSON D., [1993a], *Actions et Evénements* (tr.fr. Pascal Engel), Presses Universitaires de France, Paris (*Essays on Actions and Events* (1980), Oxford University press, Oxford).

DAVIDSON D., [1993b], *Enquêtes sur la vérité et l'inteprétation* (tr.fr. Pascal Engel), éditions Jacqueline Chambon, Nîmes (*Inquiries into truth and interpretation* (1984), Oxford University press, Oxford)

DAVIDSON D. [1982], « Empirical Content » in Davidson [2001], pp. 159-176.

DAVIDSON D., [1985], « A new basis for decision theory », *Theory and Decision* (18); pp.87-98.

DAVIDSON D.,[1986], « The Interpersonal Comparisons of Values », in Davidson [2004], pp.59-74.

DAVIDSON D. [1988], « The Myth of the Subjective », in Davidson [2001], pp.39-52.

DAVIDSON D. [1989, 2001], « What is Present to the Mind », in Davidson [2001], pp.53-68.

DAVIDSON D., [1990], The Structure and Content of Truth, *The Journal of Philosophy*, 87(6), Juin, pp.279-328.

DAVIDSON D., [1991], « Three Varieties of Knowledge », in Davidson [2001], pp.205-220.

DAVIDSON D. [1991], *Paradoxes de l'irrationalité*, Editions de l'Eclat, Nîmes.

DAVIDSON D., [1995, 2004], « The Problem of Objectivity » in Davidson [2004], pp.3-18.

DAVIDSON D., [1996], The folly of trying to Define Truth, *The Journal of Philosophy*, 93(6), Juin, pp.263-278.

DAVIDSON D., [1997], « The Emergence of Thought » in Davidson [2001], pp.123-134.

DAVIDSON D., [1999], « Autobiography » in Hahn L.E.[1999], pp.3-70.

DAVIDSON D., [2001], *Subjective, Intersubjective, Objective*, Oxford University press, Oxford.

DAVIDSON D., [2004], *Problems of Rationality*, Oxford University press, Oxford.

DAVIDSON D., MARSCHAK J. [1959], « Experimental tests of a Stochastic Decision Theory » in Churchman C.W., Ratoosh P., *Measurement : Definitions and theories*, Wiley & Sons, New-York, pp.233-270.

DAVIDSON D., McKINSEY J.C.C., SUPPES P., [1955], Outlines of a Formal Theory of Value I, *Philosophy of Science*, 22(2), Avril, pp.140-160.

DAVIDSON D., SUPPES P., [1956], *Econometrica*, 24(3), Juillet, pp. 264-275.

DAVIDSON D., SUPPES, P., SIEGEL, S. [1957], *Decision making: An experimental approach*, Stanford, 1ère edition 1957, Stanford University Press, Stanford.

DELEDALLE G., [1998 (3ème ed.)], *La philosophie américaine*, De Boeck Université, Paris.

DOKIC et ENGEL P., [2001], *Ramsey Vérité et Succès*, Presses Universitaires de France, Paris.

DOSTALLER G., [2005], *Keynes et ses combats*, Albin Michel, Paris.

EDWARDS W., [1953], Probability-Preferences in Gambling, *The American Journal of Psychology*, 66(3), Juillet, pp.349-364.

EDWARDS W.[1954a], Probability-Preferences among Bets with Differing Expected Values, *The American Journal of Psychology*, 67(1), Mars, pp.56-67.

EDWARDS W., [1954b], The Theory of Decision Making, *Psychological Bulletin*, 51(4), pp.380-417.

ENGEL, P. [1989], *La Norme du Vrai*, Gallimard, Paris.

ENGEL, P. [1991], Préface in Davidson [1991].

ENGEL, P. [1994], *Davidson et la philosophie du langage*, L'interrogation Philosophique, PUF.

ENGEL [1994b] (sous la direction de), *Lire Davidson*, Editions de l'Eclat, Combas.

ENGEL P. [1994b], « Perspectives sur Davidson » in ENGEL [1994b] (sous la direction de), *Lire Davidson*, Editions de l'Eclat, Combas.

ENGEL P., [1997] (sous la direction de), *Davidson Analysé*, Cahiers de Philosophie de l'Université de Caen (29), Presses Universitaires de Caen, Caen.

ENGEL P., [1998], Le rôle de la croyance dans l'explication de l'action in J.L. Petit, ed. *Les neurosciences et la philosophie de l'action*, Vrin, Paris, pp.327-339.

FRIEDMAN M., SAVAGE L., [1948], The Utility Analysis of Choices Involving Risk, *The journal of Political Economy*, vol. 56, n°4, pp. 279-304.

FRIEDMAN M., SAVAGE L., [1952], The Expected-Utility Hypothesis and the Measurability of Utility, *The journal of Political Economy*, vol.60, n°6, pp.463-474.

GALAVOTTI M.C., [2005], *Philosophical Introduction to Probability*, CSLI Publications, Stanford.

GAMBINO E., MARCHIONATTI R., [1997], Pareto and Political Economy as a Science: Methodological Revolution and Analytical Advances in Economic Theory in the 1890s, *Journal of Political Economy*, vol. 105(6), pp. 1322-48, December.

GILLIES D., [2000], *Philosophical Theories of Probability*, Routledge, Londres.

GRANGER G.G., [1960], *Pensée formelle et sciences de l'homme*, Aubier, Paris.

GRANGER G.G., [1988], *Essai d'une philosophie du style*, Odile Jacob, Paris.

GUALA F. [2008], « History of Experimental Economics » in *The New Palgrave Dictionary of Economics*, Steven Durlauf and Lawrence Blume (eds.) Palgrave-Macmillan, Londres.

HACKING I., [2002], *L'émergence de la probabilité*, Seuil, Paris.

HAHN L.E.(ed.) [1999], *The Philosophy of Donald Davidson*, The library of Living Philosophers, Chicago and La salle, Illinois.

HALMOS P.R., [1937], The Legend of John Von Neumann, *The American Mathematical Monthly*, 80(4), Avril, pp.382-394.

HICKS J.R., ALLEN R.G.D., [1934], A reconsideration of theory of value Part I., *Economica*, New Series, vol.1, n°1, pp.52-76.

HICKS J.R., ALLEN R.G.D., [1934], A reconsideration of theory of value Part II. A mathematical , *Economica*, New Series, vol.1, n°1, pp.52-76.

JALLAIS S., PRADIER P.C. [1997], L'erreur de Daniel Bernoulli ou Pascal incompris, *Economie et Sociétés*, Œconomia, Histoire de la pensée économique, série P. E., n° 25, pp. 17-48.

JEFFREY R., [1983], *The Logic of Decision*, University of Chicago Press, 1ère edition 1965.

KAHNEMAN D., TVERSKY A., [1979], *Prospect theory: An analysis of decisions under risk*, Econometrica, 47, 313-327

KEYNES J.M., [1921], *A Treatise on Probability*, MacMillan, Londres.

KANT E., [1781 (ed.originale), 2001 (tr.fr)], Critique de la Raison Pure, Garnier Flammarion, Paris.

KNIGHT F., 1921, *Risk, uncertainty and profit*, Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Co.

LAPIDUS A., [2000], La rationalité du choix passionnel : En quête de l'héritage de David Hume, *L'année sociologie*, 50(1), pp.8-84.

LEONARD R. [1995], From Parlor Games to Social Science: Von Neumann, Morgenstern, and the Creation of Game Theory, 1928-1994, *American Economic Association*, 33(2), Juin, pp.730-761.

LEPORE E., McLAUGHLIN B. (eds) [1985], *Actions and Events, Perspectives on The Philosophy of Donald Davidson*, Basil Blakwell, Oxford.

LEPORE E. (ed.) [1986], *Truth and Interpretation, Perspectives on The Philosophy of Donald Davidson*, Basil Blakwell, Oxford.

LEVI I., [1999] « Representing Preferences: Donald Davidson on Rational Choice » in HAHN L.E.(ed.) [1999], pp. 531-570.

LUCE D. RAIFFA H.[1956], *Games and Decisions. Introduction and Critical Survey*, Dover Publications, New-York.

MACHINA M. [2008], « Non-Expected Utility Theory » in *The New Palgrave Dictionary of Economics*, Steven Durlauf and Lawrence Blume (eds.) Palgrave-Macmillan, Londres.

MALINVAUD E., [1952], Note on von Neumann-Morgenstern's Strong Independence Axiom, *Econometrica* (20), p.679

MAY K. [1954], Intransitivity, Utility, and the Aggregation of Preference Patterns, *Econometrica*, 22, pp.1-13.

MCKINSEY J., 1952, *Introduction to the Theory of Games*, McGraw-Hill, New York

MECKLER L. [1950], The Value-Theory of C.I.Lewis, *The Journal of Philosophy*, 47(20), Septembre, pp.565-579.

MOSCATI I. [2004], Early Experiments in Consumer Demand Theory: 1930-1970, *History of Political Economy*, 39(3), pp. 359-401

MOSTELLER F., NOGEE P., [1951], *The journal of Political Economy*, vol.59, n°5, pp. 371-404.

PARETO V., [1900, 1982], *Ecrits d'économique pure*, Oeuvres complètes, tome XXVI, Librairie Droz, Genève.

PARETO V., [1909], *Manuel d'économie politique*, Giard et Brière, Paris.

PICAVET E. [1996], *Choix Rationnel et Vie Publique*, Presses Universitaires de France, Paris.

PHILIPS L.D. & WINTERFREAD D. [2006], « Reflections on the Contributions of Ward Edwards to Decision Analysis and Behavioral Research » in *Advances in decision analysis: from foundations to applications*, Cambridge University Press, Cambridge, pp. 71-80.

PRESTON M.G., et BARRATA P., 1948 , "An Experimental Study of the Auction Value of an Uncertain Outcome," *American Journal of Psychology*, vol. 61, pp.183-193.

QUINE [1960, 1977(tr.fr)], *Le mot et la chose*, Flammarion, Paris.

RAMSEY F.P., [1931], *The Foundations of Mathematics*, articles réunis par Richard B. Braithwaite, Londres, 1931, repris dans *Philosophical Papers*, édité par D.H.Mellor, Cambridge University Press, 1990.

RAWLING P., [2001], « Davidson's Measurement Theoretic Reduction of the Mind » in *Interpreting Davidson*, Kotatko P., Pagin P., Segal G. (eds), CSLI Publications, Stanford, pp.237-356.

REUCHLIN M. [1970], La mesure en psychologie in *Traité de Psychologie expérimentale I. Histoire et Méthode*, Presses Universitaires de France, Paris.

RIVENC F. [1998], *Sémantique et Vérité. De Tarski à Davidson*, Presses Universitaires de France, Paris.

ROTH A. [1995], « Introduction to Experimental Economics » in *Handbook of Experimental Economics*, edited by J.H. Kagel and A. E. Roth, Princeton University Press, Princeton, pp. 3-109.

SAHLIN N-E., [1990], *The Philosophy of F.P. Ramsey*, Cambridge university press, Cambridge.

SAMUELSON P. [1952], Probability, Utility and the Independence Axiom, *Econometrica*, 20(4), pp.670-678.

SAVAGE L., [1954], *The foundations of Statistics*, 2ème edition 1972, New-York, Dover Publications.

SCHEFFER H.M.[1913], A Set of Five Independent Postulates for Boolean Algebras, with Application to Logical Constants, *Transactions of the American Mathematical Society*, 14(4), Octobre, pp. 481-488.

SCHMIDT C. [2001], *Théorie des Jeux : Essai d'interprétation*, Presses Universitaires de France, Paris.

STIGLER G.J., [1950], The Development of Utility Theory I, *The journal of Political Economy*, vol.58, n°4, pp. 307-327.

STIGLER G.J., [1950], The Development of Utility Theory II, *The journal of Political Economy*, vol.58, n°5, pp. 373-396.

TARSKI A., 1944, "The semantic conception of truth and the foundations of semantics", *Philosophy and Phenomenological Research*, 4, 1944, pp. 341-376

TARSKI A., [1956], *Logic, Semantics, Metamathematics*, Corcoran, J., ed. Hackett. 1st edition edited and translated by J. H. Woodger, Oxford University Press.

THURSTONE L.L.1931, Attitudes can be measured, *The American Journal of Sociology*, 33(4), Janvier, pp.529-554.

TVERSKY A. [1970], La prise de decision individuelle in Coombs C., Dawes R., Tversky A. (eds.), *Mathematical Psychology*, traduction française *Psychologie Mathématique* (1975), PUF.

Von NEUMANN and MORGENSTERN [1944, 1947(2ème ed.)], *Theory of games and economic behavior*, Princeton University Press, Princeton.

WILSON N., [1959], Substances without Substrata, *Review of Metaphysics*, 12(4), pp.521-539.

TABLE DES MATIERES