

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

NORMES LOGIQUES ET MODÉLISATION DU RAISONNEMENT CONDITIONNEL:
LE CAS DE LA SUPPRESSION DES INFÉRENCES

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN PHILOSOPHIE

PAR

JANIE BRISSON

MARS 2014

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

As far as laws of mathematics refer to reality, they are not certain. As far as they are certain,
they do not refer to reality.

Albert Einstein

REMERCIEMENTS

Ce mémoire est l'achèvement d'un parcours académique et personnel plus riche et satisfaisant que je ne l'aurais imaginé lors de mes débuts en philosophie à l'UQAM. Ce parcours a été parsemé de mes plus belles réussites et aussi de mes plus grands défis. Tout au long de ces années, j'ai eu la chance d'être entourée de personnes remarquables qui ont célébré mes bons moments avec sincérité et qui m'ont généreusement offert leur support dans les moments les plus difficiles. Je remercie d'abord les professeur-e-s Josiane Boulad-Ayoub, Sara Magrin, Pierre Poirier et Mathieu Marion, qui m'ont encouragé et poussé à dépasser mes limites. Je remercie aussi le professeur Henry Markovits, pour m'avoir accueilli dans son laboratoire, où j'ai obtenu des bases solides en psychologie expérimentale ainsi qu'un nouveau regard sur mes objets de recherche. Je tiens à remercier particulièrement mon directeur de recherche, le professeur Serge Robert, pour avoir cru en mes capacités alors qu'elles n'étaient encore qu'embryonnaires, pour sa confiance et son support intellectuel inébranlables et surtout, pour m'avoir fourni un environnement de recherche où j'ai pu expérimenter et développer mes capacités à leur plein potentiel. Je remercie aussi mes collègues et ami-e-s Anne Brel Cloutier, Guillaume Beaulac et Jean-François Chartier pour leurs conseils et leurs critiques. Mon copain, collègue et complice, Maxime Sainte-Marie, mérite mes sincères remerciements pour son soutien moral, intellectuel et académique essentiel. Finalement je remercie ma famille pour le confort que seule une famille peut fournir : mes frères Jean-François et Vincent ainsi que ma mère, Martine Landry, pour avoir été présente dans les moments les plus difficiles. Je dédie ce mémoire à mon père, Gilles Brisson, dont les enseignements m'ont fourni un esprit critique qui a contribué fondamentalement à former la jeune chercheuse que je suis aujourd'hui.

RÉSUMÉ

Dans ce mémoire, nous ferons une analyse et une évaluation critique de différentes théories sur le fonctionnement non monotone du raisonnement humain et sur différentes normes du raisonnement valide. Nous analyserons d'abord la théorie des modèles mentaux (Byrne et Johnson-Laird) au sein de laquelle la logique classique est la norme du raisonnement valide. Les critiques de cette position nous mèneront à l'analyse d'une approche probabiliste du raisonnement (Oaksford et Chater). L'approche bayésienne du raisonnement est originale par la proposition d'une nouvelle norme de la rationalité. Toutefois, nous défendrons qu'une modélisation via les probabilités bayésiennes des procédures du raisonnement non monotone soit incomplète. En revanche, une modélisation par l'utilisation d'une logique non monotone (Stenning et van Lambalgen) capture mieux les procédures de raisonnement non monotone. Afin de modéliser plus fidèlement le raisonnement non monotone, Stenning et van Lambalgen doivent restreindre la portée normative de leur modèle. Plutôt que de proposer une norme universelle du raisonnement valide, leur système de logique n'est normatif que pour les procédures qu'il modélise. Finalement, la restriction observée au niveau de la portée normative des systèmes formels nous incite à remettre en question l'utilisation d'une norme universelle du raisonnement. Ainsi, nous défendrons qu'une norme adaptée aux contextes de raisonnement et aux tâches à accomplir, ce que nous appelons une *normative toolbox*, permettrait de mieux modéliser le raisonnement humain dans son ensemble.

Mots clés : Logique, implication, raisonnement conditionnel, non monotonie, modèles mentaux, probabilités, logique non monotone.

TABLE DES MATIÈRES

RÉSUMÉ.....	iv
LISTE DES TABLEAUX	viii
INTRODUCTION.....	1
0.1 La suppression des inférences : donnée empiriques.....	4
0.2 Les différents modèles.....	4
0.2.1 La théorie des modèles mentaux.....	5
0.2.2 La théorie de la logique mentale.....	6
0.2.3 Une vision probabiliste de la rationalité.....	7
0.2.4 La logique non monotone.....	8
0.2.5 Un assouplissement de la logique non monotone.....	8
0.3 Discussion sur la rationalité.....	9
0.3.1 Vers une rationalité contextualisée.....	9
0.3.2 Perspective adoptée.....	11
CHAPITRE I	
LA LOGIQUE MENTALE ET LES MODÈLES MENTAUX.....	13
1.1 Logique classique des propositions et inférences de l'implication	13
1.2 La logique mentale : une approche syntaxique.....	15
1.2.1 La procédure générale du traitement de l'information	16
1.2.2 Choix de la sémantique	16
1.2.3 La définition de l'implication et le MPP.....	17
1.2.4 Les différents modèles.....	18
1.2.5 Les procédures pour le MTT est les sophismes.....	18
1.3 La suppression des inférences : un effet de contenu.....	19
1.4 Les modèles mentaux : une approche sémantique.....	21
1.4.1 Définitions de l'implication.....	22
1.4.2 Les niveaux de sophistication du raisonnement.....	23
1.4.3 Procédure pour le MPP.....	23
1.4.4 Procédure pour le MTT.....	24
1.4.5 Procédure avec une prémisse biconditionnelle.....	24

1.4.6	Procédure pour les sophismes.....	25
1.4.7	Explication de la suppression avec les modèles mentaux.....	25
1.4.8	Critiques.....	29
CHAPITRE II RATIONALITÉ BAYESIENNE.....		33
2.1	Contre le logicisme.....	33
2.1.1	Les trois niveaux d'explication de Marr.....	33
2.1.2	Où les auteurs se situent.....	35
2.2	Redéfinition de la rationalité en termes probabilistes : Débat au niveau computationnel.....	36
2.2.1	Discussion sur la rationalité.....	36
2.2.2	Le rejet de la logique classique.....	36
2.2.3	La sélection optimale d'information.....	38
2.3	Le <i>simple probabilistic model</i>	40
2.3.1	Probabilités et théorème de Bayes.....	40
2.3.2	Modèle au niveau computationnel.....	41
2.3.3	La suppression des inférences valides : les exceptions à la règle.....	42
2.3.4	La suppression des inférences invalides : les antécédents alternatifs.....	46
2.3.5	La suppression des inférences invalides : les exceptions à la règle.....	47
2.3.6	La suppression des inférences invalides : antécédents alternatifs.....	49
2.3.7	Un modèle quantitatif, une interprétation qualitative.....	50
2.3.8	Entre norme et description.....	51
2.4	Critique : Les différentes stratégies de raisonnement	51
2.5	Probabilités et logiques non monotones : débat au niveau algorithmique.....	53
2.5.1	Les bayésiens contre les logiciens.....	53
2.5.2	Les logiciens contre les bayésiens.....	55
CHAPITRE III LOGIQUE NON MONOTONE.....		56
3.1	Réintroduction de la logique au niveau computationnel.....	56
3.1.1	Modèles multiples à portée restreinte.....	56
3.2	Le modèle.....	58

3.2.1	Le <i>closed-world reasoning</i>	58
3.2.2	Définition de l'implication.....	58
3.3	Explication de la suppression.....	59
3.3.1	Les inférences <i>forward</i> : MPP et DA.....	59
3.3.2	Les inférences <i>backward</i> : MTT et AC.....	62
3.4	Critique : La suppression est un phénomène gradué.....	65
3.5	Une flouification du modèle.....	67
3.5.1	Logique polyvalente (Lukasiewicz, Bochvar, Kleene).....	67
3.5.2	Logique floue.....	69
3.5.3	Règle générale d'inférence floue.....	70
3.5.4	Application du modèle à la suppression.....	71
3.6	Vers une rationalité multiple et contextualisée.....	73
	CONCLUSION.....	75
	BIBLIOGRAPHIE.....	83

LISTE DES TABLEAUX

Tableau 0.1 Table de vérité de l'implication matérielle.....	6
Tableau 1.1 Table de l'opérateur unaire de négation classique.....	14
Tableau 1.2 Table des opérateurs binaires principaux.....	14
Tableau 4.1 Implications floues.....	82

INTRODUCTION

Qu'est-ce que la rationalité? C'est la question à laquelle Aristote a répondu en élaborant la syllogistique. Dans cet ensemble de règles qui déterminent ce qu'est une inférence déductive valide, il a instauré le concept de contrainte logique. La capacité d'accepter une conclusion qui découle logiquement de prémisses vraies peut être perçue comme la marque de la rationalité chez l'être humain. Selon Aristote, tout individu qui refuse de se soumettre à cette contrainte est soit irrationnel, soit confus par rapport aux prémisses. Autrement dit, être rationnel, c'est savoir dire « si les prémisses suivantes sont vraie, alors la conclusion qui en découle doit l'être aussi ». C'est en ce sens que les travaux d'Aristote ont contribué à l'élaboration des sciences formelles du raisonnement et de la norme du raisonnement valide. Ainsi, dès ses premiers traités de logique, que la postérité a appelé l'*Organon*, il est possible de voir un lien direct entre rationalité et logique. Cette approche est la première, à tout le moins la plus systématique, claire et organisée de l'époque, à faire un lien entre le raisonnement humain et les systèmes formels. En essayant de rendre compte de l'utilisation du langage, Aristote a été plus que seulement descriptif et a pris une posture normative : avec la syllogistique, il devenait possible de distinguer entre le bon et le mauvais raisonnement et ce, avec des règles explicites en main, applicables systématiquement et universellement. Le lien entre logique et raisonnement humain est encore observable aujourd'hui, notamment de la part des psychologues du raisonnement qui entreprennent de modéliser le raisonnement humain à l'aide d'outils formels.

La psychologie du raisonnement teste expérimentalement le raisonnement logique des humains dans le cadre des règles syllogistiques ou bien de la logique classique des propositions. La logique classique, élaborée par Frege (1879) et Russell (début du XX^e siècle), est utilisée afin de modéliser le raisonnement comme étant bivalent (toute phrase est soit vraie, soit fausse). Elle comporte cinq opérations principales, qu'on appelle les connecteurs logiques, soit (la négation (non), la conjonction (et), la disjonction (ou), l'implication matérielle (si,...alors) et l'équivalence (si et seulement si)). Dans le cas de la logique classique, les psychologues s'intéressent particulièrement au raisonnement

conditionnel normé par l'implication matérielle, compte tenu des nombreuses difficultés que les humains rencontrent face à cet opérateur. On teste ce raisonnement via les quatre inférences de base de l'implication. Le Modus Ponendo Ponens (MPP)¹ et le Modus Tollendo Tollens (MTT)² constituent les deux inférences valides et sont généralement faits par une très grande proportion des sujets (le MPP est endossé par presque tous les sujets et un faible pourcentage d'entre eux n'endossent pas le MTT). Le sophisme de l'affirmation du conséquent (AC)³ ainsi que le sophisme de la négation de l'antécédent (DA)⁴ sont deux raisonnements invalides et sont généralement considérés comme étant valides par une proportion importante (environ 60%) des sujets, qui font alors une erreur de raisonnement dans le cadre de la logique classique.

Une des expériences les plus discutées est sans conteste la tâche de sélection de Wason (1966; 1968). Dans cette expérience, Peter Wason a testé les capacités des êtres humains à falsifier une hypothèse. Les performances furent considérablement faibles, avec un taux de réussite d'un maigre 10%. Plutôt que de falsifier les hypothèses, les sujets se contentaient en grande majorité de les vérifier. C'est dans ce contexte qu'est née l'expression « biais de confirmation », qu'on retrouve fréquemment dans la littérature sur le raisonnement et qui sert à qualifier la focalisation sur l'information nécessaire à la vérification d'une hypothèse et l'ignorance de l'information nécessaire à la falsification d'une hypothèse. L'analyse qu'on peut faire des erreurs systématiques de raisonnement fait encore objet de chauds débats aujourd'hui. Toutefois, une autre expérience, tout aussi surprenante que celle de Wason, n'est pas aussi systématiquement et ardemment discutée. Ruth Byrne (1989) a découvert que le

¹ Le Modus Ponendo Ponens est fait lorsqu'on a l'implication *Si p alors q*, qu'on sait que *p* est vrai et qu'on conclut que *q* est vrai.

² Le Modus Tollendo Tollens est fait lorsqu'on a l'implication *Si p alors q*, qu'on sait que *q* est faux et qu'on conclut que *p* est faux.

³ Ce sophisme est fait lorsqu'on a l'implication *Si p alors q*, qu'on sait que *q* est vrai et qu'on conclut que *p* est vrai.

⁴ Ce sophisme est fait lorsqu'on a l'implication *Si p alors q*, qu'on sait que *p* est faux et qu'on conclut que *q* est faux.

raisonnement conditionnel chez les adultes peut être modifiable selon le nombre de prémisses et leur mode de présentation; les quatre inférences de base de l'implication peuvent être inhibées par l'ajout de nouvelles informations sous la forme de nouvelles prémisses. Cet « effet de suppression » a été utilisé par Byrne pour attaquer la rivale classique de la théorie des modèles mentaux, soit la théorie de la logique mentale. Depuis, cette découverte est systématiquement utilisée comme base expérimentale par excellence pour trancher entre des débats théoriques.

Plusieurs psychologues ont tenté de rendre compte des résultats de la psychologie expérimentale du raisonnement, que ce soit concernant le raisonnement logique, le raisonnement probabiliste ou au niveau de la prise de décision (*decision making*). Certaines approches peuvent rendre compte des observations dans une tentative d'expliquer les mécanismes fondamentaux de la cognition humaine, le « blue-print » ou bien l'architecture de l'esprit. Parmi les principales théories qui abordent la question, on retrouve les théories à processus duaux, la théorie de la modularité de l'esprit et la psychologie évolutionniste. Ces théories sont toutes complexes et controversées et en rendre compte avec respect et rigueur serait un travail qui ferait l'objet d'un mémoire en entier. C'est pourquoi nous nous contenterons de rappeler leur existence et de mentionner qu'elles ne seront pas objet d'étude ici. Nous ne nous y attarderont pas puisqu'elles ne rendent pas compte du phénomène sur lequel nous nous penchons, soit la suppression des inférences. En revanche, nous nous pencherons sur la théorie des modèles mentaux (Johnson-Laird, Byrne), la théorie de la logique mentale (Braine, Rips) ainsi que les critiques qui leurs sont adressées dans le cadre de discussions sur le phénomène de suppression. Nous nous intéresserons aussi à des travaux (plus récemment apparus dans la littérature) qui critiquent l'utilisation de la logique classique pour appréhender le raisonnement humain, au profit des probabilités bayésiennes (Oaksford et Chater) ou d'un système de logique non monotone (Stenning et van Lambalgen). Nous verrons aussi les critiques du modèle d'Oaksford et Chater faites par Stenning et van Lambalgen. Nous défendrons que ces derniers proposent la théorie qui tient le mieux la route parmi celles passées en revue. Finalement, nous proposerons qu'une modification du modèle de Stenning et van Lambalgen soit encore plus fidèle aux observations faites à propos de la suppression des inférences de l'implication.

0.1. La suppression des inférences : données empiriques

Dans son expérience de 1989, Byrne présente la conditionnelle *Si Marie a un travail à faire, elle étudie tard à la bibliothèque* ainsi que la prémisse *Elle a un travail à faire*. Dans cette situation, 96% des participants concluent que *Elle étudie tard à la bibliothèque* et ainsi, font le MPP. Toutefois, si on présente aussi la prémisse *Si la bibliothèque est ouverte, elle étudie tard à la bibliothèque* en conjonction avec les deux précédentes, les participants font le MPP dans une proportion de 38%. Cet effet est observé pour chacune des quatre inférences de l'implication. Toutefois, les résultats diffèrent dans les cas où la prémisse ajoutée est additionnelle (en conjonction avec les autres prémisses) ou alternative (en disjonction avec les autres prémisses).

L'intérêt de ces données est qu'elles nous permettent d'observer la non-monotonie présente dans le raisonnement humain. La monotonie est une propriété métalogique qui exprime qu'une conclusion tirée à partir de prémisses est définitive. Dès qu'on retrouve ces mêmes prémisses, même si de l'information nouvelle a été ajoutée, la même conclusion sera tirée. Ainsi, un raisonnement non monotone serait de tirer une conclusion différente à partir de prémisses auxquelles on a ajouté de l'information. Il arrive souvent que les sujets ne font plus les inférences valides ainsi que les sophismes de l'implication suite à l'ajout de prémisses additionnelles ou alternatives; leur raisonnement, dans ces cas précis, n'est pas monotone.

0.2. Les différents modèles

Nous nous intéresserons à trois tentatives de modélisation du phénomène. La première est celle proposée par Byrne (1989) et est élaborée dans le cadre de la théorie des modèles mentaux de Byrne et Johnson-Laird (Johnson-Laird, 1989; Johnson-Laird et Byrne, 1991; Byrne 2005). La seconde est celle d'Oaksford et Chater (2007; 2009; 2010), qui définissent la rationalité dans le cadre du calcul bayésien des probabilités. La troisième est l'élaboration d'un système de logique non monotone proposé par Stenning et van Lambalgen (2008;

2010). Chacune des trois approches a ses avantages et ses inconvénients et chacune d'elles définit l'implication différemment. Nous terminerons en proposant que le modèle de Stenning et van Lambalgen est le plus prometteur pour l'intelligence artificielle et qu'une modification de ce système le rendrait encore plus fidèle aux observations empiriques.

0.2.1. La théorie des modèles mentaux

La théorie des modèles mentaux soutient que le raisonnement des êtres humains se fait via des modèles du monde à petite échelle. Face à une tâche logique testant le raisonnement conditionnel, les sujets auraient en tête des modèles qui représentent les trois cas vrais de la table de vérité de l'implication et non le cas faux. Un modèle initial représente explicitement le premier cas vrai de la table de vérité et implicitement les autres cas vrais possibles. Les tenants de la théorie des modèles mentaux soutiennent qu'un de ses avantages principaux est sa forte portée explicative. Notamment, elle présente une explication de plusieurs autres phénomènes observables à propos du raisonnement logique, comme le fait que la disjonction exclusive soit traitée plus facilement que la disjonction inclusive.⁵ Dans ce cadre d'analyse, l'erreur logique est une erreur de performance due au fait que la tâche en question demande un nombre élevée de modèles mentaux, qui encombrant la mémoire de travail.

Depuis la publication de « Mental models » par Philip Johnson-Laird en 1983, la théorie des modèles mentaux est bien ancrée dans la littérature sur le raisonnement logique. L'ouvrage que ce dernier a publié conjointement avec Byrne (1991) a aussi contribué à poser les bases de la version la plus répandue de la théorie des modèles mentaux. Byrne et Johnson-Laird proposent une explication du raisonnement logique comme étant une procédure en trois étapes, basée sur deux principes fondamentaux. La première étape est la construction du modèle. La seconde est la recherche de contre-exemples. La troisième, lorsqu'aucun contre-

⁵ La disjonction exclusive est vraie dans deux cas (lorsqu'une des deux prémisses est vraie) alors que la disjonction inclusive est vraie dans trois cas (lorsque l'une des deux prémisses est vraie ou lorsque les deux sont vraies).

exemple n'est trouvé, est l'adoption du modèle et de sa conclusion. Un premier principe, le principe du vrai, énonce que les humains focalisent sur le vrai plutôt que sur le faux. Un second principe, le principe de parcimonie des possibilités, énonce que les humains font une utilisation optimale de leur mémoire de travail en ne gardant en tête qu'un nombre restreint (une ou deux) de possibilités de cas vrais face à une conditionnelle. (Byrne, 2005) Considérant que cette théorie s'appuie normativement sur la logique classique, l'analyse du raisonnement conditionnel est basée sur la table de vérité de l'implication matérielle en logique classique (voir tableau 1.1).

Tableau 1.1 Table de vérité de l'implication matérielle

P	Q	→
V	V	V
V	F	F
F	V	V
F	F	V

Le premier principe énonce que le cas faux de l'implication est ignoré par les sujets et le second énonce que le foyer est fait, d'abord, sur le premier cas vrai de l'implication, soit la première rangée de sa table de vérité (le cas où l'antécédent et le conséquent sont vrais). C'est dans ce cadre théorique que Byrne (1989) rend compte du phénomène de suppression qu'elle observe.

0.2.2. La théorie de la logique mentale

La théorie des modèles mentaux est défendue en opposition à la théorie de la logique mentale (Braine et O'Brien, 1998; Rips, 1994). Selon cette dernière, les humains puisent dans leur bagage de connaissances des règles d'inférences qu'ils mettent en œuvre pour produire des raisonnements logiquement valides (dans le cadre de la logique classique). Les tenants de la théorie de la logique mentale soutiennent que l'erreur logique est due à des facteurs comme une mauvaise compréhension des directives données par l'expérimentateur, un manque de capacité de la mémoire de travail ou l'ignorance de la règle logique pertinente. Les erreurs

logiques abondamment observées ne témoignent pas d'une incapacité à raisonner logiquement dans l'absolu (donc ne témoignent pas d'un manque de compétences logiques), mais bien des limitations des êtres humains à bien comprendre et résoudre une tâche logique ou utiliser correctement les règles mentales qu'ils possèdent (donc témoigne de problèmes au niveau de la performance à résoudre des tâches). Beaucoup d'erreurs logiques seraient des erreurs de performance et non de compétence⁶.

0.2.3. Une vision probabiliste de la rationalité

Les théories classiques sur le raisonnement logique (modèles mentaux et logique mentale) perçoivent la logique classique comme seule norme du raisonnement valide. Rares sont les psychologues qui se sont intéressés à d'autres systèmes formels pour normer le raisonnement humain. Toutefois, Oaksford et Chater (2007, 2009, 2010) ont exploré une autre avenue : l'idée que les humains raisonnent conformément au calcul bayésien des probabilités. Dans ce cas, la théorie des probabilités, dans une interprétation subjectiviste⁷, est utilisée afin de décrire les performances des participants réalisant des tâches logiques. Ce qui était perçu comme une erreur logique devient alors une performance tout à fait adéquate dans le cadre d'une norme probabiliste de la rationalité. En optant pour un modèle où il y a des probabilités, les auteurs prétendent rendre compte de l'incertitude dans le raisonnement ainsi que des degrés de croyance que l'on accorde à des propositions : « [...] the problems that we have identified concern how uncertainty is handled in human inference, and probability is the

⁶ La distinction performance/compétence a été formulée pour la première fois par Noam Chomsky pour ensuite être reprise par les psychologues. La définition exacte de cette distinction varie selon les auteurs. Pour le moment, nous utilisons une définition minimale des deux notions que nous peaufinerons au cours des prochains chapitres.

⁷ Il existe deux interprétations de la théorie des probabilités; l'interprétation fréquentiste et l'interprétation subjectiviste. Les fréquentistes utilisent les probabilités comme étant la fréquence avec laquelle un événement tend à se produire. Par exemple, si on lance une pièce de monnaie, il y a 50% de chances qu'elle tombe sur pile car si on la lançait un très grand nombre de fois, elle tomberait sur pile lors de la moitié des lancers. Les subjectivistes, comme Oaksford et Chater, utilisent les probabilités comme degré de croyance d'une personne envers une théorie ou une proposition incertaine. Par exemple un médecin qui pose un diagnostic le fait avec un certain degré de croyance, considérant les symptômes de son patient.

calculus of uncertainty. » (Oaksford & Chater, 2003, p. 100) Étant donné que la logique classique est bivalente, il leur aurait été impossible de l'utiliser pour rendre compte de l'incertitude présente dans le raisonnement. Selon ces derniers, la totalité de la rationalité humaine peut être cernée par un modèle probabiliste bayésien. Nous nous intéressons à cette approche puisque les auteurs prétendent que leur modèle fournit une explication au phénomène de suppression des inférences, ce qui est le problème central de ce mémoire.

0.2.4. La logique non monotone

Stenning et van Lambalgen (2008, 2010), sans prétendre modéliser l'ensemble du raisonnement humain, proposent un système de logique non monotone qui leur permet de cerner le phénomène de suppression des inférences. Leur système génère une procédure par laquelle un raisonnement non monotone peut être fait. Comme l'information qu'on obtient sur le monde peut changer dans le temps, l'apport d'une nouvelle information change la dimension du monde dans lequel le raisonnement est fait. Ainsi, l'ensemble des informations disponibles constituerait un monde clos pour le raisonnement des sujets (« closed-world reasoning »). À partir de ces paramètres de bases, ils sont en mesure de construire un système logique qui génère des raisonnements non monotones. Sur le plan technique, leur avantage sur le modèle d'Oaksford et Chater est que leur système ne représente le phénomène que sporadiquement, mais l'englobe dans toute sa complexité; il modélise formellement le résultat ainsi qu'une procédure qui permet d'atteindre ce résultat.

0.2.5. Un assouplissement de la logique non monotone

Politzer et Bonnefon (2005) ainsi que Stevenson et Over (2001), ont fait des observations importantes à propos du phénomène qui nous intéresse. La suppression des inférences ne serait pas franche, mais graduée. Autrement dit, face à de nouvelles prémisses, les sujets seraient tantôt peu enclins, tantôt très peu enclin à tirer la conclusion qui était préalablement tirée. Ceci nous permet de questionner le modèle de Stenning et van

Lambalgen; bien que ce dernier cerne la non-monotonie de la suppression des inférences, il ne rend pas compte des différents degrés par lesquels la suppression peut être observée. Inversement, le modèle d'Oaksford et Chater modélise les degrés de croyance qu'on peut avoir en des propositions, mais ne cerne pas tout le phénomène de non-monotonie au sein d'un même système formel. La question que nous nous posons est donc la suivante : est-il possible de marier incertitude et non-monotonie dans un même modèle? Nous proposons que oui et que, pour ce faire, on doit modifier la logique non monotone proposée par Stenning et van Lambalgen. En opérant un affaiblissement de la logique non monotone de Stenning et van Lambalgen, nous défendons qu'on peut obtenir un système formel qui modélise toutes les observations à propos du phénomène de suppression des inférences, soit la non-monotonie via la logique non monotone, l'incertitude et les degrés de suppression via une sémantique polyvalente ainsi que certaines règles d'inférence qu'on retrouve en logique floue.

Les logiques polyvalentes, comme celles élaborées par Lukasiewicz, Bochvar ou Kleene, traitent des indéterminations rencontrées régulièrement par les êtres humains en rendant possible un nombre infini de valeurs de vérité pour une proposition (dans l'intervalle $[0, 1]$) (Resher, 1969). Les logiques floues (Zadeh, 1965 et plusieurs autres depuis) traitent non seulement de l'indétermination d'une proposition mais aussi de différents degrés de corrélation entre les propositions, c'est-à-dire qu'elles traitent de différents degrés pour les connecteurs logiques comme la conjonction, la disjonction et l'implication (Klir & Yuan, 1995). Ainsi, nous défendons que l'ajout d'une sémantique polyvalente et de règles d'inférences floues à la logique non monotone de Stenning et van Lambalgen permette de générer un système formel qui appréhende le phénomène de suppression dans toute sa complexité.

0.3. Discussion sur la rationalité

0.3.1. Vers une rationalité contextualisée

Autant dans la théorie des modèles mentaux que dans la théorie de la logique mentale, nous pouvons observer des définitions *a priori* de la compétence logique. Dans le

cas des modèles mentaux, les théoriciens ont postulé que le raisonnement logique se fait via la construction de modèle et ils présentent ainsi une définition de la rationalité comme compétence sémantique *a priori* sous forme de procédure. Dans celui de la logique mentale, des règles logiques sont postulées *a priori* dans la cognition, et la rationalité humaine y est décrite comme une compétence syntaxique *a priori*. C'est ainsi que l'une comme l'autre de ces théories sont appuyées sur une conception *a priori* de la rationalité. Cette vision de la rationalité s'oppose à une vision contextuelle de la rationalité comme celle que présentent Oaksford et Chater (2007; 2009). Ces derniers ont tenté de redéfinir la rationalité humaine en termes d'un traitement optimal de l'information exempt des règles de la logique formelle et adapté au contexte dans lequel on raisonne. Stenning et van Lambalgen, dans la construction de leur modèle du raisonnement, tiennent compte du contexte dans lequel un sujet pensant se trouve dans la vie de tous les jours; une adaptation du raisonnement face à de l'information nouvelle. C'est là tout l'intérêt de leur clause des mondes clos du raisonnement : ils peuvent, par cette clause, modéliser une procédure non monotone de traitement de l'information par les sujets humains. Dans les prochains chapitres, nous montrerons qu'on observe, dans les travaux d'Oaksford et Chater ainsi que dans ceux de Stenning et van Lambalgen, une rupture dans le concept de rationalité : en sortant des modèles classiques du raisonnement, on s'est dirigé vers une vision contextualisée de la rationalité.

D'autre part, les deux théories classiques du raisonnement logique adoptent la logique classique comme norme *a priori* du raisonnement valide et testent le raisonnement humain dans le cadre de cette norme. En créant de nouveaux modèles formels pour le raisonnement humain, Oaksford et Chater (2007; 2009; 2010) ainsi que Stenning et van Lambalgen (2008; 2010) modifient la norme du raisonnement valide. On observe, dans cette évolution de l'interprétation des résultats, une interaction entre la norme et la description; une norme du raisonnement valide a été postulée (logique classique) et elle a été modifiée afin de rendre compte des performances des êtres humains. Au cours de ce mémoire, nous défendrons qu'à travers les approches présentées, on passe d'une normativité purement *a priori* du raisonnement logique vers une norme plus souple, corrigible et adaptée aux observations empiriques. De plus, nous en tirerons des conséquences importantes sur la manière dont on devrait tester le raisonnement humain en psychologie expérimentale.

0.3.2. Perspective adoptée

L'approche de Stenning et van Lambalgen partage deux caractéristiques avec celle d'Oaksford et Chater et en possède une qui lui est propre. D'abord, en utilisant de nouvelles logiques dans leur cadre d'analyse, ce qui était autrefois perçu comme une erreur logique est maintenant perçu comme une performance logiquement valide. Ainsi, ces approches brisent le lien entre rationalité et conformité à la logique classique en utilisant des définitions de l'implication qui diffèrent de l'implication matérielle qu'on retrouve en logique classique. En second lieu, ces approches ont des ambitions explicatives restreintes; alors que les théories des modèles mentaux et de la logique mentale prétendent fournir une modélisation de ce qui se passe explicitement dans la tête des humains, les modélisations bayésiennes et non monotones se restreignent à fournir une manière logiquement viable de générer des réponses similaires à celles observées par les psychologues expérimentaux du raisonnement. Toutefois, une troisième caractéristique est propre à l'approche de Stenning et van Lambalgen. Leur définition de la rationalité est plus parcimonieuse et prudente que celle d'Oaksford et Chater dans la mesure où leur modèle ne porte que sur le raisonnement non monotone. Leurs aspirations étant restreintes, ils n'excluent pas les cas où le raisonnement humain est monotone, déductif, fidèle à la logique classique, comme c'est souvent le cas dans le cadre de raisonnements déontiques (où la prémisse majeure est une règle à suivre).

Notre apport à la modélisation du raisonnement humain se fera dans une perspective qui se rapproche des travaux d'Oaksford et Chater ainsi que de ceux de Stenning et van Lambalgen en ce qui est lié au fossé ontologique entre les modèles et la réalité. Nous défendrons que la norme *a priori* du raisonnement doit être remplacée par un modèle qui peut constituer un ensemble de normes flexibles *a posteriori*. Cette question en soulève une autre que nous tenterons de traiter tout au long de ces pages, soit la question du pluralisme logique appliqué à la psychologie du raisonnement. Autrement dit, nous nous poserons des questions comme : quels sont les critères de désirabilité d'une logique pour la rationalité humaine? Existe-t-il des logiques plus désirables que d'autres dans l'*a priori*? Peut-on et devrait-on adopter une vision de la rationalité comme traitement optimal de l'information?

Une précision quant aux limites des modèles du raisonnement nous semble de mise. Lorsque l'on propose une modélisation du raisonnement humain, en aucun cas il ne devrait être prétendu que les humains utilisent explicitement le système formel mis de l'avant. Il est seulement prétendu que les performances des sujets sont conformes à ce système. Afin de mieux comprendre cette relation entre le système formel et l'agent cognitif, pensons au langage naturel. La langue française, par exemple, constitue un système formel complet (elle est constituée d'un alphabet, d'une grammaire et de règles sémantiques). Les francophones sont, de façon générale, de bons utilisateurs de la langue française. Ainsi, on peut prétendre qu'ils performent conformément au système formel qu'est le français. Toutefois, ils n'ont pas besoin de réciter les règles de la grammaire ni les définitions du dictionnaire afin d'entretenir une conversation avec un locuteur francophone et même que, dans certains cas, ils ne les connaissent pas du tout. Par analogie, on peut concevoir la possibilité pour un agent cognitif de raisonner conformément à un système formel sans même connaître ce système. (Stenning & van Lambalgen, 2008) Ces considérations se rapprochent d'une des thèses que nous défendrons dans ce mémoire; une part du contenu ontologique d'un phénomène ne se retrouve pas dans les modèles qu'on en fait et cette frontière entre modèle et réalité peut servir de restriction quant aux portées explicatives de l'entreprise de modélisation du raisonnement humain.

CHAPITRE 1

LA LOGIQUE MENTALE ET LES MODÈLES MENTAUX

1.1 Logique classique des propositions et inférences de l'implication

Élaborée par Frege (1842-1925) et Russell (1872-1970), la logique classique est bivalente¹, assertorique², vérifonctionnelle³ et monotone⁴. Elle constitue un langage symbolique; les propositions sont représentées par des symboles (p , q , r , etc.) ainsi que les opérations interpropositionnelles ($\&$, \rightarrow , etc.). Une expression complexe comprend également des parenthèses pour ordonner les opérations qui sont faites sur les propositions. Contrairement à la syllogistique aristotélicienne, la logique classique des propositions n'est pas sujette à l'interprétation de la langue naturelle. Par exemple, les propositions ne doivent plus être traduites dans la forme « Sujet copule Prédicat ». Ainsi, la proposition *Marie mange* devient p plutôt que d'être traduite par *Marie est une mangeuse*.

La vérifonctionnalité des opérateurs permet de générer la technique des tables de valeurs de vérité élaborée par Wittgenstein (Wittgenstein, 1993). Chacun des opérateurs de la logique classique a sa table de vérité propre qui constitue sa définition sémantique. Les tableaux ci-dessous sont une représentation exhaustive des tables de vérité des opérateurs.

¹Le vrai et le faux sont ses deux seules valeurs de vérité.

²Elle ne modélise pas les modulateurs d'assertion comme « je pense que », « demain », « il est possible que », etc.

³Ses opérateurs sont des foncteurs de vérité; la valeur de vérité de l'opération est déterminée en fonction des valeurs de vérité des propositions.

⁴Une conclusion tirée est définitive; elle demeure valide indépendamment de l'apport de nouvelles informations.

Tableau 1.2 Table de l'opérateur unaire de négation classique

P	$\neg P$
V	F
F	V

Tableau 1.3 Tables des opérateurs binaires principaux

P	Q	$P * Q$	$P \rightarrow Q$	$P \vee Q$	$P \wedge Q$	$P \leftrightarrow Q$	$P \& Q$	$P \circ Q$
V	V	V	V	V	F	V	V	F
V	F	V	F	V	V	F	F	F
F	V	V	V	V	V	F	F	F
F	F	V	V	F	F	V	F	F

La logique classique comporte de nombreuses lois logiques. Sur le plan sémantique, leur table de vérité est équivalente à la tautologie ($P * Q$). Seulement certaines d'entre elles seront discutées dans ce mémoire :

La loi du tiers exclu : $(P \vee \neg P)$

La loi de non-contradiction : $\neg (P \& \neg P)$

La loi de double négation : $(P \leftrightarrow \neg \neg P)$

Les lois de dualité:

- $(P \vee Q) \leftrightarrow \neg (\neg P \& \neg Q)$
- $(P \& Q) \leftrightarrow \neg (\neg P \vee \neg Q)$
- $\neg (P \vee Q) \leftrightarrow (\neg P \& \neg Q)$

- $\neg(P \& Q) \leftrightarrow (\neg P \vee \neg Q)$

Les 4 types d'inférences fondamentales que l'on peut faire avec des implications (\rightarrow) sont les plus étudiés dans la littérature expérimentale sur le raisonnement logique. Les deux inférences valides (*MPP*, *MTT*) sont des lois logiques alors que les sophismes n'en sont pas (*AC*, *DA*) :

Le *modus ponendo ponens* (*MPP*) : $((P \rightarrow Q) \& P) \rightarrow Q$

Le *modus tollendo tollens* (*MTT*) : $((P \rightarrow Q) \& \neg Q) \rightarrow \neg P$

Le sophisme de l'affirmation du conséquent (*AC*) : $((P \rightarrow Q) \& Q) \rightarrow P$

Le sophisme de la négation de l'antécédent (*DA*) : $((P \rightarrow Q) \& \neg P) \rightarrow \neg Q$

1.2 La logique mentale : une approche syntaxique

Depuis plus de quarante ans, les psychologues du raisonnement étudient les quatre inférences de l'implication. Les observations montrent qu'une majorité des participants sont en mesure de faire les inférences valides, mais une proportion importante d'entre eux ne parviennent pas à éviter les inférences invalides. Plusieurs théories du raisonnement déductif ont été élaborées pour identifier les processus cognitifs impliqués dans les performances observées. L'une des premières qui a été élaborée est la théorie de la logique mentale. La première version de la théorie (Braine, 1978) est une entreprise de modélisation du raisonnement humain qui se base sur la logique classique. La syllogistique aristotélicienne est rejetée puisque trop restreinte; elle ne rend pas compte d'une grande quantité d'énoncés à partir desquels on raisonne. Le modèle proposé par Braine (1978) est un ensemble de règles d'inférence distinctes pour chaque opérateur qu'on retrouve dans le raisonnement déductif de tous les jours (soit les opérateurs de base de la logique classique; non, et, ou, si...alors, si et seulement si). Lorsqu'un raisonneur détient une règle d'inférence dans son bagage mental, il est en mesure de l'utiliser pour faire des inférences valides. Dans cette perspective, l'erreur

logique est due à l'absence de règle d'inférence adéquate dans le bagage de connaissances de l'agent cognitif. De plus, chaque règle constitue un schéma d'inférence qu'on peut combiner avec d'autres pour dériver une conclusion à partir de prémisses. Le point de départ pour la construction des règles d'inférences sont les règles de la déduction naturelle de Gentzen (Gentzen, 1964). Dans son article fondateur de 1978, Braine pose les bases d'un programme de recherche nouveau; la construction d'une logique naturelle. Alors que la logique classique a été construite par les logiciens comme norme du raisonnement valide, la logique naturelle de Braine s'appuie sur la logique classique et la modifie dans le but de construire une théorie du raisonnement déductif humain.

1.2.1 La procédure générale du traitement de l'information

La procédure générale du raisonnement déductif proposée par les théoriciens de la logique mentale comprend trois étapes. D'abord, il y a la reconnaissance de la forme logique des prémisses. Ensuite, les raisonneurs recherchent dans leur répertoire mental les règles d'inférence dont ils croient qu'elles s'appliquent à cette forme logique dans leur répertoire mental et procèdent à la dérivation d'une conclusion qu'ils jugent valide. Finalement, ils traduisent cette conclusion abstraite dans le contenu des prémisses de départ.

1.2.2 Choix de la sémantique

La bivalence est un choix sémantique explicitement assumé par Braine:

« The assumption that natural logic has three truth values has some very counterintuitive consequences for reasoning. The reductio-ad-absurdum method becomes less useful. [...] I know of no empirical data that test whether subject restrict their modes of reasoning in the manner predicted by the tree-valued theory. However, the predictions appear highly counterintuitive, and it seems appropriate that a two-valued logic be assumed until evidence compels otherwise. » (Braine, 1978, p. 11)

Il faudra attendre Stenning et van Lambalgen (2008; 2010) pour voir un choix de sémantique polyvalente dans la construction d'une logique naturelle. Il semblerait donc que la polyvalence soit indésirable étant donné qu'elle peut invalider la preuve par l'absurde⁵. Or, le fondement de la preuve par l'absurde est le caractère involutif⁶ de la négation, qui est conservé dans plusieurs logiques polyvalentes. Avec une sémantique polyvalente, on peut dériver des quasi-tautologies (comme par exemple un énoncé jamais faux, tantôt vrai, tantôt indéterminé) et des quasi-contradictions (comme par exemple un énoncé jamais vrai, tantôt indéterminé, tantôt faux).

1.2.3 La définition de l'implication et le MPP

Dans le cadre de la logique naturelle proposée par Braine, l'implication n'est pas l'implication matérielle de la logique classique. Plutôt que de constituer un opérateur vérifonctionnel, l'implication correspond à la règle d'inférence de base suivante : sur la base de p , on peut conclure q . Cette définition large permet les interprétations de *Si p alors q* comme étant p est la cause de q , q découle de p ou il y a un lien quelconque entre p et q qui permet de conclure q dès que p est établi. Dans ce cadre théorique, le *MPP* est équivalent à la définition de l'implication et constitue la base de tout raisonnement conditionnel. Le *MPP* étant l'inférence la plus rapide et aisée chez les raisonneurs, il devient la règle par excellence pour la construction d'une logique naturelle.

⁵ On fait une preuve par l'absurde lorsqu'on démontre que la négation d'une hypothèse mène à une contradiction. Dans un cadre classique, étant donné la bivalence de la logique, on peut conclure que l'hypothèse non niée est une loi logique.

⁶ L'involution est la propriété logique qui appuie la loi de double négation ($P \leftrightarrow \neg \neg P$) et ce, indépendamment du nombre de valeurs de vérité d'un système logique. La négation des logiques polyvalentes de Lukaziewicz, Bochvar et Klenne ainsi que plusieurs logiques floues est involutive. La preuve par l'absurde (selon laquelle une formule niée menant à une contradiction permet de déterminer que la formule non-niée est tautologique) demeure possible au sein de ces logiques non-classiques.

1.2.4 Les différents modèles

Le modèle « A Natural Deduction System » (ANDS) élaboré dans Rips (1983) s'inscrit dans le programme proposé par Braine, mais contient une procédure explicite pour le *MPP*. Plutôt que de reprendre l'analogie entre l'inférence et la définition de l'implication, Rips propose deux règles distinctes pour la procédure (Rips, 1983, p. 46) :

R1. *Modus Ponendo Ponens*, version *backward*

Conditions: 1. But courant = q

2. L'arbre d'assertion contient SI p, q

Actions: 1. Poursuivre le but pour déduire p

2. Si le but 1 est atteint, ajouter q à l'arbre d'assertion

R1' *Modus Ponendo Ponens*, version *forward*

Conditions: 1. L'arbre d'assertion contient la proposition $x = \text{SI } p, q$

2. x n'a pas été utilisée par R1 ou R1'

3. L'arbre d'assertion contient p

Actions: 1. Ajouter q à l'arbre d'assertion

La distinction *backward/forward* marque une différence sur le plan de la finalité entre deux procédures inférentielles. Dans la version *backward*, dériver le conséquent (q) est un but préétabli et des mesures sont prises pour l'atteindre (dériver p préalablement). Dans la version *forward*, le raisonnement ne dépend que des prémisses, sans avoir de but préétabli.

1.2.5 Les procédures pour le *Modus Tollendo Tollens* et les sophismes

Les sophismes de l'implication et le *MTT* peuvent être le résultat de deux types d'interprétations des prémisses. Dans un cas, les raisonneurs interprètent p et q comme étant

co-occurents. Ainsi, avec la majeure *Si p alors q*, ils concluent *p* sur la base de *q* (AC), non-*q* sur la base de non-*p* (DA) et non-*p* sur la base de non-*q* (MTT). Cette mauvaise interprétation des prémisses conduit à faire les sophismes mais aussi à faire l'inférence valide (MTT) pour les mauvaises raisons (Braine, 1978). Dans le cas où les sujets interprètent bien les prémisses, ils ne font plus les sophismes et produisent le MTT en faisant une preuve par l'absurde (Wason & Johnson-Laird, 1972).

Bien que le cadre théorique soit partagé chez les théoriciens de la logique mentale, plusieurs auteurs proposent des ensembles de règles d'inférences qui diffèrent. Notamment, Johnson-Laird (1975)⁷ propose 12 règles, Braine (1978) en élabore 18, et Rips (1983), 14. Chacun de ces auteurs présente des divergences au niveau de la définition de l'implication ou de l'ensemble des schémas d'inférences.⁸ Malgré les divergences, les théories ont plusieurs points en commun, notamment, qu'il n'existe pas de règle pour les sophismes et que le *MPP* est fondamental dans la cognition humaine. Une autre convergence concerne les causes des erreurs du raisonnement conditionnel, considérées comme étant occasionnées par un manque d'accès à la règle adéquate, ou par la surcharge de la mémoire de travail ou enfin par une mauvaise interprétation des prémisses.

1.3 La suppression des inférences : un effet de contenu

Tel que mentionné en introduction, l'effet de suppression des inférences montre que les quatre inférences de l'implication peuvent être inhibées, tout dépendant des prémisses qui sont présentées (Byrne 1989; Byrne 1991; Dieussaert & al, 2000; Markovits & Potvin, 2001).

⁷ Avant la publication de *Mental Models* (1983), Johnson-Laird défendait un modèle du raisonnement par schémas d'inférences.

⁸ Passer en revue chacune des positions et les détails des divergences entre les auteurs nécessiterait un chapitre entier. Comme ces détails ne sont pas déterminants pour le sujet central de ce mémoire, nous nous contentons de les mentionner ici.

Le caractère additionnel ou alternatif d'une prémisse dépend de l'interprétation que fait le sujet du contenu des prémisses. Par exemple, à partir de la majeure *Si elle a un travail à faire, elle étudiera tard à la bibliothèque*, on ajoute *Si la bibliothèque est ouverte, elle étudiera tard à la bibliothèque*, les inférences valides ne sont plus acceptées. Toutefois, lorsqu'on ajoute à la prémisse majeure *Si elle a un livre à lire, elle étudiera tard à la bibliothèque*, on observe une inhibition des sophismes. Or, dans les deux cas, les prémisses ont la même forme logique ($\text{Si } p \text{ alors } q$, $\text{Si } r \text{ alors } q$) mais les stratégies de traitement de l'information qu'elles engendrent sont différentes. Dans le premier cas, l'antécédent de la prémisse ajoutée (r) semble être interprété comme une condition additionnelle ($p \ \& \ r$) alors que dans le second cas, l'antécédent serait interprété comme une condition alternative ($p \ \vee \ r$). Dès lors, Byrne utilise ces données pour critiquer la théorie de la logique mentale. D'abord, la suppression est un effet de contenu : deux inférences de même forme logique, dont seul le contenu diffère, sont interprétés différemment. Or, si le raisonnement était purement syntaxique, cet effet ne serait pas observé. De plus, les défenseurs de la théorie de la logique mentale soutiennent qu'il n'existe aucune règle d'inférence pour les inférences invalides puisque celles-ci peuvent être corrigées. Ainsi, le même argument peut être fait pour les inférences valides : aucune règle ne peut y correspondre puisque celles-ci peuvent être inhibées. Avant la parution de cette étude, les effets de contenu étaient bien connus (Kuhn, 1977; Wason & Johnson-Laird, 1972). Cependant, les effets observés avaient toujours mené à l'amélioration de la performance des sujets; la falsification d'hypothèse pour les tâches de sélection et le refus des sophismes pour les syllogismes catégorico-déductifs et hypothético-déductifs. Ces données n'étaient donc pas utilisables contre une théorie de la performance qui veut que la compétence logique soit l'utilisation adéquate de règles d'inférences qui respectent l'implication matérielle classique. Toutefois, suite à l'observation de l'effet de suppression des inférences valides comme dans les deux exemples précédents, la théorie de la logique mentale doit être revue pour intégrer cet effet de contenu qui nuit aux performances des raisonneurs. Finalement, cet effet lance plusieurs défis de taille pour l'entreprise de modélisation du raisonnement humain : comment rendre compte de ces effets de contenu par une approche exclusivement syntaxique? ; Comment utiliser des règles d'inférences issues de la logique classique pour modéliser un raisonnement non monotone? Prétendant que ces questions ne sont pas résolubles par une approche syntaxique, Byrne défend que seule une

approche sémantique peut inclure ces phénomènes dans une entreprise complète de modélisation du raisonnement humain.

1.4 Les modèles mentaux de Johnson-Laird et Byrne : une approche sémantique

La toute première version de la théorie des modèles mentaux a été élaborée par Kenneth Craik (1914-1945) et elle postule que les fondements des procédures du traitement de l'information sont la représentation mentale de modèles du monde à petite échelle. Cette notion de modèle a une très grande extension parce que la seule caractéristique qu'ils partagent est un isomorphisme avec les objets ou concepts qu'ils représentent. Cette idée est reprise par la version défendue par Byrne et Johnson-Laird : « Mental models may occur as visual images, or they may not be accessible to consciousness » (Johnson-Laird & Byrne, 1991). Cette version de la théorie, apparue pour la première fois dans la littérature chez Johnson-Laird (1983), est celle sur laquelle nous nous pencherons. Ces auteurs sont encore des références en matière de modèles mentaux en psychologie du raisonnement. La théorie est aujourd'hui bien plus complexe que dans sa première version, elle présente plusieurs ramifications⁹ et rend compte d'un éventail élaboré de phénomènes à propos du raisonnement humain : elle s'applique maintenant aux syllogismes catégorico-déductifs, au raisonnement conditionnel, au raisonnement contrefactuel, au raisonnement modal, à l'imagination¹⁰, etc. En ce qui nous concerne ici, le raisonnement conditionnel est considéré comme se faisant via la construction de modèles mentaux des possibilités dans lesquelles l'antécédent et le conséquent sont avérés ou non. Ces modèles s'apparentent aux trois rangées vraies de la table de vérité de l'implication. Toutefois, les auteurs défendent que la procédure n'est pas vérifonctionnelle; alors qu'une table de vérité représente les valeurs de vérité des

⁹ Pour les débats récents, voir Byrne & Johnson-Laird, 2002; Evans & al. 2005; Barouillet & al. , 2008.

¹⁰ Voir Byrne, 2005.

propositions, les modèles mentaux représentent des états possibles du monde¹¹ (Johnson-Laird & Byrne, 2002).

La procédure du traitement de l'information lors du raisonnement déductif est entièrement sémantique. D'abord, le raisonneur crée un modèle initial où l'information des prémisses est représentée explicitement. Ensuite, une conclusion partielle est tirée. À la suite de cette conclusion, le raisonneur entame une procédure de recherche de contre-exemples. À cette étape, les modèles implicitement présents dans les prémisses sont explicités. Si un contre-exemple est trouvé, la conclusion du modèle initial est refusée. Lorsque le raisonneur croit qu'il n'y a plus de contre-exemple à trouver, il adopte le modèle ainsi que sa conclusion.

1.4.1 Définition de l'implication

La définition de l'implication adoptée est l'implication matérielle de la logique classique. Raisonner conformément à l'implication est l'explicitation des trois modèles implicites qui concordent avec les trois rangées vraies de la table de vérité de l'implication.

Le premier modèle correspond à la première rangée de la table de vérité :

$$p \quad q$$

...

Le second modèle correspond à la rangée 3 :

$$\neg p \quad q$$

Le troisième modèle correspond à la rangée 4 :

$$\neg p \quad \neg q$$

¹¹ Nous reviendrons sur cette question dans la partie critique de la théorie.

Le premier modèle est le premier à être explicité. Les autres sont rendus explicites par un processus moins facile, plus long et plus coûteux que celui du modèle initial.

1.4.2 Les niveaux de sophistication du raisonnement

Entre la formation du premier modèle et l'explicitation complète, Johnson-Laird et Byrne (1991) proposent un niveau intermédiaire de sophistication du raisonnement déductif. Il y aurait donc trois niveaux de sophistication du raisonnement au sein de la théorie. Le premier niveau correspond à la première étape mentionnée ci-dessus (un seul modèle explicite). Le second niveau est l'ajout de notes mentales (« mental footnotes ») qui rappellent les autres possibilités d'occurrence de l'antécédent et du conséquent et qui mentionnent que dans ces possibilités, l'antécédent est nié. Autrement dit, il s'agit de réaliser que l'antécédent affirmé a été traité exhaustivement dans le premier modèle, mais que l'antécédent nié n'est pas encore traité. Le troisième et dernier niveau de sophistication est la représentation explicite de tous les cas vrais de la table de vérité de l'opérateur, y compris les deux cas où l'antécédent est nié.

1.4.3 Procédure pour le *MPP*

Dans le cas du raisonnement *MPP*, la théorie prédit une procédure simple. Devant une prémisse *Si p, alors q*, le raisonneur construit d'abord le premier modèle de la conditionnelle :

p *q*

...

Devant la prémisse *Marie a un travail à faire* (qui correspond à *p* dans le modèle), il est aisé de conclure que *q* est le cas; *Marie étudie tard à la bibliothèque*. La simplicité de cette procédure expliquerait la quasi-universalité du *MPP* dans le raisonnement humain.

1.4.4 Procédure pour le *MTT*

La procédure pour le *MTT*, beaucoup moins répandue dans le raisonnement, s'expliquerait d'une manière un peu plus complexe. Alors que le *MPP* peut se faire avec le premier modèle qui vient à l'esprit des raisonneurs, le *MTT* demande qu'on rende explicite les modèles encore implicites à la première étape du raisonnement conditionnel :

$$\begin{array}{ll} p & q \\ \neg p & q \\ \neg p & \neg q \end{array}$$

Ce n'est qu'en apportant le troisième modèle dans sa mémoire de travail que le raisonneur peut faire une association entre la négation du conséquent (non- q) et la négation de l'antécédent (non- p) et alors conclure que non- p est le cas. Cette différence en matière de difficulté expliquerait le fait que les psychologues observent systématiquement une meilleure performance pour les tâches du type *MPP* que pour celles du type *MTT*.

1.4.5 La procédure avec une prémisse biconditionnelle

La biconditionnelle ($q \leftrightarrow p$) demande l'explicitation des deux modèles qui correspondent aux deux rangées vraies de sa table de vérité :

$$\begin{array}{ll} p & q \\ \neg p & \neg q \end{array}$$

Ce qui entraîne que les raisonnements par *Modus Ponendo Ponens* et par le *Modus Tollendo Tollens* à partir d'une prémisse majeure de type biconditionnelle sont de même difficulté. Cette prédiction est corroborée expérimentalement par Evans (1977).

1.4.6 La procédure pour les sophismes

Les sophismes de l'implication s'expliquent par une procédure inachevée d'explicitation des trois modèles de l'implication, ce qui fait qu'on la traiterait comme une biconditionnelle. Dans ce cas, on conclut p sur la base de q (AC) et $\text{non-}p$ sur la base de $\text{non-}q$ (DA) puisqu'aucun contre-exemple ne vient invalider la conclusion dans les deux modèles de la biconditionnelle. De cette façon, on exclut le modèle $\text{non-}p$ et q pertinent avec une conditionnelle et non avec une biconditionnelle.

Johnson-Laird et Byrne soutiennent que la théorie des modèles mentaux présente plusieurs avantages sur leur rivale classique. Ils soulignent que leur modèle explique mieux et plus de phénomènes, notamment les différences de difficulté entre le *MPP* et le *MTT* à partir d'une majeure conditionnelle, ainsi que le fait que cette différence s'annule lorsque la majeure est une biconditionnelle. La théorie fournit une explication et prédit la différence de difficulté entre les disjonctions exclusives et inclusives; la première requiert la construction de deux modèles mentaux (les deux rangées vraies de sa table de vérité) alors que la seconde en requiert trois (les trois rangées vraies de sa table de vérité).

Le point de rupture entre les deux rivales se trouve, selon nous, au niveau de la suppression des inférences. La théorie de la logique mentale, telle que présente dans la littérature lors de la publication de 1989, est silencieuse quant à la variation des performances à propos de tâches syntaxiquement équivalentes, mais sémantiquement différentes.

1.4.7 Explication de la suppression avec les modèles mentaux

La suppression des inférences ne peut pas s'expliquer par la simple construction des modèles mentaux. La composante interprétative des prémisses est essentielle pour fournir une explication complète du phénomène :

« Background knowledge can affect the construction, fleshing out, and elimination of models (Johnson-Laird & Byrne, 1991), and the interpretative component is crucial to a theory of reasoning if it is to explain the suppression effect. Our key suggestion is that the

provision of an alternative or additional antecedent enables people to represent counterexamples to conclusions explicitly. » (Byrne & al., 1999)

Byrne et al. (1999) soutiennent que dépendamment du contenu d'une prémisse additionnelle, cette dernière peut être interprétée comme condition supplémentaire pour que l'antécédent de la majeure implique son conséquent (ce qui inhibe les raisonnements valides). Incidemment, l'antécédent de la majeure devient une condition nécessaire mais insuffisante pour impliquer le conséquent. Par exemple, avoir un travail à faire est nécessaire pour aller à la bibliothèque car cette dernière doit aussi être ouverte. L'antécédent de la prémisse ajoutée peut aussi être interprété comme condition alternative à l'antécédent de la prémisse majeure. Par exemple, avoir un livre à lire est aussi une bonne raison pour aller à la bibliothèque. Ces différences d'interprétation seraient dues au bagage de connaissances des sujets. Toutefois, aucune procédure n'est proposée pour cette partie du raisonnement qu'on peut associer à un raisonnement vers une interprétation des prémisses.¹²

Suivant l'interprétation des prémisses, la construction du premier modèle, dans le cas de la prémisse additionnelle dont l'antécédent est interprété comme condition supplémentaire, commence par un premier niveau d'analyse :

p r q

...

Ici, r représente l'antécédent de la prémisse ajoutée. Cet ajout rend alors possible la production de contre-exemples pour les inférences valides; les cas où p est le cas et q n'est pas le cas invalident le *MPP* et le *MTT*. Ces contre-exemples seraient partiellement représentés au second niveau d'analyse du raisonnement par l'ajout des « mental footnotes » qui rappellent les possibilités d'occurrence des antécédents dans les cas où le conséquent est faux. Les trois dernières rangées représentent les cas qui doivent encore être traités:

¹² La distinction entre le raisonnement vers une interprétation et le raisonnement à partir d'une interprétation proposée par Stenning et van Lambalgen (Stenning et van Lambalgen, 2008) n'était pas présente dans la littérature lors de la publication citée ici, mais ne semble pas avoir été reprise par les tenants des modèles mentaux encore aujourd'hui. Nous verrons cette distinction, ses avantages et comment Stenning et van Lambalgen l'utilisent comme critique des modèles mentaux au chapitre 3.

p	r	q
...	$\neg p$	$\neg r$
	p	$\neg r$
	$\neg p$	r

Le troisième niveau d'analyse (l'explicitation complète des modèles des cas vrais de la situation) demanderait la construction de sept modèles correspondant aux cas vrais de la table de vérité de la formule $((p \& r) \rightarrow q)$:

p	r	q
$\neg p$	$\neg r$	$\neg q$
$\neg p$	r	$\neg q$
p	$\neg r$	$\neg q$
$\neg p$	$\neg r$	q
p	$\neg r$	q
$\neg p$	r	q

Il est clair que la représentation complète de ces modèles dépasse largement les capacités de la mémoire de travail¹³. Cependant, le second niveau d'analyse est suffisant pour se rappeler la possibilité que p soit le cas mais que r ne soit pas le cas. Ces « mental footnotes » suffisent donc pour refuser de conclure q sur la base de p (et inhiber le *MPP*) ou pour refuser de conclure non- p sur la base de non- q (et inhiber le *MTT*). Dans le cas qui nous intéresse, il s'agit de se rappeler que même si Marie a un travail à faire, il se peut que la bibliothèque soit fermée et dans ce cas, on refuserait de conclure que Marie étudie à la bibliothèque sur la base qu'elle a un travail à faire.

Lorsque l'antécédent de la prémisse additionnelle est interprété comme une condition alternative pour la majeure (un livre à lire comme alternative au travail à faire pour aller à la bibliothèque), le modèle du premier niveau d'analyse est simple:

¹³ Selon la loi de Miller sur la mémoire de travail, un sujet humain peut garder à l'esprit 7 +/- 2 éléments (Miller, 1956).

p r

Au second niveau, le modèle représente les possibilités où le conséquent est le cas alors qu'un seul des deux antécédents est avéré

 $p \quad q$ $r \quad q$

...

Ce modèle fournit déjà le contre-exemple nécessaire à l'inhibition des sophismes; la possibilité où r et q sont avérés est un contre-exemple à la conclusion de p sur la base de q . Par exemple, ce modèle permet de ne pas conclure que Marie a un travail à faire (p) sur la base qu'elle se trouve à la bibliothèque (q) puisqu'elle pourrait y être parce qu'elle a un livre à lire (r).

Le troisième niveau d'analyse de la situation demanderait la construction de 5 modèles correspondants aux cas vrais de la table de vérité de la formule $((p \vee r) \rightarrow q)$:

 $p \quad r \quad q$ $p \quad \neg r \quad q$ $\neg p \quad r \quad q$ $\neg p \quad \neg r \quad \neg q$ $\neg p \quad \neg r \quad q$

Ces modèles excèdent aussi la capacité de la mémoire de travail, mais le second niveau d'analyse est suffisant pour fournir les contre-exemples nécessaires à la suppression des AC et DA. Étant donné la place centrale accordée aux contre-exemples¹⁴, la théorie prédit que les

¹⁴ Cette position a été comparée expérimentalement à d'autres explications du phénomène qui se rangent du côté de la logique mentale. Nous traiterons de cette question, notamment des travaux de Stevenson et Over (1995) et Politzer et Braine (1991) au chapitre 3.

raisonneurs ne sentiront pas le besoin de dépasser le second niveau d'analyse pour résoudre la tâche en question.

1.4.8 Critiques

1.4.8.1 Les niveaux de sophistication du raisonnement

Les niveaux de sophistication du raisonnement sont présentés comme étant trois étapes qui se succèdent au cours d'un même raisonnement déductif. Idéalement, la procédure devrait être complète mais les limites de la mémoire de travail balisent la progression d'un niveau inférieur à un niveau supérieur, empêchant même souvent l'accès au troisième niveau. Les théoriciens des modèles mentaux soutiennent que le niveau intermédiaire est marqué par l'ajout de « mental footnotes » qui indiquent que l'antécédent affirmé a été représenté exhaustivement. Ce dernier point nous semble très problématique; pour savoir que l'antécédent n'est le cas dans aucun autre modèle, on doit préalablement avoir scanné mentalement toutes les possibilités permises par les prémisses. Or, ceci n'est possible que lorsqu'on atteint l'explicitation complète des modèles. Autrement dit, le second niveau de sophistication du raisonnement nécessiterait que le troisième niveau, souvent inatteignable étant donné la lourdeur de la tâche donnée, ait déjà été accompli.

1.4.8.2 Le principe du vrai

L'hypothèse selon laquelle les humains ne construisent que les modèles qui correspondent aux cas vrais de la table de vérité des opérateurs est devenue l'un des principes fondamentaux d'une version récente de la théorie des modèles mentaux

« The principle implies that mental models represent only what is true and not what is false. Moreover, each mental model represents a clause in the premises only when the clause is true within the possibility that the model represents. » (Johnson-Laird & Byrne, 2002, p. 653)

Ce principe du vrai est appuyé par l'idée que les humains auraient tendance à faire le foyer sur l'information vraie et ignoreraient l'information fausse. Cette idée présente, selon nous, un problème conceptuel fondamental et suscite plusieurs questionnements qui demeurent sans réponse. D'abord, comment expliquer que les humains ne construisent pas de modèle du cas faux de l'implication sans prétendre qu'ils aient *a priori* une connaissance de la vérifonctionnalité de l'opérateur qui appuie la détermination de la vérité et de la fausseté de chacune des quatre possibilités de la table? Autrement dit, pour s'intéresser au vrai et non au faux, encore faut-il être en mesure de déterminer ce qui est vrai et ce qui est faux. Or, pour maintenir ce postulat théorique, on doit aussi postuler que les humains ont une connaissance *a priori* de la vérifonctionnalité de l'implication matérielle et qu'ils mettent ces connaissances en œuvre dans chacune de leurs procédures de raisonnement conditionnel déductif. Stenning et van Lambalgen (2008) soutiennent que toute procédure de raisonnement logique contient deux composantes; le raisonnement vers une interprétation des prémisses et le raisonnement à partir de l'interprétation des prémisses. Pour reprendre ces termes, la théorie des modèles mentaux ne fournit aucune hypothèse explicite quant à la procédure vers une interprétation des prémisses qui permettrait aux raisonneurs de faire le foyer sur les modèles mentaux qu'ils construisent, qui correspondent aux cas vrais des tables de vérités des connecteurs qu'ils rencontrent. Or, il appert que la littérature sur la théorie des modèles mentaux, y compris la littérature actuelle, ne peut fournir de réponse claire à ces questionnements.

1.4.8.3 Refus de la vérifonctionnalité

Dans leurs récents travaux, Byrne et Johnson-Laird tentent de faire une distinction entre les tables de vérité des opérateurs et les modèles mentaux des raisonneurs : alors que les tables de vérité représentent des valeurs de vérité, les modèles représentent des états du monde possibles: « Those possibilities are psychologically basic, not truth values ». (Johnson-Laird & Byrne, 2002, p.4) Toutefois, ils affirment que leur modèle se dissocie de la vérifonctionnalité des opérateurs logiques tout en maintenant le principe du vrai. Or, cette position suscite d'autres questionnements à propos du postulat : si les valeurs de vérité sont

étrangères à la cognition humaine, alors comment expliquer que les humains aient une tendance fondamentale à se concentrer sur les états possibles du monde qui rendent l'implication matérielle vraie? En quoi est-ce que la construction de modèles qui correspondent aux cas vrais des tables de vérité n'est elle-pas vérifonctionnelle? Ici aussi, la littérature ne nous apparaît pas en mesure de répondre à ces questions.

1.4.8.4 L'utilisation exclusive de la logique classique

Finally, toute la littérature couverte jusqu'à maintenant dans le présent travail ne considère que la logique classique comme outil formel pour aborder le raisonnement humain. Les cadres théoriques ainsi que les protocoles expérimentaux sont élaborés sans considérer sérieusement des logiques plus complexes. La question soulevée par Braine (1978) à propos de la polyvalence semble avoir été élaguée de la réflexion sur le raisonnement logique. Ces théoriciens n'ont pas utilisé une gamme d'outils formels comme les logiques polyvalentes (Lukasiewicz, Bochvar et Kleene), qui traitent des indéterminations rencontrées régulièrement par les êtres humains, ou les logiques floues (Zadeh 1965), qui traitent non seulement des indéterminations, mais aussi de différents degrés de corrélation entre les propositions¹⁵. Ces logiques sont notamment utilisées en informatique cognitive pour modéliser plus fidèlement le raisonnement humain (Klir et Yuan, 1995). Par exemple, les moteurs de recherche comme Google doivent utiliser des logiques floues afin de modéliser les différents niveaux de corrélation entre les mots clés qu'on leur fournit afin d'ordonner les réponses par ordre de pertinence.

Cependant, la critique envers l'utilisation de la logique classique n'est pas sans voix dans la littérature des 20 dernières années. Notamment, Oaksford et Chater (1991, 1996) posent un regard critique sur l'association entre logique classique et raisonnement dans la littérature dominante en psychologie ainsi que chez certains philosophes (comme Fodor et

¹⁵ Par « degrés de corrélation entre les propositions », nous entendons différents degrés pour les connecteurs logiques comme la conjonction, la disjonction et l'implication.

Pylyshyn). Dès 1991, ils proposent de repenser ce qu'ils nomment l'approche logiciste en sciences cognitives. Ils ont élaboré un programme de recherche où les probabilités bayésiennes seraient l'outil idéal pour aborder la rationalité humaine (Oaksford & Chater 2003, 2007, 2011).

Dans ce chapitre, nous avons vu que la suppression des inférences pose un problème majeur à la théorie de la logique mentale. Cependant, la théorie des modèles mentaux comporte elle aussi des problèmes importants, dont l'utilisation de la logique classique comme norme du raisonnement valide et comme seul outil formel pour modéliser le raisonnement en général. Dans les prochains chapitres, nous examinerons des approches qui ouvrent la porte à l'utilisation d'autres outils formels pour modéliser plus fidèlement le phénomène. L'une de ces approches est le modèle bayésien d'Oaksford et Chater. Le programme bayésien ainsi que son origine, sa portée et ses limites feront l'objet du prochain chapitre.

CHAPITRE 2

RATIONALITÉ BAYÉSIENNE

2.1. Contre le logicisme

Le logicisme est la vision de la cognition humaine comme procédure d'opérations à partir de formules logiques internes. Cette approche comporte trois composantes; i) un langage formel; ii) un ensemble de règles d'inférence syntaxiques; iii) une implémentation « mécanistique » de i) et ii) (Oaksford & Chater, 1991). C'est par le biais du phénomène de la non monotonie dans le raisonnement humain qu'Oaksford et Chater adressent leur critique à l'approche logiciste du raisonnement. En effet, la non monotonie, c'est-à-dire la possibilité de réviser une conclusion tirée à partir d'un ensemble de prémisses face à de la nouvelle information, est fortement présente dans le raisonnement humain de la vie de tous les jours. Une approche logiciste doit donc rendre compte de ce phénomène, ce qu'elle fait en construisant des logiques non monotones. Cependant, selon Oaksford et Chater, les modélisations par le biais des logiques non monotones rencontrent un problème majeur : ces logiques produisent des conclusions disjonctives trop faibles sur le plan informatif (Oaksford & Chater, 2007). Plus loin dans ce chapitre, nous verrons en détail l'étendue et la portée de cette critique.

2.1.1. Les trois niveaux d'explication de Marr

David Marr a proposé que trois niveaux d'explication deviennent les conditions nécessaires pour qu'une théorie puisse prétendre expliquer avec exhaustivité un processus quelconque (Marr, 1982). En accord avec lui, Oaksford et Chater soutiennent qu'une modélisation des processus cognitifs qui se veut explicative doit comporter les deux premiers niveaux proposés.

2.1.1.1. Le premier niveau : le computationnel

Le plus haut niveau d'explication est le niveau computationnel. Il s'agit du « quoi » et du « pourquoi » de la computation; quel est son but et pourquoi est-elle appropriée. Pour ce niveau, Oaksford et Chater prescrivent la spécification d'un langage logique ainsi qu'une théorie de la preuve définie à partir de ce langage. (Oaksford & Chater, 1991, p. 6) Les auteurs présentent l'exemple de la caisse enregistreuse. Le niveau computationnel d'explication du fonctionnement d'une caisse enregistreuse est la concordance entre nos intuitions à propos du fonctionnement de la caisse et la théorie de l'arithmétique. Ainsi, l'élément neutre (0) équivaut à ne pas payer quand on n'achète rien. L'inverse (la négation) est équivalent au remboursement. La commutativité et la distributivité de l'addition sont équivalentes au fait que l'ordre des achats ou le nombre d'additions ne change pas le paiement final. En ce qui concerne les théories du raisonnement évaluées dans ce mémoire, le niveau computationnel correspond à la théorie formelle utilisée comme norme du raisonnement valide. Ainsi, tout système formel complet et cohérent peut tenir lieu de norme du raisonnement valide. Pour la théorie des modèles mentaux, cette norme est la logique classique alors que la théorie présentée dans ce chapitre est normée par le calcul bayésien des probabilités. Au chapitre suivant, nous présenterons une théorie normée par une logique non monotone.

2.1.1.2. Le second niveau : l'algorithmique

Le niveau d'explication intermédiaire est le niveau algorithmique. Il s'agit du « comment » de la computation; comment le processus est-il implémenté? Il s'agit de déterminer comment les intrants et extrants sont représentés et par quel(s) algorithme(s) de transformation ils sont traités. Dans l'exemple de la caisse enregistreuse, le niveau algorithmique est le système de représentation des nombres, soit les nombres arabes. En ce qui concerne ce mémoire, le niveau algorithmique correspond aux processus du raisonnement humain. Ce niveau est ouvert à toutes théories qui prétendent expliquer le traitement de l'information par les sujets humains. Ainsi, serait une explication du raisonnement au niveau

algorithmique peut soutenir que les humains traitent l'information de manière syntaxique, sémantique, heuristique, etc. La théorie des modèles mentaux propose la construction de modèles du monde à petite échelle alors que la théorie présentée ici propose une évaluation des probabilités via un test de Ramsey. Au prochain chapitre, nous présenterons une théorie qui suggère un traitement de l'information au sein d'un monde fermé.

Dans ce chapitre ainsi que le suivant, nous évaluerons les théories du raisonnement humain selon leurs penchants computationnel et algorithmique étant donné la complémentarité de ces deux niveaux. Alors que le premier fournit une norme nécessaire à l'évaluation des performances, le second propose une explication des processus par lesquels le raisonnement conditionnel s'opère.

2.1.1.3. Le troisième niveau : la réalisation physique

Le dernier niveau d'explication est au niveau de la réalisation physique dans laquelle les algorithmes peuvent être implémentés. Dans l'exemple de la caisse, cette étape est la construction de la caisse elle-même.

2.1.2. Où les auteurs se situent

Oaksford et Chater soutiennent que la théorie des modèles mentaux se situe au niveau algorithmique alors que leur théorie se situe surtout au niveau computationnel. Selon eux, toute théorie qui prétend expliquer des phénomènes cognitifs doit comporter une composante computationnelle et une composante algorithmique. Dans la théorie des modèles mentaux (et sa rivale classique, la logique mentale) la logique classique est la norme du rationnel et la théorie qui la soutient est le niveau computationnel d'explication. Les procédures sémantiques, d'une part, ou les règles d'inférence, d'autre part, sont les composantes algorithmiques des théories. Cependant, Oaksford et Chater défendent que les débats classiques se situent exclusivement au niveau algorithmique (la logique classique est la

norme implicite de la rationalité et cette position n'est pas attaquée). En revanche, Oaksford et Chater fournissent une théorie du raisonnement au niveau computationnel, proposant ainsi de changer le niveau explicatif dans lequel le débat se situe. Cette position aura des conséquences au niveau algorithmique de leur théorie et les amènera à introduire les formules du calcul bayésien des probabilités dans leur modélisation du raisonnement. (Oaksford et Chater, 2003; 2007)

2.2. Redéfinition de la rationalité en termes probabilistes : débat au niveau computationnel

2.2.1. Discussion sur la rationalité

Les expériences en psychologie du raisonnement, comme la tâche de sélection de Wason (Wason, 1978), démontrent que les humains se conforment difficilement à la logique classique. Dans un contexte où cette logique est normative pour le raisonnement, on peut conclure que les humains ne sont pas performants logiquement. Toutefois, dans la dernière décennie, le discours qui perçoit la logique classique comme seule norme possible du raisonnement valide a commencé à être questionné. Par exemple, chez Oaksford et Chater (2007, 2010) une autre avenue a été explorée; l'idée que les humains raisonnent conformément au calcul bayésien des probabilités. En créant de nouveaux modèles logiques du raisonnement, Oaksford et Chater tentent de redéfinir la rationalité et ont pour motif la création de modèles qui rendent compte des performances observées. Dans cette initiative, on observe une interaction entre la norme et la description; une norme du raisonnement valide a été postulée et on l'a modifiée afin de se rapprocher des résultats de l'expérimentation.

2.2.2. Le rejet de la logique classique

Oaksford & Chater (2007; 2009, 2010) critiquent d'abord l'utilisation de la logique classique pour modéliser le raisonnement humain. En effet, l'expérience de Byrne est formulée dans le cadre du raisonnement classique. Ceci dit, la logique classique est monotone et, dans ce cadre d'analyse, on peut conclure que les sujets font des erreurs logiques. Dans le

cadre de la logique classique un MPP ou un MTT faits à partir de prémisses données doivent être faits dès que ces mêmes prémisses sont présentes, peu importe le nombre ou la nature des prémisses qu'on ajoute. En ce sens, la logique classique est monotone et tout raisonnement non monotone représente un raisonnement invalide. Or, plutôt que de conclure à l'erreur humaine, les auteurs tentent de redéfinir la rationalité dans un nouveau cadre d'analyse : les probabilités bayésiennes. Ils proposent qu'un modèle probabiliste bayésien du concept de causalité rende compte du raisonnement non monotone. En optant pour un modèle probabiliste, les auteurs prétendent rendre compte de l'incertitude dans le raisonnement ainsi que des degrés de croyance qu'on accorde à des propositions : « ...the problems that we have identified concern how uncertainty is handled in human inference, and probability is the calculus of uncertainty. » (Oaksford & Chater, 2010, p.92). La logique classique étant binaire, elle ne permet que deux valeurs de vérité pour les propositions : vrai et faux, ce qui ne laisse aucune place pour l'incertitude. Or, les humains font face à l'incertitude quotidiennement. Par exemple, l'indétermination des événements futurs ou l'incertitude en mécanique quantique sont des contextes qui poussent à croire en des propositions avec incertitude. Par le biais des probabilités bayésiennes, Oaksford et Chater soutiennent que l'incertitude du raisonnement humain est captée par un modèle.

Dans un cadre bayésien, l'implication matérielle est modulée en tant que probabilité conditionnelle; l'implication *Si p alors q* est interprétée comme étant la probabilité de *q* étant donné *p*. Par exemple, l'implication « Si cet oiseau est un cygne (*p*), alors il est blanc (*q*) » devient « La probabilité qu'un oiseau soit blanc étant donné que c'est un cygne ». Dans ce cas, si les événements sont dépendants, la probabilité qu'un oiseau soit blanc étant donné que c'est un cygne est plus grande que la seule probabilité qu'un oiseau soit blanc. En termes formels, cette probabilité est représentée de la manière suivante; $\Pr(q|p)$. Toutefois, si les événements sont indépendants, la probabilité qu'un oiseau soit blanc étant donné que c'est un cygne est la même que la probabilité qu'un oiseau soit blanc. En termes formels, on représente cette probabilité de la manière suivante : $\Pr(q|p) = \Pr(q)$.

2.2.3. Sélection optimale d'information

Avant de modéliser les inférences conditionnelles classiques (le *Modus Ponendo Ponens*, le *Modus Tollendo Tollens*, le sophisme de l'affirmation du conséquent et le sophisme de la négation de l'antécédent), les auteurs proposent de modéliser les performances aux tâches de sélection de Wason (exposées au chapitre 1) à la lumière de ces notions probabilistes. Ils appellent leur modèle le modèle ODS (sélection optimale d'information). Selon ce modèle, on considère la dépendance et l'indépendance entre deux événements comme équiprobables (0,5 – 0,5). On essaie ensuite de déterminer quelle possibilité est le cas en faisant le moins possible de manipulations (en tournant le moins possible de cartes). Pour ce faire, on tourne les cartes qu'on considère susceptibles de donner le plus d'information. Pour maximiser le gain d'information, on doit tourner les cartes A et 2 (p et q : dans l'exemple qui suivra, p équivaut à « être mordu par une chauve-souris vampire » et q équivaut à « développer des dents pointues »). Cette stratégie de traitement d'information peut être optimale lorsque le raisonnement approprié est abductif¹ et dans des contextes précis, comme lorsque l'hypothèse à tester est un événement rare. La citation qui suit, quoique longue, exprime de manière très explicite ce genre de contexte et est, nous croyons, convaincante quant à la pertinence du modèle ODS :

« Suppose we consider the (rather implausible) conditional: If a person is bitten by a vampire bat (p), they will develop pointed teeth (q). Clearly, we should check people who we know to have been bitten to see if their teeth are pointed (i.e. turn the p card); and, uncontroversially, we can learn little from people who have not been bitten (i.e. do not turn the $\neg p$ card). If we see someone with pointed teeth, it is surely worth finding out whether they have been bitten – if they have, this raise our belief in the conditional, according to a Bayesian analysis (this is equivalent to turning the q card). But it seems

¹ Un raisonnement abductif consiste à fournir la meilleure explication disponible pour un effet observé. Ayant un conséquent comme évidence, on postule un antécédent connu (lien implicatif préalablement fait entre un antécédent et un conséquent). En revanche, un raisonnement déductif consiste à expliciter de l'information présente dans les prémisses (raisonnement fait en logique classique). Dans la vie de tous les jours, la déduction sert à l'organisation des connaissances (tirer de l'information à partir de connaissances) Dans la conclusion, l'information implicitement présente dans le bagage de connaissances devient évidente.

scarcely productive to investigate someone without pointed teeth (i.e. do not turn the $\neg q$ card) to see if they have been bitten. To be sure, it is possible that such a person might have been bitten, which would disconfirm our hypothesis, and lead to maximum information gain; but this has an almost infinitesimal probability. Almost certainly, we shall find that they have not been bitten, and learn nothing. Hence, with rarity, the expected informativeness of the q card is higher than of the $\neg q$ card, diverging sharply from the falsificationist perspective, but agreeing with the empirical data. » (Oaksford et Chater, 2009)

Dans cet exemple, il y a un lien causal entre p (être mordu par une chauve-souris vampire) et q (avoir les dents pointues). Or, l'implication matérielle ne peut modéliser la relation entre p et q . Le fait que, selon la définition de l'implication matérielle, le faux implique tout (ne pas être mordu impliquerait qu'on ait ou non les dents pointues) et le vrai est impliqué par tout (avoir les dents pointues peut être impliqué par le fait qu'on ait été mordu ou non) ne permettrait pas de s'intéresser au lien causal entre p et q . De plus, lorsque l'antécédent (être mordu) est rare, bien que le raisonnement par *Modus Tollendo Tollens* soit possible (cette personne n'a pas les dents pointues, alors elle n'a pas été mordue) il fournit très peu d'information comparativement au nombre de personnes qu'on devrait examiner. Ainsi, il est préférable, dans un premier temps, d'examiner les personnes qui ont été mordues et voir si elles ont les dents pointues. Dans les tâches de sélections de Wason, ceci correspond à tourner la carte A. Dans un second temps, on doit examiner les personnes qui ont les dents pointues et voir si elles ont été mordues, ce qui correspond à tourner la carte 2. Cette attitude, bien qu'elle ne fournisse pas toute l'information qu'on peut retirer du monde, est celle qui a le plus de chance de fournir de l'information sur ce monde. Il est important de rappeler que tourner la carte A et la carte 2 fait partie des réponses typiques. Rappelons que cette réponse est non conforme à la logique classique et omet de falsifier l'hypothèse.

Bien que la notion de biais cognitif ne soit pas, pour tous les auteurs, la marque de l'irrationalité humaine, dans le cadre du falsificationnisme popperien, le biais de confirmation est clairement un producteur d'erreur logique. La « bonne réponse » aux tâches étant de vérifier l'hypothèse (tourner la carte p) et chercher des falsificateurs à cette hypothèse (tourner la carte $\neg q$), un raisonnement conforme au biais de confirmation est illogique et fait fi du critère popperien de scientificité. Ceci dit, Oaksford et Chater montrent qu'en changeant les critères de scientificité et par le fait même, les normes du raisonnement valide,

ce qui était vu auparavant comme une marque d'irrationalité devient tout à fait rationnel. On observe alors, dans les deux interprétations des résultats aux tâches, une corrélation entre la norme scientifique et la norme logique du raisonnement valide. Popper propose que le véritable critère de scientificité d'une théorie soit la falsifiabilité (Popper, 2002), ce qui a des conséquences directes sur la norme logique du raisonnement valide : dans le cadre de la logique classique, raisonner selon l'implication matérielle et rechercher à falsifier des hypothèses en utilisant le MTT. Oaksford et Chater, en revanche, proposent que, dans certains contextes où on recherche des causes, un raisonnement conforme aux probabilités bayésiennes maximise nos chances d'obtenir de l'information sur le monde. Les conséquences sur la norme logique sont que la recherche de falsificateurs est parfois très coûteuse et peu informative.

2.3. *Le simple probabilistic model*

2.3.1. Probabilités et théorème de Bayes

Afin de bien saisir le modèle probabiliste d'Oaksford et Chater, on doit comprendre les notions de base de la théorie des probabilités. Considérons d'abord la loi générale de la probabilité conditionnelle :

$$\Pr(q | p) = \frac{\Pr(p \& q)}{\Pr(p)}$$

Cette loi énonce que la probabilité conditionnelle de q étant donné p est égale à la probabilité de la conjonction de p et q divisée par la probabilité de p . Le tout n'est possible que si la probabilité de p est plus grande que zéro puisqu'une division par zéro est impossible.

Par exemple, si on lance un dé non pipé, la probabilité que le dé tombe sur un 6 est de 1/6. Toutefois, la probabilité conditionnelle que le dé tombe sur un 6 étant donné que le dé est tombé sur un chiffre pair est différente. En effet, il y a trois possibilités que le dé tombe sur un chiffre pair, soit 2, 4 et 6. Ainsi, la probabilité que le dé soit tombé sur 6 s'il est tombé sur

l'un des trois chiffres est de 1/3. Or, cette probabilité correspond au résultat de la loi générale de la probabilité conditionnelle. En effet, $\Pr(6 \& \text{pair})$ est de 1/6 et $\Pr(\text{pair})$ est de 1/2. Suivant la loi suivante (Hacking, 2001, pp. 49-50):

$$\Pr(6 | \text{pair}) = \frac{\Pr(6 \& \text{pair})}{\Pr(\text{pair})} = \frac{1/6}{1/2} = 1/3$$

Une des conséquences de la loi générale des probabilités est la loi de la probabilité totale :

$$\Pr(p) = \Pr(p \& q) + \Pr(p \& \neg q)$$

Selon la loi de la probabilité totale, la probabilité de p est équivalente la probabilité de la conjonction ($p \& q$) additionnée à la probabilité de la conjonction ($p \& \neg q$). Par exemple, la probabilité qu'il pleuve demain est équivalente à la probabilité qu'il pleuve demain et que la terre soit ronde additionnée à la probabilité qu'il pleuve demain et que la terre ne soit pas ronde. Étant donné que l'énoncé ajouté est affirmé une fois et nié une fois, sa probabilité n'influe pas sur la probabilité de l'événement initial.

2.3.2. Règles de Bayes

Alors que la règle générale de la probabilité conditionnelle permet de calculer la probabilité d'un événement étant donné une hypothèse, les règles de Bayes permettent de calculer la probabilité d'une hypothèse étant donné un événement.

La version pour deux hypothèses mutuellement exclusives et exhaustives est la suivante :

$$\Pr(H | E) = \frac{\Pr(H) \times \Pr(E | H)}{(\Pr(H) \times \Pr(E | H)) + (\Pr(\neg H) \times \Pr(E | \neg H))}$$

Autrement dit, la probabilité d'une hypothèse étant donné un événement est équivalente à la probabilité de l'hypothèse multipliée par la par la probabilité de l'événement étant donné l'hypothèse, le tout divisé par la probabilité de l'hypothèse multipliée par la

probabilité de l'événement étant donné l'hypothèse et additionné à la probabilité de l'hypothèse contraire multipliée par la probabilité de l'événement étant donné cette hypothèse contraire. La version généralisée pour un nombre indéterminé d'hypothèses mutuellement exclusives et conjointement exhaustives est (Hacking, 2001):

$$\Pr(H_j | E) = \frac{\Pr(H_j) \times \Pr(E | H_j)}{\sum \Pr(H_i) \times \Pr(E | H_i)}$$

2.3.3. Théorème de Bayes

Dérivé des règles de Bayes, le théorème de Bayes est utilisé pour la modélisation de 3 inférences de l'implication. On ne peut modéliser le *MTT* ainsi que les sophismes AC et DA via la loi générale de la probabilité conditionnelle comme on le fait pour le *MPP*. On doit plutôt utiliser le théorème suivant (Oaksford & Chater, 2007):

$$\Pr(p | q) = \frac{\Pr(q | p) \times \Pr(p)}{\Pr(q)}$$

2.3.3.1. Modèle au niveau computationnel

2.3.3.1.1. Modélisation du *MPP*

D'abord les auteurs mentionnent que la probabilité de la conclusion d'une inférence conditionnelle est équivalente à la probabilité conditionnelle de la conclusion étant donné la prémisse catégorique (Oaksford et Chater, 2007). Pour le *MPP*, cela signifie que la probabilité de la conclusion (q) est équivalente à la probabilité conditionnelle de cette conclusion (q) étant donné la prémisse catégorique (p) :

$$\Pr(q) = \Pr(q | p)$$

La probabilité de la conclusion devient donc équivalente à la probabilité de la prémisse majeure. Or, la probabilité de la prémisse majeure est estimée via le test de Ramsey². La prémisse mineure a d'emblée une probabilité de 1 puisqu'elle est affirmée. Ainsi, la modélisation de l'inférence *MPP* se résume à la probabilité conditionnelle $\Pr(q|p)$ ³.

En ce qui concerne les 3 autres inférences, la formule qui les modélise ne peut être déduite de la même manière. En effet, contrairement à la mineure du *MPP* ($\Pr(p) = 1$), la prémisse mineure du *AC* ($\Pr(q) = 1$), celle du *DA* ($\Pr(\neg p) = 1$) et celle du *MTT* ($\Pr(\neg q) = 1$) ne se réduisent pas à la simple probabilité conditionnelle de la majeure (Oaksford et Chater, 2007, p. 119).

2.3.3.1.2. Modélisation du AC

On fait le sophisme de l'affirmation du conséquent lorsqu'on conclut p sur la base de q . En termes probabilistes, il s'agit d'évaluer la probabilité de p étant donné q . Oaksford et Chater considèrent que les raisonneurs ont une évaluation initiale de la probabilité de l'antécédent ($\Pr(p)$), la probabilité du conséquent ($\Pr(q)$) et de la probabilité conditionnelle

² Si deux personnes discutent de la conditionnelle $A \rightarrow C$ et sont incertaines à propos de A , elles font l'hypothèse que A est le cas et attribuent un degré de croyance en la probabilité conditionnelle de C étant donné A . Pour ce faire, elles ajoutent A à leur ensemble de croyances, ajuste cet ensemble pour le rendre cohérent avec ce nouvel ajout (de manière conservatrice) et déterminent si, au sein de leur nouveau système de croyance, C est probable ou non. Cependant, les auteurs mentionnent un bémol quant à l'ajout hypothétique de A à l'ensemble des croyances. L'ajout hypothétique de A , ou la supposition que A est vrai, n'a pas nécessairement les mêmes conséquences que la croyance réelle en A . Par exemple, une athée qui croit réellement qu'elle a une maladie mortelle peut soudainement croire en l'existence d'un Dieu ou d'une vie après la mort. Cependant, cette même athée ne révisera peut-être pas sa croyance qu'il n'y a pas de vie après la mort si elle suppose qu'elle a une maladie mortelle. Dans ce cas, la croyance en A ou l'ajout hypothétique de A à son ensemble de croyances n'a pas le même effet sur cet ensemble.

³ Sur ce point, les auteurs assument la rigidité. La rigidité est une propriété du calcul des probabilités conditionnelles qui signifie que la valeur de la probabilité conditionnelle $\Pr(q|p)$ avant de savoir que $\Pr(p) = 1$ est la même que postérieurement à cette connaissance. Autrement dit, la probabilité accordée à la prémisse majeure $\Pr(q|p)$ ne dépend pas de la probabilité accordée à la prémisse mineure $\Pr(p)$ (Oaksford & Chater, 2007, p.119).

de la majeure ($\Pr(q|p)$). Ainsi, la formule qui modélise l'évaluation de la probabilité du AC, s'avère être le théorème de Bayes présenté précédemment.

$$\Pr(p|q) = \frac{\Pr(q|p) \times \Pr(p)}{\Pr(q)}$$

2.3.3.1.3. Modélisation du DA

Le sophisme de la négation de l'antécédent est la conclusion que q est faux sur la base du fait que p est faux. En termes probabilistes, il s'agit d'évaluer la probabilité de $\neg q$ étant donné $\neg p$.

$$DA : \Pr(\neg q | \neg p) = \frac{\Pr(\neg q) - ((1 - \Pr(q|p)) \times \Pr(p))}{\Pr(\neg p)}$$

Comme les auteurs ne précisent pas les procédures de dérivations par lesquelles ils ont construit la modélisation suivante, nous en proposons une preuve formelle :

Preuve partie 1

Prouvons d'abord le théorème suivant :

$$TA : \Pr(\neg q \& \neg p) = \Pr(\neg q) - \Pr(p) + \Pr(q \& p)$$

$$1. \Pr(p) - \Pr(\neg q) = (\Pr(p \& q) + \Pr(p \& \neg q)) - (\Pr(\neg q \& p) + \Pr(\neg q \& \neg p))$$

Loi de la probabilité totale

$$2. \Pr(p) - \Pr(\neg q) = \Pr(p \& q) + \Pr(p \& \neg q) - \Pr(p \& \neg q) + \Pr(\neg q \& \neg p)$$

Commutativité de &

$$3. \Pr(p) - \Pr(\neg q) = \Pr(p \& q) - \Pr(\neg p \& \neg q) \quad \text{Élimination de } \Pr(p \& \neg q) - \Pr(p \& \neg q)$$

$$4. \Pr(\neg q \ \& \ \neg p) = \Pr(\neg q) - \Pr(p) + \Pr(q \ \& \ p)$$

Interversion

Preuve partie 2

$$\text{TB : } \Pr(\neg q \ \& \ \neg p) = \Pr(\neg q) - (1 - \Pr(q \mid p)) \times \Pr(p)$$

Supposons que :

$$1. \Pr(\neg q \ \& \ \neg p) = \Pr(\neg q) - (1 - \Pr(q \mid p)) \times \Pr(p)$$

$$2. \Pr(\neg q \ \& \ \neg p) = \Pr(\neg q) - \left(1 - \frac{\Pr(q \ \& \ p)}{\Pr(p)}\right) \times \Pr(p) \quad \text{Loi de la probabilité conditionnelle}$$

$$3. \Pr(\neg q \ \& \ \neg p) = \Pr(\neg q) - \left(\Pr(p) - \frac{\Pr(q \ \& \ p) \times \Pr(p)}{\Pr(p)}\right) \quad \text{Distributivité de } \Pr(p)$$

$$4. \Pr(\neg q \ \& \ \neg p) = \Pr(\neg q) - (\Pr(p) - \Pr(q \ \& \ p)) \quad \text{Élimination de } \Pr(p)/\Pr(p)$$

$$5. \Pr(\neg q \ \& \ \neg p) = \Pr(\neg q) - \Pr(p) + \Pr(q \ \& \ p) \quad \text{Double négation}$$

$$6. \Pr(\neg q \ \& \ \neg p) = \Pr(\neg q \ \& \ \neg p) \quad \text{Remplacement par TA}$$

Preuve partie 3

$$1. \Pr(\neg q \mid \neg p) = \frac{\Pr(\neg q \ \& \ \neg p)}{\Pr(\neg p)} \quad \text{Loi de la probabilité conditionnelle}$$

$$2. \Pr(\neg q \mid \neg p) = \frac{\Pr(\neg q) - ((1 - \Pr(q \mid p)) \times \Pr(p))}{\Pr(\neg p)} \quad \text{Remplacement par TB}$$

CQFD

2.3.3.1.4. Modélisation du *MTT*

Le *MTT* est l'inférence valide consistant à conclure à la négation de p , en se basant sur la négation de q . En termes probabilistes, il s'agit d'évaluer la probabilité de $\neg p$ étant donné $\neg q$. Oaksford et Chater le modélisent via la formule suivante :

$$\Pr(\neg p | \neg q) = \frac{\Pr(\neg q) - ((1 - (\Pr(q | p) \times \Pr(p)))}{\Pr(\neg q)}$$

Comme dans le cas précédent, aucune procédure de dérivation n'est présentée par les auteurs. Nous avons donc construit une preuve de la formule de modélisation du *MTT* :

2.3.4. Preuve du *MTT*

$$\text{MTT : } \Pr(\neg p | \neg q) = \frac{\Pr(\neg q) - ((1 - (\Pr(q | p) \times \Pr(p)))}{\Pr(\neg q)}$$

$$1. \Pr(\neg p | \neg q) = \frac{\Pr(\neg q | \neg p) \times \Pr(\neg p)}{\Pr(\neg q)} \quad \text{Théorème de Bayes}$$

$$2. \Pr(\neg p | \neg q) = \frac{(\Pr(\neg q \& \neg p) / \Pr(\neg p)) \times \Pr(\neg p)}{\Pr(\neg q)} \quad \text{Remplacement par loi de la probabilité conditionnelle}$$

$$3. \Pr(\neg p | \neg q) = \frac{(\Pr(\neg q \& \neg p) \times \Pr(\neg p)) / (\Pr(\neg p) \times \Pr(\neg p))}{\Pr(\neg q)} \quad \text{Distributivité de } \Pr(\neg p)$$

$$4. \Pr(\neg p | \neg q) = \frac{\Pr(\neg q \& \neg p)}{\Pr(\neg q)} \quad \text{Élimination de } \Pr(\neg p) / \Pr(\neg p)$$

$$5. \Pr(\neg p | \neg q) = \frac{\Pr(\neg q) - ((1 - (\Pr(q | p) \times \Pr(p)))}{\Pr(\neg q)} \quad \text{Remplacement par TB}$$

CQFD

2.3.5. La suppression des inférences valides : les exceptions à la règle

Selon Oaksford et Chater, les prémisses additionnelles sont des exceptions à la règle (implication) qui ont une incidence sur les probabilités pertinentes aux raisonnements valides, soit $\Pr(q|p)$, $\Pr(p)$ et $\Pr(q)$. Ainsi, la probabilité conditionnelle de q étant donné p baisse, ce qui fait baisser la probabilité que l'inférence valide (*MPP* ou *MTT*) soit valide aussi.

2.3.5.1. La suppression du *MPP*

La suppression du *Modus Ponendo Ponens* se fait via l'ajout d'une prémisses additionnelle et Oaksford et Chater (2003) font entrer cette prémisses à l'intérieur du calcul des probabilités. Par exemple, si on pense à la conditionnelle *Si je tourne la clé, la voiture démarre*, une prémisses additionnelle peut représenter quelconque événement qui puisse empêcher l'avènement de q bien que p soit le cas, comme le fait que la batterie soit à plat ou que la voiture manque d'essence. Par exemple, si on ajoute la prémisses *Le réservoir à essence est vide*, à la prémisses *Je tourne la clé*, il est moins probable que les raisonneurs concluent que la voiture démarrera, puisque la prémisses additionnelle constitue une exception à l'implication présentée. Cette exception a pour conséquence de réduire l'évaluation de la probabilité conditionnelle de la majeure $\Pr(q|p)$. Or, tel qu'exposé précédemment, cette probabilité s'avère aussi être la probabilité de l'inférence *MPP*. Ainsi, les prémisses additionnelles, en tant qu'exception à l'implication de la majeure, ont pour effet de baisser la probabilité de l'inférence *MPP*. Dans le cas de l'expérience de Byrne, la prémisses majeure est *Si Marie a un travail à faire, alors elle étudie tard à la bibliothèque*, et la mineure est *Marie a un travail à faire*, ce qui engendre le *MPP*. Toutefois, la prémisses additionnelle *Si la bibliothèque est ouverte, alors Marie étudie tard à la bibliothèque* serait interprétée comme une exception possible à l'implication initiale (il est possible que la bibliothèque soit fermée, donc que Marie ne puisse étudier à la bibliothèque même si elle a un travail à faire). Cette interprétation entraînerait la suppression du *MPP*. (Oaksford & Chater, 2007, pp. 123-124)

2.3.5.2. La suppression du *MTT*

L'exception du réservoir à essence vide explique aussi la suppression du *MTT* pour la majeure *Si je tourne la clé, alors la voiture démarre* et la mineure *La voiture ne démarre pas*. En effet, le fait de tourner la clé peut être insuffisant pour faire démarrer la voiture dans le cas exceptionnel où la voiture manque d'essence. Ainsi la probabilité que la voiture démarre puisqu'on a tourné la clé $\Pr(q|p)$ baisse, ce qui a pour effet de réduire la probabilité conditionnelle qui modélise le *MTT*, soit $\Pr(p|q)$. En effet, selon la règle de modélisation du *MTT*, une diminution de la probabilité conditionnelle $\Pr(q|p)$ entraîne une diminution de la probabilité de l'inférence (Oaksford & Chater, 2007, pp 123-124) :

$$\Pr(\neg p | \neg q) = \frac{\Pr(\neg q) - ((1 - \Pr(q|p)) \times \Pr(p))}{\Pr(\neg q)}$$

Supposons que $\Pr(p) = 1$, $\Pr(q) = 1$ et que $\Pr(q|p) = 0.9$

$$\Pr(\neg p | \neg q) = \frac{1 - ((1 - .9) \times 1)}{1} = \frac{1 - (0,1 \times 1)}{1} = 0,9$$

Supposons maintenant que $\Pr(q|p)$ baisse à 0,4 alors que les autres probabilités demeurent les mêmes.

$$\Pr(\neg p | \neg q) = \frac{1 - ((1 - 0,4) \times 1)}{1} = \frac{1 - (0,6 \times 1)}{1} = 0,4$$

On voit donc que la probabilité de l'implication initiale a un effet sur la probabilité de l'inférence *MTT*. Dans l'expérience de Byrne, la prémisses majeure est *Si Marie a un travail à faire, alors elle étudie tard à la bibliothèque*, et la mineure est *Marie n'étudie pas tard à la bibliothèque*, ce qui engendre le *MTT*. Toutefois, la prémisses additionnelle *Si la bibliothèque est ouverte, alors Marie étudie tard à la bibliothèque* serait interprétée comme une exception possible à l'implication initiale (il est possible que la bibliothèque soit fermée, donc que Marie ne puisse étudier à la bibliothèque même si elle a un travail à faire). Comme dans le cas de la suppression du *MPP*, cette interprétation entraînerait la suppression du *MTT*.

2.3.6. La suppression des inférences invalides : antécédents alternatifs

2.3.6.1. La suppression du AC

Les prémisses alternatives sont des antécédents alternatifs (d'autres manières d'impliquer q) qui deviennent des contre-exemples aux raisonnements sophistiques, dont le AC. Par exemple, une alternative à l'implication *Si on tourne la clé, alors la voiture démarre* serait le fait qu'on ait accès aux fils de contacts pour la démarrer. Oaksford et Chater soutiennent que face à ce genre de prémisses, les humains ont moins tendance à conclure que la clé a été tournée étant donné que la voiture démarre. Autrement dit, les antécédents alternatifs tendent à faire baisser l'évaluation de la probabilité que p soit le cas étant donné que q est le cas. Ainsi, ils font baisser la probabilité de faire le AC $\Pr(p|q)$. Dans l'expérience de Byrne, lorsqu'on donne la prémisse majeure *Si Marie a un travail à faire, elle étudie tard à la bibliothèque* ainsi que la mineure *Elle étudie tard à la bibliothèque*, les participants ont tendance à conclure qu'elle a un travail à faire et ainsi commettre le AC. Toutefois, la prémisse alternative *Si Marie a un livre à lire, elle étudie tard à la bibliothèque* ferait baisser l'évaluation de la probabilité que Marie ait un travail à faire étant donné qu'elle étudie à la bibliothèque, ce qui supprimerait le sophisme AC. (Oaksford et Chater, 2007, pp. 122-123)

2.3.6.2. La suppression du DA

Les prémisses alternatives peuvent aussi tenir lieu de contre-exemple au sophisme DA. Tel qu'énoncé précédemment, une alternative à l'implication *Si on tourne la clé, alors la voiture démarre* serait le fait qu'on ait accès aux fils de contacts pour la démarrer. Face à ce genre de prémisses, les humains ont moins tendance à conclure que la voiture ne démarrera pas étant donné qu'on n'a pas tourné la clé (il devient probable que la voiture soit démarrée par l'utilisation des fils de contact). Ainsi, les antécédents alternatifs font augmenter la probabilité que q soit le cas même si p n'est pas le cas ($\Pr(q|\neg p)$). Ceci a comme conséquence de baisser l'évaluation de la probabilité que q ne soit pas le cas étant donné que p n'est pas le cas. Autrement dit, l'augmentation de la probabilité conditionnelle $\Pr(q|p)$ fait baisser la probabilité de faire le DA $\Pr(\neg q|\neg p)$. Dans l'expérience à l'étude, ceci voudrait

dire que lorsqu'on présente *Si Marie a un travail à faire, elle étudie tard à la bibliothèque*, ou *Si Marie a un livre à lire, elle étudie tard à la bibliothèque*, et qu'elle n'a pas de travail à faire, les participants ne concluent pas qu'elle n'étudiera pas tard à la bibliothèque. Le DA serait donc supprimé. (Oaksford & Chater, 2007, pp. 122-123)

2.3.7. Un modèle quantitatif, une interprétation qualitative

La modélisation exposée est sans conteste un calcul trop lourd pour que les humains l'utilisent dans leur raisonnement quotidien. Les auteurs en sont bien conscients et c'est pourquoi ils spécifient que leur modèle (de niveau computationnel) est quantitatif alors que les humains, traitent qualitativement des probabilités (au niveau algorithmique). En aucun cas les auteurs ne prétendent que les humains effectuent un réel calcul des probabilités. Ils considèrent plutôt que l'interprétation des probabilités conditionnelles se rapproche du test de Ramsey :

« Probabilistic models of conditionals reasoning do not, of course, require that people actually carry out the mathematical computations [...] as if they were doing mental arithmetic. Rather we assume that if the processes that implement a Ramsey test respect the laws of probability, at least to some degree of approximation, then, assuming rigidity, people's estimates for the corresponding conditional probabilities should be close to those indicated in these equations. » (Oaksford & Chater, 2007, p. 121)

Ils soutiennent donc que les humains ont la capacité d'évaluer approximativement et qualitativement la probabilité conditionnelle de q étant donné p ($P(q|p)$) ainsi que la probabilité d'occurrence de l'antécédent $P(p)$ et du conséquent $P(q)$. Ces estimations, approximatives et non computationnelles, seraient alimentées par le bagage de connaissances qu'une personne a acquis par ses expériences antérieures.

2.3.8. Entre norme et description

En créant de nouveaux modèles pour le raisonnement humain, Oaksford et Chater ont tenté de redéfinir la rationalité humaine. Plutôt qu'être exclusivement normatifs, ils ont proposé une nouvelle norme du raisonnement valide à la lumière des observations empiriques. On observe donc, dans cette évolution de l'interprétation des résultats, une interaction entre la norme et la description. Au sein de la psychologie traditionnelle du raisonnement (la théorie de la logique mentale et la théorie des modèles mentaux), on a postulé une norme *a priori* (la logique classique). Suite aux performances aux tâches élaborées dans un cadre classique, une approche nouvelle propose une nouvelle norme du raisonnement valide (le raisonnement bayésien) selon laquelle des raisonnements erronés deviennent rationnels et modélisables au niveau computationnel. En ce sens, Oaksford et Chater ont montré que la description expérimentale du raisonnement humain peut s'avérer être un outil fertile et peut alimenter les débats sur la détermination de ce qu'est le rationnel.

2.4. Critique : Les différentes stratégies de raisonnement

Tel qu'exposé dans (Markovits, Lorties Forgues & Brunet, 2010) et (Markovits, Brunet, Thompson, Brisson, 2013), il est possible d'identifier deux types de stratégie de raisonnement conditionnel; une stratégie probabiliste et une stratégie par contre-exemples. La première, liée à la position d'Oaksford et Chater, est une évaluation des probabilités conditionnelles pertinentes à l'inférence. La seconde, plus conforme à la logique classique, consiste à refuser l'inférence sophistique (AC) dès qu'un contre-exemple à l'inférence (même improbable) soit présent. Markovits *et al.* (2013) ont présenté le type de problème suivant:

Sur la planète Kronus, une équipe de géologues ont découvert une variété de roche intéressante, appelée la Trolyte. Ils affirment que sur Kronus, si une Trolyte est chauffée, alors elle dégage du gaz Philoben.

Parmi les 1000 dernières fois qu'ils ont observé la Trolyte, les géologues ont fait les observations suivantes:

900 Trolytes ont été chauffées et ont dégagé du Philoben.

100 Trolytes n'ont pas été chauffées et ont dégagé du Philoben.

À partir de ces informations, Jean a raisonné de la façon suivante :

Les scientifiques affirment que: Si une Trolyte est chauffée, alors elle dégagera du Philoben. Observation: Une Trolyte dégage du Philoben.

Conclusion : La Trolyte a été chauffée.

Le raisonnement présenté est le sophisme AC, et les participants devaient évaluer si le raisonnement est valide ou non. Une personne qui adopte une stratégie probabiliste aura tendance à accepter la conclusion puisque l'exception à la règle est peu probable ($\Pr(\neg p) = 0,1$). Toutefois, une personne qui adopte une stratégie par contre-exemples refusera la conclusion puisqu'un seul contre-exemple à l'implication est suffisant pour ne pas conclure p sur la base de q . Ainsi, dans le cas suivant, la stratégie probabiliste mène au sophisme de l'affirmation du conséquent alors que la stratégie par contre-exemples mène à un raisonnement conforme à la logique classique. De plus, ces observations ne sont pas qu'une différence individuelle. Les deux types de stratégies peuvent être utilisés par un même sujet, dépendamment des circonstances dans lesquelles il se trouve (Markovits, Brunet, Thompson, Brisson, 2013). Au cours d'une même expérience divisée en deux parties, un même participant peut adopter une stratégie probabiliste quand, dans la première partie, un temps de réponse limite lui est imposé. Il peut ensuite adopter une stratégie par contre-exemples dans la seconde partie où il dispose d'un temps de réponse illimité. Ces résultats suggèrent que le traitement probabiliste des inférences conditionnelles est bien présent dans le raisonnement humain, sans toutefois représenter la totalité du raisonnement conditionnel humain. De plus, la présence des deux stratégies n'est pas seulement le résultat de différences individuelles mais est aussi accessible à au moins une certaine partie de la population. Finalement, l'adoption de l'une ou l'autre de ces stratégies semble influencée par le contexte dans lequel le raisonneur se trouve.

Ainsi, pour prétendre modéliser le raisonnement conditionnel dans son ensemble, une théorie du raisonnement doit tenir compte des différentes stratégies de raisonnement entre les individus et chez une même personne, mais aussi de l'influence que le contexte peut avoir sur un raisonneur.

2.5. Probabilités et logique non monotone: débat au niveau algorithmique

2.5.1. Les bayesiens contre les logiciens

Une vaste littérature en psychologie du raisonnement montre que les raisonnements de la vie de tous les jours sont révisables et non monotones. Conséquemment une approche logiciste du raisonnement humain doit rendre compte de ce phénomène. De ce fait, la modélisation du raisonnement non monotone est un problème central en intelligence artificielle et plusieurs auteurs ont tenté d'y répondre en construisant des logiques non monotones. Parmi ces travaux, on compte la « circumscription logic » (McCarthy, 1980), la « default logic » (Reiter, 1980, 1985) et la « non-monotonic logic I » (McDermott & Doyle, 1980). Selon Oaksford et Chater, ces approches présentent un problème majeur : au sein du système formel, les inférences non monotones produisent des conclusions trop faibles sur le plan informatif. (Oaksford & Chater, 1991, p. 14). Afin d'exemplifier cette critique, considérons les prémisses suivantes :

1. Si un œuf qui cuit 5 minutes dans l'eau bouillante est mollet & qu'un œuf (i) est mis dans l'eau bouillante au temps (t), alors l'œuf (i) sera mollet au temps $t + 5$.
2. Si un œuf qui cuit 5 minutes dans une cocotte-minute est cuit dur & qu'un œuf (i) est mis dans la cocotte-minute au temps (t), alors l'œuf (i) sera cuit dur au temps $(t + 5)$.
3. Si un œuf (i) est mollet au temps $(t + 5)$, alors l'œuf (i) n'est pas cuit dur au temps $(t + 5)$.
4. Si un œuf (i) est cuit dur au temps $(t + 5)$, alors l'œuf (i) n'est pas mollet au temps $(t + 5)$.

Les prémisses 3 et 4 énoncent qu'un même œuf ne peut pas être à la fois mollet et cuit dur.

Considérons maintenant que Martin dit à son frère Jean qu'il a mis un œuf à cuire dans l'eau bouillante. Jean ajoute alors la prémisses 5 aux 4 prémisses précédentes :

5. L'œuf (i) est dans l'eau bouillante.

Supposons maintenant que Jean entre dans la cuisine et fait l'observation que la seule casserole présente sur la cuisinière est la cocotte-minute. Il en déduit donc que l'œuf (i) est dans la cocotte-minute. Ainsi, Jean ajoute la prémisse 6 aux 5 précédentes :

6. L'œuf (i) est dans la cocotte-minute.

Ainsi, les prémisses 1 à 6 devraient impliquer que l'œuf est cuit dur au temps $(t + 5)$ et n'est pas cuit dur au temps $(t + 5)$ et donc que les prémisses sont inconsistantes. Cependant, dans un système de logique non monotone, les prémisses ne sont pas considérées inconsistantes. Il s'ensuit plutôt que les conclusions 5 et 6 sont des extensions distinctes des prémisses 1 à 4. En fait, la seule conclusion valide sera une conclusion consistante avec toutes les extensions des prémisses. Le système déduira que la conjonction entre 5 et 6 est invalide et que la disjonction inclusive entre 5 et 6 est aussi invalide puisque les deux situations laissent place à la possibilité que les prémisses 5 et 6 soient vraies en même temps alors que ces dernières s'excluent mutuellement. La seule conclusion qui soit valide devient donc la disjonction exclusive entre 5 et 6, ce qui résulte en la prémisse 7 :

7. Ou bien l'œuf (i) est cuit dur, ou bien l'œuf (i) est mollet.

Cette conclusion semble effectivement très faible sur le plan informatif. Sur la base de l'information « J'ai mis un œuf à cuire », un raisonneur s'attend à savoir quelle sera sa cuisson dans 5 minutes. On peut aisément faire l'inférence dès qu'on connaît le mode de cuisson et les humains résolvent ce genre de problème très aisément. En effet, la croyance que l'œuf est dans l'eau bouillante est aisément révisable; sachant que la seule casserole sur la cuisinière est une cocotte-minute, on la révisé par la croyance que l'œuf est dans la cocotte-minute. Ainsi, la conclusion initiale (l'œuf sera mollet dans 5 minutes) est remplacée par une nouvelle conclusion (l'œuf sera cuit dur dans 5 minutes). L'intelligence artificielle, pour modéliser le raisonnement non monotone de façon satisfaisante, doit fournir un système formel qui expose ce type d'inférence, c'est-à-dire qui a la capacité de réviser une conclusion préalablement tirée sur la base de nouvelles informations. Or, tel que mentionné par Oaksford et Chater, les logiques non monotones citées plus haut ne rencontrent pas ce critère, ce qui les rend inadéquates pour une entreprise de modélisation du raisonnement humain.

2.5.2. Les logiciens contre les bayesiens

Stenning et van Lambalgen (2010) soulignent que le modèle bayésien ne permet pas de capturer la non-monotonie. Dans le cas de la suppression du MPP, le changement de probabilité accordée aux prémisses n'est pas modélisable en termes bayésiens. Oaksford et Chater ont besoin de clauses supplémentaires à leur modèle afin de capturer ce phénomène; ils doivent passer par le langage naturel pour expliquer en quoi certains raisonnements sont non monotones. En effet, le doute envers la probabilité conditionnelle ($\Pr(q|p)$) n'est pas présent au sein de leur modèle. Ces derniers semblent même d'accord avec la critique qui leur est adressée :

« ...we view this as illustrating the knowledge-rich character of reasoning, rather than challenging a Bayesian account. [...] But this does not imply that $P(p|q)$ remains the same if we are told that p , because pragmatic factors allows us to infer a great deal of additional information; and the information can legitimately change $P(p|q)$. » (Oaksford & Chater, 2009)

Cette réponse peut sembler très surprenante, considérant les motivations qui leur ont fait rejeter le cadre de la logique classique. Rappelons que les auteurs refusent d'utiliser la logique classique pour modéliser la tâche de suppression étant donné que c'est une logique monotone. Or, en quoi un modèle bayésien qui ne capture pas la non-monotonie est-il préférable? Pour ce qui est de la non-monotonie, il ne semble y avoir aucun avantage. En revanche, l'utilisation de probabilités comme degrés de croyance afin de modéliser l'incertitude envers une implication semble être un avantage comparé au cadre classique. Toutefois, une question demeure : peut-on trouver un système formel qui rend compte des raisonnements non monotones? Oaksford et Chater semblent se ranger du côté de la pragmatique et du langage naturel et ainsi, restreindre leur entreprise de modélisation. C'est à ce niveau que Stenning et van Lambalgen (une équipe composée d'un logicien et d'un informaticien) ont proposé une logique non monotone qui capturerait le phénomène au sein d'un même système formel et qui ne tomberait pas sous la critique qu'Oaksford et Chater ont fait à propos des approches précédentes. Au prochain chapitre, nous nous pencherons sur ce système de logique non monotone, nous en évaluerons la portée et les limites et nous en proposerons une modification formelle.

CHAPITRE 3

LOGIQUE NON MONOTONE

Après avoir vu l'approche probabiliste du raisonnement dans le chapitre précédent, nous passons maintenant une étude critique du modèle de Stenning et van Lambalgen. Par le moyen de la logique non monotone, ce modèle est une représentation computationnelle du raisonnement. Nous examinerons ce modèle, sa portée, ses limites et nous en proposerons une modification formelle. Finalement, nous proposerons une nouvelle approche normative du raisonnement valide, inspirée des travaux de modélisation examinés dans ce chapitre et le précédent.

3.1. Une réintroduction de la logique au niveau computationnel

3.1.1. Modèles multiples à portée restreinte

Guy Politzer et Jean-François Bonnefon adressent une critique au modèle probabiliste du raisonnement qui se rapproche de celles que nous avons adressées à Oaksford et Chater dans le précédent chapitre :

« [...] a purely probabilistic model lacks the power of executing one basic human skill, the generation of proof. [...] We believe that the solution to this quandary is to be found in a mixed model, accommodating both uncertainty and deduction [...] we know far too little about human uncertainty to definitely adopt a normative model; and this matter should be solved empirically, rather than a priori. » (Politzer et Bonnefon, 2011)

Ici, Politzer et Bonnefon soutiennent qu'un seul modèle du raisonnement est insuffisant. Ensuite, ils se positionnent quant à l'utilisation normative *a priori* de la logique. Nous rappelant que la rationalité humaine est méconnue à bien des égards, ils plaident pour la prudence et la flexibilité des normes que nous imposons aux observations afin de mieux les comprendre.

De plus, Politzer souligne que d'autres outils logiques existent et qu'ils pourraient être utiles pour modéliser les raisonnements déductifs : « To develop new theoretical approaches, psychologists could borrow conceptual tools from formalisms such as non-monotonic logics, possibilistic logic, fuzzy logic, multi-valued logics and the like [...] » (Politzer, 2007)

Comme Politzer et Bonnefon, Stenning et van Lambalgen (2008) critiquent l'application d'un modèle probabiliste à l'ensemble du raisonnement. De plus, Stenning et van Lambalgen soulignent que le modèle d'Oaksford et Chater ne permet pas de modéliser dans son ensemble la non monotonie du raisonnement humain. En effet, Oaksford et Chater ont besoin de clauses supplémentaires à leur modèle. Ils doivent passer par le langage naturel pour expliquer en quoi certains raisonnements sont non monotones. Tel que vu en détail dans le chapitre précédent, le modèle d'Oaksford et Chater ne représente pas les procédures par lesquelles un raisonnement peut être non monotone.

Pour pallier à ces problèmes, Stenning et van Lambalgen proposent un recours à la logique non monotone afin de modéliser ces types de raisonnements, mais sans l'appliquer à tout le raisonnement humain et évitent, par le fait même, l'erreur fondamentale d'Oaksford et Chater. Ainsi, le raisonnement humain n'est plus perçu comme étant représentable par un seul modèle, mais sous plusieurs : « [...] we believe discourse processing has to be understood as at least two-component processes, and the rational analyses of these two processes cannot be the same. » (Stenning & van Lambalgen, 2010). Selon ces auteurs, face à des prémisses (ou de l'information) données, les humains raisonneraient en deux temps; d'abord, ils feraient une interprétation du contexte dans lequel ils se trouvent ainsi que la détermination du type de raisonnement qu'on doit adopter et ensuite, ils raisonnent à propos des prémisses. Ces deux processus de raisonnement seraient fondamentalement orientés par l'information sur le monde qui est mise à la disposition de l'agent cognitif. Autrement dit, l'interprétation des prémisses et la résolution du problème ne peuvent être effectués indépendamment du bagage de connaissances et du contexte dans lequel se trouve un raisonneur.

3.2. Le modèle

Au moyen d'une logique non monotone, Stenning et van Lambalgen prétendent réussir à capturer le phénomène de non-monotonie au sein d'un même modèle, ce que le modèle bayésien d'Oaksford et Chater ne pouvait pas faire. Les auteurs défendent que le raisonnement se ferait au sein d'un monde sémantique (le *closed-world reasoning*). Une fois incluse dans leur système formel, cette notion est essentielle pour capturer la non-monotonie du raisonnement.

3.2.1. Le *closed-world reasoning*

Tel qu'énoncé précédemment, l'élément central du modèle de Stenning et van Lambalgen est le *closed-world reasoning*. Ce type de raisonnement en monde fermé veut que l'information qui n'est pas mentionnée soit considérée comme inexistante par défaut. Par exemple, si une personne est à la gare et consulte l'horaire des départs de trains, elle considérera qu'il n'y a aucun départ de train entre deux départs listés : dans son raisonnement, seuls les départs listés sont considérés comme existants. Nous verrons que cette notion de raisonnement au sein d'un monde fermé, une fois incluse dans un système formel, est essentielle pour capturer la non-monotonie du raisonnement : « A form of closed world reasoning [...] must be developed which allows an agent to set the conditional probabilities to [...] change [...] value if relevant new information comes in. » (Stenning et van Lambalgen, 2010) (mon emphase).

3.2.2. Définition de l'implication

Stenning et van Lambalgen proposent que l'implication *Si p, alors q* est habituellement interprétée par les agents cognitifs comme signifiant *Si p et rien d'anormal, alors q*. En termes formels, ils représentent l'implication comme étant $(p \ \& \ \neg ab) \rightarrow q$. Étant donné la notion du *closed-world reasoning*, si rien n'est dit à propos de *ab*, on considère qu'il n'y a

rien d'anormal : si l'implication est fautive (c'est-à-dire que l'antécédent est vrai mais le conséquent est faux), on considère que quelque chose d'anormal s'est passé. À cette définition, les auteurs ajoutent qu'on considère que les antécédents mentionnés sont les seuls antécédents qui peuvent impliquer le conséquent en question. Ainsi, pour la formule $p \rightarrow q$, si p est faux, alors q est faux. De plus, si q est faux, p est faux. Il y a donc une rupture avec la définition de l'implication matérielle en logique classique. Sur le plan formel, le symbole de l'implication devient $p \leftrightarrow q$. Cette définition de l'implication peut sembler être la définition de la biconditionnelle classique, mais les auteurs définissent la conditionnelle comme étant non seulement différente de l'implication matérielle, mais comme étant des « licences for inferences rather than truth-functional connectives ». (Stenning et van Lambalgen, 2008, p. 196). Cette clause est une forme du *closed-world reasoning* que les auteurs appellent la complétion. En ajoutant la clause relative à l'anormalité, la conditionnelle devient $(p \& \neg ab) \leftrightarrow q$. De plus, lorsque q a plusieurs antécédents possibles connus (par exemple, lorsqu'on sait $p \rightarrow q$ et $r \rightarrow q$ et que q est le cas, on considère que nécessairement, l'un des antécédents connus en est la cause (p ou r) : on ne postule pas l'existence d'un autre antécédent inconnu. Autrement dit, on ferme l'univers sémantique des possibilités à celles qu'on connaît déjà. Cette notion est une forme du *closed-world reasoning* que les auteurs appellent la contrainte d'intégrité.

Cette définition de l'implication sera essentielle pour modéliser les quatre inférences conditionnelles ainsi que leur suppression. De plus, cette définition a comme conséquence que ces inférences et leur suppression, incluant les deux sophismes, deviennent des inférences valides au sein de leur modèle.

3.3. Explication de la suppression

3.3.1. Les inférences *forward* : MPP et DA

3.3.1.1. Modélisation du MPP

Selon le modèle de Stenning et van Lambalgen, lorsqu'on présente la prémisse majeure $p \rightarrow q$ ainsi que la mineure p , le raisonneur entame un raisonnement vers une interprétation

des prémisses. Représentées par le système logique de Stenning et van Lambalgen, les prémisses deviennent:

$$\begin{array}{l} (p \& \neg ab) \leftrightarrow q \\ p \end{array}$$

Comme aucune anormalité possible n'est mentionnée, le *closed-world reasoning* permet de considérer qu'il n'y en a aucune. Formellement, cette clause se représente ainsi:

$$\perp \leftrightarrow ab$$

Étant donné que p est affirmé dans la mineure et que ab est nié, l'antécédent $(p \& \neg ab)$ est vrai. Comme p et q doivent avoir la même valeur de vérité, q découle directement du raisonnement à partir de l'interprétation des prémisses. Ainsi, on conclut q sur la base des prémisses majeures et mineures, ce qui équivaut au raisonnement *MPP* classique (Stenning & van Lambalgen, 2008).

3.3.1.2. Suppression du MPP

Dans le cas de la suppression du *Modus Ponendo Ponens*, lorsque les sujets sont devant l'implication *Si Marie a un travail à faire, elle étudie tard à la bibliothèque*, ils l'interprèteraient comme signifiant *Si Marie a un travail à faire et que rien d'anormal ne s'est produit, elle étudie tard à la bibliothèque* $((p \& \neg ab) \leftrightarrow q)$. Lors de l'ajout d'une prémisses additionnelle telle que *Si la bibliothèque est ouverte, elle étudie tard à la bibliothèque* $((r \& \neg ab) \leftrightarrow q)$, l'antécédent de chaque implication devient une anormalité possible pour l'autre implication. Autrement dit, le fait que Marie ait un travail à faire n'implique pas nécessairement qu'elle soit à la bibliothèque, car la bibliothèque peut être fermée. De même, le fait que la bibliothèque soit ouverte n'implique pas nécessairement que Marie s'y trouve, car il se peut qu'elle n'ait pas de travail à faire. Formellement, les deux implications $(p \& \neg ab) \leftrightarrow q$ et $(r \& \neg ab) \leftrightarrow q$ deviennent $(p \& r) \leftrightarrow q$. Ainsi, lorsque les sujets raisonnent avec une prémisses additionnelle, ils traitent de $((p \& r) \leftrightarrow q)$ et p . Ils ne concluent pas q sur la base de p car rien n'assure que la condition supplémentaire soit remplie

(rien n'assure que r soit vrai). Le *MPP* est alors supprimé (Stenning & van Lambalgen, 2008).

3.3.1.3. Modélisation du DA

En ce qui a trait au raisonnement DA, la majeure $p \rightarrow q$ et la mineure $\neg p$, suite au raisonnement vers une interprétation des prémisses, deviennent :

$$\begin{array}{l} (p \ \& \ \neg ab) \leftrightarrow q \\ \neg p \end{array}$$

Tel qu'énoncé précédemment, le *closed-world reasoning* fait en sorte que face à ces seules prémisses, on conclut qu'aucune anomalie possible n'est le cas :

$$\perp \leftrightarrow ab$$

Ainsi, comme l'antécédent (p) est faux et que q et p doivent avoir la même valeur de vérité, le raisonnement à partir d'une interprétation des prémisses se fait directement et on conclut que q est faux, ce qui équivaut au DA. Rappelons qu'au sein du programme de Stenning et van Lambalgen, compte tenu de la complétion, ce raisonnement est logiquement valide.

Les nombreux cas où les raisonneurs refusent le DA s'expliqueraient par un refus d'appliquer le *closed-world reasoning* à q . Autrement dit, les raisonneurs refuseraient de considérer que si aucune autre cause possible pour q n'est mentionnée, alors il n'y en a aucune. Ils postuleraient donc la possibilité d'antécédents alternatifs, ce qui élimine la possibilité de conclure que q est faux sur la base que p est faux (Stenning & van Lambalgen, 2008).

3.3.1.4. Suppression du DA

Selon plusieurs travaux en psychologie du raisonnement, le DA est supprimé lors de l'ajout d'une prémisse alternative. Contrairement à une prémisse additionnelle, l'antécédent d'une prémisse alternative ne constitue pas une anomalie possible. En effet, selon Stenning et van Lambalgen, l'antécédent de la prémisse alternative ne serait pas interprété comme une condition supplémentaire, mais plutôt comme une alternative à l'antécédent de la majeure. Or, devant les prémisses *Si Marie a un travail à faire, elle va à la bibliothèque* ou *Si elle a un livre à lire, elle va à la bibliothèque* les sujets interprètent *Si Marie a un travail à faire ou un livre à lire, elle va à la bibliothèque*. Lors de l'ajout de la mineure *Elle n'a pas de travail à faire*, les sujets ne concluent pas *Elle n'étudie pas à la bibliothèque*, car elle pourrait y être parce qu'elle a un livre à lire. Sur le plan formel, lors du raisonnement vers une interprétation des prémisses, les prémisses $p \rightarrow q$ et $r \rightarrow q$ deviennent d'abord $(p \& \neg ab) \leftrightarrow q$ et $(r \& \neg ab) \leftrightarrow q$ et ensuite $(p \vee r) \leftrightarrow q$. Comme l'antécédent de l'implication résultante est une disjonction, la négation d'une seule proposition ne permet pas de déterminer la valeur de vérité de la formule $(p \vee r)$. On ne peut donc rien conclure sur la base de $\neg p$ et le DA est alors supprimé. (Stenning & van Lambalgen, 2008),

3.3.2. Les inférences *backward* : MTT et AC

3.3.2.1. Modélisation du MTT

Tel qu'énoncé précédemment, la majeure $p \rightarrow q$, suite au raisonnement vers une interprétation des prémisses, devient $(p \& \neg ab) \leftrightarrow q$. De plus, la contrainte d'intégrité permet de déduire qu'étant donné qu'aucun autre antécédent alternatif n'est mentionné, aucun autre n'est le cas. Avec l'ajout de la mineure $\neg q$, étant donné la définition de l'implication, le raisonneur déduit que l'antécédent doit aussi être faux. Pour ce faire, p peut être faux, $\neg ab$ peut être faux, ou les deux. Comme le *closed-world reasoning* permet de déduire qu'aucune anomalie n'est le cas, $\neg ab$ doit être vrai, ce qui ne laisse qu'une seule possibilité : p est faux. Ainsi, on conclut que p est faux sur la base de la négation de q , ce qui équivaut au raisonnement MTT classique.

3.3.2.2. Suppression du *MTT*

Tel qu'énoncé précédemment, le *MTT* est supprimé lors de l'ajout d'une prémisse additionnelle. Ainsi, les prémisses *Si Marie a un travail à faire, alors elle étudie tard à la bibliothèque* et *Si la bibliothèque est ouverte elle étudie tard à la bibliothèque*, se représentent ainsi :

$$(p \ \& \ \neg ab) \leftrightarrow q$$

$$(r \ \& \ \neg ab') \leftrightarrow q$$

Conformément au *closed-world reasoning*, r est interprété comme une condition supplémentaire pour l'implication entre p et q . De même, puisque les deux prémisses sont commutatives, p est aussi une condition supplémentaire pour l'implication entre r et q . Ce faisant, si r n'est pas le cas, alors il y a anormalité (ab), ce qui inhibe l'implication entre p et q . Autrement dit, le fait que la bibliothèque ne soit pas ouverte ($\neg r$) est une condition qui empêche Marie d'étudier à la bibliothèque même si elle a un travail à faire. De même, si Marie n'a pas de travail à faire ($\neg p$), elle n'étudiera pas à la bibliothèque même si cette dernière est ouverte. Sur le plan formel, la négation de l'antécédent d'une prémisse est équivalente à une anormalité au sein de la prémisse jointe :

$$\neg p \leftrightarrow ab$$

$$\neg r \leftrightarrow ab'$$

Lorsqu'on ajoute la prémisse mineure $\neg q$, étant donné la contrainte d'intégrité, le raisonneur déduit que l'un des antécédents (p ou r) doit être faux, mais sans savoir lequel. Si r est faux, il y a anormalité dans la prémisse majeure et si p est faux, il y a anormalité dans la prémisse additionnelle. Ne sachant pas laquelle de ces conditions désamorçantes est à l'œuvre (Marie n'a pas de travail à faire ou la bibliothèque est fermée) on ne peut conclure que Marie n'a pas de travail à faire ($\neg p$) sur la base qu'elle n'est pas à la bibliothèque ($\neg q$). Le *MTT* est alors supprimé. (Stenning & van Lambalgen, 2008).

3.3.2.3. Modélisation du AC

Tel qu'énoncé précédemment, le raisonnement vers une interprétation des prémisses mène à l'interprétation de la majeure $p \rightarrow q$ comme étant $(p \& \neg ab) \leftrightarrow q$. Comme cette seule conditionnelle est présentée lors du raisonnement AC, le raisonneur considère que p est le seul antécédent possible pour q . Sachant, par la mineure, que q est vrai, il déduit que l'antécédent doit être vrai aussi. Comme *closed-world reasoning* entraîne que $\neg ab$ est vrai, le raisonneur peut conclure que p est vrai sur la base de q et ainsi faire le raisonnement AC. Ce raisonnement, comme les trois autres présentés, est un raisonnement valide au sein du système de logique non monotone de Stenning et van Lambalgen.

3.3.2.4. Suppression du AC

Tel qu'exposé précédemment, l'inférence AC est supprimée lors de l'ajout d'une prémisses alternative. Suite au raisonnement vers une interprétation des prémisses, la majeure *Si Marie a un travail à faire, elle étudie tard à la bibliothèque* et la prémisses additionnelle *Si Marie a un livre à lire, elle étudie tard à la bibliothèque*, sont interprétées comme étant $(p \& \neg ab) \leftrightarrow q$ et $(r \& \neg ab) \leftrightarrow q$ respectivement. Comme aucune anomalie n'est mentionnée, on les considère inexistantes. Sachant, par la mineure, que q est vrai (que Marie étudie à la bibliothèque), on considère que l'un des deux antécédents (p ou r) doit être la cause de p . Comme on ne peut déterminer lequel des deux est la cause effective, on en déduit la conclusion valide en logique non monotone, soit $(p \vee r)$. Ainsi, la suppression du AC serait le résultat d'un processus plus complexe que le simple refus de conclure p sur la base de q . Il s'agirait de conclure p ou r plutôt que de ne rien conclure du tout. Toutefois, comme cette option n'est pas offerte dans les questionnaires traditionnels, les participants choisiraient de ne rien conclure. Pour appuyer cette partie du modèle, les auteurs réfèrent à une étude (Dieussart & al., 2000) dans laquelle 90,7% des participants choisissent de conclure $(p \vee r)$ lorsqu'on leur en donne l'opportunité (Stenning & van Lambalgen, 2008).

L'avantage de ce modèle sur celui d'Oaksford et Chater est, d'abord dans l'ajout de la clause d'anormalité au sein de la définition de l'implication. Cette légère modification formelle permet à deux raisonnements de même forme logique, qui diffèrent seulement par l'ajout d'une information nouvelle, de mener à des conclusions différentes tout en demeurant logiquement valides. Alors qu'au sein du modèle probabiliste, l'explication de la non-monotonie repose entièrement sur une interprétation du problème par l'agent cognitif, le modèle de Stenning et van Lambalgen la représente formellement dans sa définition de l'implication. Dans un second temps, leur entreprise reconnaît la complexité du raisonnement humain et ne prétend modéliser que le phénomène qui nous intéresse. Malgré ces avantages, le modèle non monotone présente des limites importantes que nous verrons dans les pages suivantes.

3.4. Critique : La suppression est un phénomène gradué

Bien que l'approche de Stenning et van Lambalgen modélise le processus de raisonnement non monotone plus finement que l'approche d'Oaksford et Chater, leur modèle élague tout de même une composante importante du phénomène. Politzer et Bonnefon (2005), ainsi que Stevenson et Over (2001), ont observé que la suppression des inférences conditionnelles ne serait pas franche, mais graduée. Autrement dit, face à de nouvelles prémisses, les sujets seraient tantôt peu enclins, tantôt très peu enclins à tirer la conclusion qui était préalablement tirée. Ceci nous permet de questionner le modèle de Stenning et van Lambalgen : bien que ce dernier cerne la non-monotonie dans la suppression des inférences, il ne rend pas compte des différents degrés par lesquels la suppression peut être observée.

Politzer propose que le phénomène de suppression des inférences soit dû à un processus de raisonnement incertain à partir de prémisses incertaines. Plusieurs études sur le raisonnement montrent que, lors du raisonnement inférentiel comme le *MPP* ou le *MTT*, l'incertitude contenue dans les prémisses se retrouve dans la conclusion (Politzer, 2005).

« The breakdown for the 61% who did not endorse the conclusion in the latter condition was as follows: False 5%; various degrees of doubt 56% (very doubtful 3%;

doubtful 25%; slightly doubtful 28%). In brief, participants who did not endorse the conclusion while being offered response options formulated in terms of doubt almost always used these options rather than the false option. » (Politzer, 2005, p. 6)

Ces observations ont été faites dans le cadre d'une expérience où Politzer a reproduit les tâches de suppression en demandant aux sujets (chez qui le *MPP* a été supprimé) d'évaluer le doute qu'ils ont envers la conclusion faite au préalable, une fois la prémisse additionnelle ajoutée. Les résultats sont concluants : seulement 3% des participants expriment que la conclusion est devenue fausse. Ainsi, il appert que lorsqu'on leur donne la latitude d'exprimer de l'incertitude à propos du *MPP*, les participants en expriment en grande proportion. Cette position est empiriquement supportée par Stevenson et Over (1995), qui ont observé que les différents degrés d'incertitude à propos de l'antécédent de la prémisse additionnelle correspondent aux différents degrés d'incertitude à propos de la conclusion issue du *MPP* :

« When a premise added to the major premise if A, C introduces doubt about the satisfaction of a condition (N) that is necessary for C to hold, the credibility of the major premise is questioned and the conclusion inherits a level of credibility which depends on that of N. » (Politzer, 2005, p. 7-8)

Ces observations permettent de supposer que lorsqu'un participant interprète une conditionnelle, il puisse attribuer différents degrés de connexion entre l'antécédent et le conséquent de la prémisse. Par exemple, dans le cas où seule la prémisse majeure est mentionnée, le fait que Marie ait un travail à faire pourrait être interprété comme impliquant complètement (par exemple, à un degré 1) qu'elle étudie à la bibliothèque. Dans le cas où on mentionne aussi que si la bibliothèque est ouverte, elle étudie tard à la bibliothèque, un doute pourrait être porté sur la force avec laquelle l'antécédent implique le conséquent de la prémisse majeure. On pourrait ainsi dire que le fait que Marie ait un travail à faire implique partiellement qu'elle étudie tard à la bibliothèque. Ceci aurait pour conséquence que la conclusion du *MPP* soit faite avec une incertitude qui dépend du doute qu'a engendré l'ajout de r comme autre condition nécessaire à la réalisation de q .

On peut alors conclure qu'il manque un élément au modèle de Stenning et van Lambalgen. Bien que ce dernier capture la non-monotonie des tâches de suppression, il ne

rend pas compte des différents degrés par lesquels la suppression peut être effectuée et ne modélise pas l'incertitude dans le raisonnement humain. Inversement, le modèle bayésien d'Oaksford et Chater modélise les degrés d'incertitude, mais ne modélise pas toute la procédure du raisonnement non monotone. Or, est-il possible de marier incertitude et non-monotonie au sein d'un même modèle? Nous proposons que oui et que pour ce faire, on doive assouplir la logique non monotone proposée par Stenning et van Lambalgen par le biais des logiques polyvalentes et floues.

Les logiques polyvalentes (Lukasiewicz, Bochvar, Kleene, etc.) peuvent modéliser différents degrés de certitude accordés à une proposition. Les logiques floues (Zadeh, 1965, etc) traitent non seulement de l'indétermination d'une proposition mais aussi de différents degrés de corrélation entre les propositions, c'est-à-dire qu'elles traitent de différents degrés pour les connecteurs logiques comme la conjonction, la disjonction et l'implication. Ces logiques peuvent donc devenir un outil pertinent pour modéliser la graduation du phénomène de suppression. Les prochaines pages seront consacrées à exposer que la combinaison de certaines composantes des logiques polyvalentes et floues ainsi que la logique non monotone de Stenning et van Lambalgen permettrait de générer un système formel qui appréhende le phénomène de suppression dans toute sa complexité.

3.5. Une flouification du modèle

3.5.1. Logique polyvalente (Lukasiewicz, Bochvar, Kleene, etc.)

Les logiques polyvalentes sont un assouplissement de la logique classique dans la mesure où elles permettent de traiter des propositions qui sont indéterminées. Ainsi, on passe de deux valeurs de vérité (0 pour le faux et 1 pour le vrai) à plusieurs valeurs, ou même une infinité de valeurs de vérité. Par exemple, dans une logique trivalente, une proposition peut être vraie (1), fausse (0) ou indéterminée (0,5). Dans une logique infinivalente, une proposition peut prendre une valeur quelconque dans l'intervalle $[0, 1]$. Les connecteurs de base (la négation, la disjonction, la conjonction et l'implication) sont définis en fonction de la nature de l'indétermination dont on traite. Par exemple, les logiques polyvalentes de

Lukasiewicz traitent de l'indétermination ontique¹ alors que celles de Kleene traitent de l'indétermination épistémique². Puisque les phénomènes qui nous intéressent tombent sous le concept de l'indétermination épistémique, la logique trivalente forte de Kleene (Kleene's strong three-valued logic) sera utilisée dans le modèle proposé. Au sein de cette logique, les connecteurs de base ne sont pas définis classiquement. Néanmoins, les cas classiques (les cas dont les propositions ont la valeur 0 ou 1) sont traités comme en logique classique. Les 5 connecteurs de base sont définis ainsi :

La négation : $\neg p = 1 - \text{val}(p)$

Cette définition signifie que la négation d'une proposition équivaut à l'inverse de sa valeur. Ainsi, la négation demeure classique pour les cas classiques : la négation d'une proposition fautive (de valeur 0) équivaut à une proposition vraie (de valeur 1) et inversement. Toutefois, la négation de Kleene permet de nier une proposition plus ou moins fautive (par exemple, de valeur 0,2) qui deviendra une proposition de valeur plus ou moins vraie (de valeur 0,8) et inversement.

La disjonction : $\text{val}(p \vee q) = \max(\text{val}(p), \text{val}(q))$

La disjonction classique équivaut à la plus élevée des valeurs de vérité des propositions. Par exemple, lorsque p est vrai (de valeur 1) et que q est faux (de valeur 0), la disjonction est vraie (1). La disjonction de Kleene respecte ce principe en énonçant que la disjonction équivaut à la plus élevée des valeurs entre p et q . Par exemple, une disjonction entre p (0,2) et q (0,8) équivaut à la valeur maximale, soit 0,8.

La conjonction : $\text{val}(p \& q) = \min(\text{val}(p), \text{val}(q))$

¹ L'indétermination ontique est une indétermination à propos de l'état du monde. Par exemple, la couleur des vêtements que je porterai demain est actuellement indéterminée.

² L'indétermination épistémique est due à l'incapacité humaine de connaître certaines choses du monde. Par exemple, l'incapacité de connaître à la fois la vitesse et la position d'un électron en physique quantique est une indétermination épistémique.

La conjonction classique équivaut aussi à la moins élevée des valeurs de vérité des propositions. Par exemple, lorsque p est vrai (de valeur 1) et q est faux (de valeur 0), la conjonction est fautive (0). La conjonction de Kleene respecte ce principe en énonçant qu'elle équivaut à la moins élevée des valeurs entre p et q . Par exemple, une conjonction entre p (0,2) et q (0,8) équivaut à la moins élevée des valeurs, soit 0,2.

L'implication (ou conditionnelle) : $\text{val}(p \rightarrow q) = \max((1 - \text{val}(p)), \text{val}(q))$

L'implication de Kleene est définie à partir de son équivalence avec l'expression $(\neg p \vee q)$. Pour cette raison, l'implication de Kleene est la disjonction entre la négation de p et q . Ceci entraîne que pour que l'implication soit vraie, la valeur de p ne peut être supérieure à la valeur de q (qui serait le seul cas faux de la disjonction entre non- p et q). On retrouve aussi ce principe en logique classique. En effet, la table de vérité de l'implication classique montre que le seul cas faux (p vrai et q faux) est aussi le seul cas où la valeur de p est supérieure à la valeur de q . Ce principe est respecté par la définition de Kleene, tout en s'appliquant aux propositions de toute valeur de vérité possible.

L'équivalence (ou biconditionnelle) : $\text{val}(p \leftrightarrow q) = \min([\max((1 - \text{val}(p)), \text{val}(q))], [\max(1 - \text{val}(q), \text{val}(p))])$

Tout comme en logique classique, la définition de la biconditionnelle est la définition de l'implication, allant dans les deux sens, soit p implique q et q implique p .

3.5.2. Logique floue

Les logiques floues permettent non seulement une infinité de valeurs de vérité et de degrés de connexion entre les propositions, mais aussi des degrés par lesquels des sujets sont affectés par des prédicats au sein d'une proposition. Un prédicat flou serait par exemple, être jeune, d'âge moyen ou vieux. Ainsi, avec les logiques floues, on peut traiter de propositions comme *Marie n'est plus très jeune*, qui exprime qu'une personne (x) est affectée par un

prédicat (A) à un certain degré, qu'on appelle mesure d'appartenance, et dont la valeur se situe dans l'intervalle $[0,1]$.

3.5.3. Règle générale d'inférence floue

Une flouification d'un système logique permet d'attribuer un degré aux inférences de l'implication comme le *MPP*. Alors qu'en logique classique, soit on fait le *MPP*, soit on ne le fait pas, en logique floue, on peut le faire à un certain degré. Ce degré est exprimé par une valeur qui se situe dans l'intervalle $[0, 1]$ et dépend des valeurs de p et de q dans la majeure, de la valeur de la mineure ainsi que de la définition de l'implication. Étant donné que les logiques floues modélisent des prédicats flous qui affectent des sujets à différents degrés, les propositions p et q deviennent respectivement x est A et y est B . Ainsi, la forme générale du *MPP* flou est la suivante :

$$\frac{\begin{array}{l} \text{Si } x \text{ est } A, \text{ alors } y \text{ est } B \\ x \text{ est } A' \end{array}}{y \text{ est } B'}$$

Le calcul de la valeur de vérité de la conclusion se fait via la règle d'inférence suivante :

$$I. B'(y) = \max (\min [A'(x), R(x, y)])$$

Cette règle exprime que la valeur de vérité de la conclusion est égale à la disjonction de la conjonction entre les mesures d'appartenance de chacun des x à A et leur relation R avec les mesures d'appartenance de chacun des y à B . Toutefois, dans le cas qui nous intéresse, les prédicats ne sont pas des prédicats flous. En effet, il n'y a que deux mesures d'appartenance possibles; la bibliothèque est soit fermée, soit ouverte et Marie a soit un travail à faire, soit pas de travail à faire. On peut donc exprimer les propositions par p et q .

Ainsi, afin d'alléger la notation, la forme générale du *MPP* flou peut se réduire à celle-ci :

$$\frac{\text{Si } p, \text{ alors } q}{p'}$$

$$q'$$

Et la règle d'inférence à ceci :

$$2. q' = \min [(p', (p \rightarrow q))]$$

Dans ce qui suit, nous proposerons un modèle hybride pour représenter une suppression graduée des inférences de l'implication. Notre modèle reprend la définition de l'implication de Stenning et van Lambalgen, traite des opérateurs et des propositions selon la logique trivalente forte de Kleene et traite des inférences en utilisant la règle générale d'inférence floue. Tel qu'énoncé précédemment, la logique polyvalente de Kleene est préférée aux logiques de Lukasiewicz et de Bochvar puisqu'elle traite de l'indétermination épistémique.

3.5.4. Application du modèle à la suppression

3.5.4.1. La suppression du *MPP*

Si on reprend l'explication de Stenning et van Lambalgen pour la suppression du *MPP*, on modéliserait le phénomène comme suit : $(p \& r) \leftrightarrow q$ et p , mais rien n'assure que r est le cas, donc on ne peut conclure que q . Cette explication rendrait bien compte du phénomène si les participants jugeaient que dans ce cas, conclure q est totalement faux. Cependant, il semblerait que la suppression soit due à une incertitude par rapport à la conclusion plutôt qu'à un refus catégorique de conclure. De plus, on a observé que l'incertitude envers la conclusion est proportionnelle à l'incertitude envers l'antécédent de la prémisse additionnelle (Politzer, 2005). Or, en combinant l'explication de Stenning et van Lambalgen avec la forme générale du *MPP* flou énoncée ci-haut, l'antécédent p devient la conjonction entre p et r et l'implication devient une biconditionnelle. La règle d'inférence qui permet de calculer la valeur de vérité de la conclusion du *MPP* devient alors :

$$3. q' = \min [(p' \& r'), ((p \& r) \leftrightarrow q)]$$

Pour le cas qui nous intéresse, nous considérons que la valeur de vérité de p , q et r est 1 dans la mesure où les prémisses sont interprétées comme étant *Si p est vrai, alors q est vrai* et *si r est vrai, alors q est vrai*. La valeur de p' est 1 puisqu'il est énoncé que p est vrai. La valeur de r' dépend du degré de croyance que les sujets lui accordent. En utilisant la règle d'inférence 3, on remarque que la valeur de q' est proportionnelle au degré de croyance en r' . Par exemple, dans un cas où la valeur accordée à r' est .1, la valeur de q' se calculerait ainsi :

$$q' = \min((p' \& r'), ((p \& r) \leftrightarrow q))$$

$$q' = \min((p' \& r'), \max(1 - (p \& r), q) \& \max(1 - q, (p \& r)))$$

$$q' = \min((\min(1, .1)), \max(1 - \min(1, 1), 1) \& \max(1 - 1, \min(1, 1)))$$

$$q' = \min(.1, \max(1 - 1, 1) \& \max(1 - 1, 1))$$

$$q' = \min(.1, \max(0, 1) \& \max(0, 1))$$

$$q' = \min(.1, 1 \& 1)$$

$$q' = \min(.1, \min(1, 1))$$

$$q' = \min(.1, 1)$$

$$q' = .1$$

Lorsque la valeur de r' est .9, q' se calculerait ainsi :

$$q' = \min((p' \& r'), (p \& r) \leftrightarrow q)$$

$$q' = \min((p' \& r'), \max(1 - (p \& r), q) \& \max(1 - q, (p \& r)))$$

$$q' = \min((\min(1, .9)), \max(1 - \min(1, 1), 1) \& \max(1 - 1, \min(1, 1)))$$

$$q' = \min(.9, \max(1 - 1, 1) \& \max(1 - 1, 1))$$

$$q' = \min(.9, \max(0, 1) \& \max(0, 1))$$

$$q' = \min(.9, 1 \& 1)$$

$$q' = \min(.9, \min(1, 1))$$

$$q' = \min(.9, 1)$$

$$q' = .9$$

Contrairement aux modèles auxquels nous nous sommes intéressés dans les précédents chapitres, au sein de ce modèle, l'incertitude envers la conclusion est proportionnelle à l'incertitude envers l'antécédent de la prémisse additionnelle. Ce modèle est moins une

modélisation complète qu'une nouvelle avenue de recherche. Dans des travaux ultérieurs, il serait intéressant d'en vérifier la portée et les limites. D'une part, il serait intéressant de voir dans quelle mesure ce modèle est implémentable dans un programme informatique. D'autre part, l'avenue de recherche proposée par cette ébauche de modélisation est une explication computationnelle du raisonnement non monotone et sa composante algorithmique reste encore à être développée. Dans la mesure où le modèle proposé est trop complexe pour être mis en œuvre par des sujets humains, il ne peut être considéré que comme computationnel et non algorithmique. Il appert néanmoins qu'en opérant une flouification de la logique non monotone de Stenning et van Lambalgen, on obtienne un modèle qui capture toutes les observations à propos du phénomène de suppression des inférences, soit la non-monotonie via la logique non monotone, l'incertitude via un assouplissement des connecteurs de base et les degrés de suppression via la règle générale d'inférence floue. De plus, un tel modèle est adaptable à différents types de contenu dans la mesure où il peut tenir compte des cas où les prémisses présentent des prédicats flous comme ceux qui expriment des propriétés qualitatives telles que jeune, vieux, froid, chaud, etc.

3.6. Vers une rationalité multiple et contextualisée

Outre leur différence formelle claire, les modèles examinés dans ce chapitre et les précédents sont supportés par des conceptions distinctes du rationnel. Dans les théories classiques du raisonnement telles que les modèles mentaux et la logique mentale, la seule norme du raisonnement valide est la logique classique. Chez Oaksford et Chater, on observe une redéfinition de la rationalité humaine en termes d'un traitement optimal de l'information exempt des règles de la logique formelle et adapté au contexte dans lequel on raisonne. Stenning et van Lambalgen, dans la construction de leur modèle du raisonnement, tiennent compte du contexte dans lequel un sujet pensant se trouve dans la vie de tous les jours. C'est là tout l'intérêt de leur clause des mondes clos du raisonnement; ils modélisent une procédure non monotone de traitement de l'information pour un sujet humain. On observe donc, dans les travaux d'Oaksford et Chater ainsi que dans ceux de Stenning et van Lambalgen, une rupture avec le concept classique de rationalité et une tendance vers une vision contextualisée de la rationalité. Oaksford et Chater ont une conception de la rationalité contextualisée en ce

qu'ils proposent une modélisation adaptée à la recherche des meilleures explications possibles dans un environnement incertain. Stenning et van Lambalgen vont en ce sens par la clause du *closed-world reasoning* et en proposant une norme applicable au phénomène qui nous intéresse, mais laissant ouverte la possibilité d'utiliser d'autres logiques pour modéliser d'autres phénomènes du raisonnement humain.

CONCLUSION

Le point de départ de ce mémoire est la suppression des inférences conditionnelles observée pour la première fois par Byrne (1989). L'ajout d'une prémisse additionnelle (en conjonction avec la majeure) a pour effet d'inhiber les inférences valides, alors que l'ajout d'une prémisse alternative (en disjonction avec la majeure) inhibe les sophismes chez une proportion importante de participants.

Nous avons examiné trois tentatives d'explication et de modélisation du phénomène, toutes intégrées dans des visions distinctes des processus du raisonnement humain et de la norme du rationnel. La première, la théorie des modèles mentaux de Byrne et Johnson-Laird a été présentée comme une théorie classique du raisonnement dont la rivale était la théorie de la logique mentale. Cette dernière a été présentée comme introduction historique à la théorie des modèles mentaux puisqu'elle ne fournit pas d'explication au phénomène qui nous intéresse et n'est plus défendue aujourd'hui. La seconde tentative de modélisation, plus récente, est l'approche probabiliste d'Oaksford et Chater. La dernière, aussi récente, est le programme de logique non monotone de Stenning et van Lambalgen. Nous avons soutenu que cette dernière représente le plus fidèlement le phénomène à l'étude et ce, sans tomber sous les critiques que nous avons adressé à ses théories concurrentes. Finalement, nous avons proposé une modification à l'approche de Stenning et van Lambalgen pour pallier à la critique que nous leur avons adressée.

Selon la théorie des modèles mentaux, face à une tâche de raisonnement conditionnel, les participants auraient en tête des modèles qui représentent une partie ou la totalité des trois cas vrais de la table de vérité de l'implication matérielle et non le cas faux. Un modèle initial représente explicitement le premier cas vrai de la table de vérité et implicitement les autres cas vrais possibles. Le premier modèle correspond à la première rangée de la table de vérité ($p \ \& \ q$), le second modèle correspond à la quatrième rangée ($\neg p \ \& \ \neg q$) et le troisième modèle à la troisième rangée ($\neg p \ \& \ q$). Le premier modèle est le premier à être explicité. Les autres sont rendus explicites par un processus moins facile, plus long et plus coûteux que les auteurs appellent le *fleshing out*. Ainsi, on raisonne conformément à l'implication matérielle lorsqu'on se représente explicitement les trois modèles qui concordent avec les trois rangées

vraies de la table de vérité de l'opérateur. De plus, le raisonnement logique s'opère selon une procédure en trois étapes basée sur deux principes fondamentaux. La première étape est la construction d'un premier modèle. La seconde est la recherche de contre-exemples. La troisième, lorsqu'aucun contre-exemple n'est trouvé, est l'adoption du modèle et de sa conclusion. Dans ce cadre d'analyse, l'erreur logique est une erreur de performance due au fait que la tâche en question demande un nombre de modèles mentaux qui dépasse les capacités de la mémoire de travail.

Byrne et al. (1999) soutiennent que dans le cas d'une prémisse additionnelle, l'antécédent de la majeure soit une condition insuffisante pour impliquer le conséquent. Toutefois, cette explication ne distingue pas la théorie des modèles mentaux de ses concurrentes puisqu'on la retrouve chez Oaksford et Chater ainsi que chez Stenning et van Lambalgen. L'approche des modèles mentaux se distingue par l'explication des procédures algorithmiques de résolution de problème : une analyse complète du problème demanderait la construction de sept modèles mentaux correspondant aux cas vrais de la table de vérité de la formule $((p \ \& \ r) \rightarrow q)$. Le *fleshing out* complet de ces modèles dépasse largement les capacités de la mémoire de travail mais le second niveau d'analyse est suffisant pour se rappeler la possibilité que p soit le cas mais que r ne soit pas le cas. Ces *mental footnotes* suffisent donc pour refuser de conclure q sur la base de p (et inhiber le *MPP*) ou pour refuser de conclure non- p sur la base de non- q (et inhiber le *MTT*).

L'avantage principal de la théorie des modèles mentaux est qu'elle fournit une explication complète du raisonnement au niveau algorithmique alors que les autres approches que nous avons vues sont incomplètes à ce niveau.

Outre les critiques que nous avons formulé à l'égard de cette approche, nous avons soutenu la critique d'Oaksford et Chater en regard à l'association *a priori* entre logique classique et raisonnement dans la littérature dominante en psychologie, dont la théorie des modèles mentaux fait partie. Le problème central de cette association est qu'elle est faite sans évaluer l'intérêt des autres outils formels à notre disposition pour approcher le raisonnement, ce qui a pour conséquence deux problèmes importants : d'abord, les performances non conformes à la logique classique sont considérées comme erronées sans justification explicite

et ensuite, la corrélation entre le modèle et les performances n'est pas aussi grande qu'elle peut l'être en utilisant des logiques différentes. Cette critique a été un point tournant dans ce mémoire puisqu'elle marque la transition entre une théorie classique du raisonnement et des travaux plus récents comme la théorie probabiliste d'Oaksford et Chater et le modèle non monotone de Stenning et van Lambalgen.

Chez Oaksford et Chater, l'idée que les humains raisonnent conformément au calcul bayésien des probabilités est élaborée en continuité avec leur critique de la logique classique comme norme par défaut du rationnel. En créant de nouveaux modèles logiques pour le raisonnement humain, ils redéfinissent la rationalité humaine et modifient la norme du raisonnement valide. Cela est motivé par le désir de créer des modèles qui rendent compte des performances observées. Dans cette initiative, on observe une interaction entre la norme et la description; une norme du raisonnement valide a été postulée et on l'a modifiée afin de se rapprocher des résultats. La définition de l'implication passe alors de l'implication matérielle classique à la probabilité conditionnelle, soit la probabilité de q étant donné p ($P(p | q)$) incluse dans un raisonnement abductif.

Le modèle bayésien du raisonnement comporte trois points forts. D'abord, Oaksford et Chater ont montré que leur modèle est fidèle aux observations empiriques de suppression des inférences. Ensuite, la logique probabiliste permet de représenter l'incertitude présente dans le raisonnement, ce que la logique classique ne peut pas faire. Finalement, Oaksford et Chater ont été les premiers à fournir une explication du raisonnement au niveau computationnel. Cependant, nous avons vu que leur explication au niveau algorithmique doit être élaborée, ce qui rend les modèles mentaux plus intéressants que l'approche probabiliste à ce niveau. En plus de cette critique, nous avons défendu qu'un modèle probabiliste ne cerne pas les procédures par lesquelles un raisonnement non monotone est orchestré : l'interprétation lors de l'ajout de nouvelles informations n'est pas formellement représentée.

L'approche de Stenning et van Lambalgen ne tombe pas sous cette critique et est celle qui modélise le plus fidèlement le phénomène à l'étude. Leur programme de logique non monotone a donc été présenté comme détenant l'explication computationnelle la plus complète. La capacité à raisonner sur un monde fermé est incluse dans leur définition de

l'implication et est essentielle pour capturer la non-monotonie du raisonnement. Les auteurs proposent que l'implication *Si p, alors q* soit interprétée comme signifiant *Si p et rien d'anormal, alors q*. Si rien n'est dit à propos de *ab*, on considère qu'il n'y a rien d'anormal : si l'implication est fautive (c'est-à-dire que l'antécédent est vrai mais le conséquent est faux), on considère que quelque chose d'anormal s'est passé. Dans ce contexte, une prémisse additionnelle devient une anomalie possible pour la majeure, ce qui affaiblit le lien implicatif en l'antécédent et le conséquent de la majeure. Les inférences valides sont alors supprimées. Ainsi, par le biais d'une logique non monotone augmentée de la clause de « closed-world reasoning » ($\neg ab$), Stenning et van Lambalgen ont réussi à capturer le phénomène de non-monotonie au sein d'un même modèle, ce que le modèle bayésien d'Oaksford et Chater ne pouvait pas faire. De plus, leur utilisation parcimonieuse de la logique non monotone comme norme du raisonnement rend possible l'utilisation d'autres outils formels pour modéliser fidèlement d'autres phénomènes du raisonnement.

Bien que nous ayons présenté cette approche comme la meilleure explication computationnelle du raisonnement non monotone, nous avons formulé une critique importante à son égard. Nous basant sur les travaux de Politzer et Bonnefon ainsi que ceux de Stevenson et Over, nous avons souligné que la suppression des inférences est plus complexe que ce que le modèle appréhende. Face à de nouvelles prémisses, les participants auraient une confiance variable en la conclusion. Or, le modèle de Stenning et van Lambalgen ne rend pas compte des différents degrés par lesquels la suppression peut être observée.

Nous avons défendu que l'ajout de règles d'inférences floues à leur modèle permette de générer un système formel qui appréhende le phénomène de suppression dans toute sa complexité. En combinant la définition non monotone de l'implication, la règle générale d'inférence floue et la sémantique polyvalente de Kleene, nous avons présenté l'ébauche d'une modélisation de la suppression graduée des inférences. Nous avons soutenu que ce modèle peut combiner les avantages des approches bayésienne et non monotone, soit les représentations formelles de l'incertitude et des procédures non monotones de raisonnement. Comme Stenning et van Lambalgen, nous soutenons que la logique qui soutient ce modèle ne doit pas être présentée comme la norme de l'ensemble du raisonnement humain. En effet, suite à l'évaluation critique des trois théories du raisonnement effectuée dans les chapitres

précédents, nous avons observé une tendance progressive à définir la rationalité comme étant multiple et contextualisée. La norme *a priori* du raisonnement (la logique classique) a été remplacée par un modèle qui peut constituer une norme flexible *a posteriori*. Les logiques utilisées constituent des normes puisqu'au sein d'elle-même, une logique est normative¹. Par ailleurs, la norme doit être flexible puisqu'ultimement, elle n'est plus utilisée avec la prétention de normer la totalité du raisonnement. Le modèle qu'elle soutient est construit *a posteriori* puisqu'il est utilisé dans un contexte où on tient compte des observations empiriques. Pour reprendre la métaphore de la vision de l'esprit comme une *adaptive toolbox* (Gigerenzer & Selton, 2001), nos conclusions sont qu'une vision de la norme du raisonnement valide comme une *normative toolbox* pourrait être plus fertile et mieux adaptée au raisonnement humain que la recherche d'une norme universelle du raisonnement valide.

Sur le plan formel, nous avons vu trois manières de définir l'implication. Chacune d'entre elles est adaptée à la théorie explicative des auteurs qui l'utilisent. Byrne et Johnson-Laird utilisent l'implication matérielle classique pour construire les modèles mentaux à l'œuvre lors du raisonnement conditionnel, Oaksford et Chater utilisent la probabilité conditionnelle pour modéliser un raisonnement probabiliste et Stenning et van Lambalgen construisent leur propre définition pour introduire la non-monotonie dans une interprétation non monotone de l'implication. Toutefois, nous avons vu que ces derniers, appuyés par Politzer et Bonnefon, laissent le champ libre pour l'utilisation d'autres outils formels afin de modéliser différents types de raisonnements. Ils sont conscients de l'éventail d'outils logiques à leur disposition et en proposent l'utilisation pour des entreprises de modélisation futures. Un premier pas dans cette voie serait, selon nous, d'examiner les outils déjà à notre disposition ainsi que ce qui les distingue. Par exemple, à partir des opérateurs flous et de la

¹ Afin de modéliser le raisonnement, on utilise des systèmes formels comme la logique classique, des logiques polyvalentes, des logiques non monotones, du calcul bayésien des probabilités, etc. Dans tous les cas, ces systèmes sont autosuffisants et, de par leurs règles, ils déterminent la validité ou l'invalidité des manipulations qui y sont faites. Par exemple, prétendre qu'une phrase est à la fois vraie et fausse est une manipulation logiquement invalide en logique classique, mais peut être valide en logique polyvalente. C'est en ce sens qu'un système formel est normatif au sein de lui-même.

définition de l'implication matérielle, il est possible de construire 14 définitions de l'implication au sein desquelles l'opérateur se comporte de manière distincte.

Nous partons de l'idée que l'implication matérielle peut être représentée avec la disjonction ainsi que la conjonction. Par exemple, l'expression $p \rightarrow q$ est équivalente à l'expression $(\neg p \vee q)$, ce qui signifie que l'implication est vraie quand p est faux, quand q est vrai ou les deux. L'implication est aussi équivalente à l'expression $\neg(p \& \neg q)$, ce qui signifie que l'implication est fautive dans le seul cas où p est vrai et que q est faux. En logique floue, la conjonction et la disjonction ont chacune quatre définitions distinctes qui les font se comporter différemment. Les deux opérateurs peuvent être définis de façon standard, algébrique, bornée ou drastique. Les quatre définitions de la conjonction sont les suivantes :

$$\text{Intersection standard: } (p \& q) = \min(p, q)$$

$$\text{Produit algébrique: } (p \& q) = pq$$

$$\text{Différence bornée: } (p \& q) = \max(0, p + q - 1)$$

$$\text{Intersection drastique: } (p \& q) = \begin{cases} p \text{ quand } q = 1 \\ q \text{ quand } p = 1 \\ 0 \text{ autrement} \end{cases}$$

Les quatre définitions de la disjonction sont les suivantes :

$$\text{Union standard : } (p \vee q) = \max(p, q)$$

$$\text{Produit algébrique : } (p \vee q) = p + q - pq$$

$$\text{Union bornée : } (p \vee q) = \min(1, p + q)$$

$$\text{Union drastique : } (p \vee q) = \begin{cases} p \text{ quand } q = 0 \\ q \text{ quand } p = 0 \\ 1 \text{ autrement} \end{cases}$$

En combinant les différentes définitions de l'implication par l'utilisation des conjonction et disjonctions floues, on peut obtenir 16 combinaisons, exposées dans le tableau ci-dessous. De ces 16 combinaisons, 2 apparaissent à deux reprises. Ce tableau expose donc 14 définitions distinctes de l'implication.

Tableau 3.1

Inférences floues

	$p \rightarrow q \leftrightarrow \neg p \vee q$	$p \rightarrow q \leftrightarrow 1 \vee q$	$p \rightarrow q \leftrightarrow \neg p \vee (p \& q)$	$p \rightarrow q \leftrightarrow (\neg p \& \neg q) \vee q$
Intersection et union standard	$\max(1-p, q)$	$\begin{cases} 1 \text{ quand } p \leq q \\ q \text{ quand } p > q \end{cases}$	$\max[1-p, \min(p, q)]$	$\max[\min(1-p, 1-q), q]$
Intersection et union algébriques	$1-p+(p \times q)$	$\begin{cases} 1 \text{ quand } p \leq q \\ \frac{q}{p} \text{ quand } p > q \end{cases}$	$1-p+(p^2 \times q)$	$1-p+q+2pq+p^2-pq^2$
Intersection et union bornées	$\min(1, 1-p+q)$	$\min(1, 1-p+q)$	$\max(1-p, q)$	$\min(1, \max(0, (1-p)-(q-1)))+q)$
Intersection et union drastiques	$\begin{cases} q \text{ quand } p = 1 \\ 1-p \text{ quand } q = 0 \\ 1 \text{ autrement} \end{cases}$	$\begin{cases} q \text{ quand } p = 1 \\ 1 \text{ autrement} \end{cases}$	$\begin{cases} q \text{ quand } p = 1 \\ 1-p \text{ quand } p \neq 1, q \neq 1 \\ 1 \text{ quand } p \neq 1, q = 1 \end{cases}$	$\begin{cases} 1-p \text{ quand } q = 0 \\ q \text{ autrement} \end{cases}$

La définition de l'implication proposée par Stenning et van Lambalgen semble être adéquate pour modéliser le raisonnement non monotone. Leur entreprise montre que pour modéliser un phénomène précis, la simple définition de l'implication matérielle peut être caduque. Pour palier à ce problème, il est possible, et même nécessaire, de construire une définition de l'implication adaptée au phénomène qu'elle modélise. Or, une vaste littérature en psychologie du raisonnement expose des phénomènes tout aussi complexes que la suppression des inférences. En utilisant les outils de la logique floue, il devient possible de construire de nombreuses définitions distinctes de l'implication. Selon nous, une entreprise de modélisation de ces phénomènes pourrait profiter d'outils logiques aussi complexes et diversifiés que semble être le raisonnement humain. En ce sens, les définitions précédentes sont présentées comme l'ébauche d'un coffre à outil diversifié qui pourrait nourrir les entreprises ultérieures d'explication computationnelle du raisonnement humain.

BIBLIOGRAPHIE

- Barouillet, P. Gauffroy, C. & J. F. Lecas. 2008. « Mental models and the suppositional account of conditionals ». *Psychological Review*, vol 115, n° 3, p.760-771
- Beaulac, G. & S. Robert.2011. « Les théories de l'éducation à l'ère des sciences cognitives : le cas de l'enseignement de la pensée critique et de la logique » *Les ateliers de l'éthique numéro special : Le point sur la recherche québécoise sur la philosophie et les théories de l'éducation*. vol. 6, n° 1. p. 63-77.
- Bonnefon, J-F. & G. Politzer. 2011. « Pragmatics, mental models and one paradox of the material conditional » *Mind and Language*, vol. 26, n° 2. p. 141-155.
- Braine M. D .S. & D.P. O'Brien dir., 1998. *Mental Logic*. Mahwah. NJ: Erlbaum.
- Byrne, R. M. J. 1989. « Suppressing valid inferences with conditionals » *Cognition*, vol. 31. p. 61-83.
- Byrne R. M. J. 1991. « Can valid inferences be suppressed? » *Cognition*, vol. 39, n° 1, p. 71-78.
- Byrne, R. M. J. 2005. *The rational imagination: How people create alternatives to reality*. Cambridge: The MIT Press.
- Byrne, R. M. J., Espino, O. & C. Santamaria. 1999. « Counterexamples and the suppression of inferences ». *Journal of Memory & Language*, vol 40, p. 347-373.
- Cummins D. D., Lubart T, Alksnis O, Rist R.1991. « Conditional reasoning and causation ». *Memory Cognition*. vol. 19, n° 3, p. 274-282.
- Dieussaert, K., Schaeken, W., Schroyens, W. & G. D'Ydewalle. 2000. « Strategies during complex conditional inferences ». *Thinking and reasoning*, vol 6, n° 2. p. 125-160.
- Evans, J. S. B. T. 1977. « Linguistic factors in reasoning » *The Quarterly Journal of Experimental Psychology*, vol 29, n° 2, p. 297-306.
- Evans, J. S. B. T., S.E. Newstead & R.M.J. Byrne. 1993. *Human reasoning: The psychology of deduction*. Hove: Lawrence Erlbaum Associates.
- Evans, J. S. B. T. 1998. « Matching bias in conditional reasoning: Do we understand it after 25 years? ». *Thinking and Reasoning*, vol. 4, p. 45-82.b. New York: Oxford University Press.
- Evans, J. S. B. T. & D.E. Over. 2004. *If*. New York: Oxford University Press.

- Evans, J. S. B. T., Over, D. E. & S. J. Handley. 2005. « Suppositions, extensionality, and conditionals: A critique of the mental model theory of Johnson-Laird and Byrne (2002) ». *Psychological Review*, vol 112, no 4, p. 1040-1052.
- Gabbay, D., C. Hogger & J. Robinson. 1994. *Handbook of logic in artificial intelligence and logic programming*, vol. 3, Oxford et New York: Oxford University Press.
- Gentzen, G. 1964. « Investigations into logical deduction ». *American Philosophical Quarterly*, vol 1, n° 4, p. 288-306.
- Gigerenzer, G. 2007. *Gut feelings. The intelligence of the unconscious*. New York: Penguin Books.
- Gigerenzer, G. & Selton, R (2001). « Bounded rationality » in Gigerenzer, G. Selton R. *Bounded Rationality, The adaptive toolbox*, Cambridge, MA, MIT Press.
- Gigerenzer, G., P.M. Todd & ABC Research Group. 1999. *Simple heuristics that make us smart*. New York: Oxford University Press.
- Gilovich, T., D. Griffin & D. Kahneman, dir. 2002. *Heuristics and biases: The psychology of intuitive judgment*. New York: Cambridge University Press.
- Ginsberg, M. dir., 1987. *Readings in non monotonic reasoning*. Los Altos. CA: Morgan Kauffman.
- Hacking, I. 2001. *An introduction to probability and inductive logic*. New York: Cambridge University Press.
- Johnson Laird, P. 1975. « Models of deduction ». In Falmagne, R. (Ed.) *Reasoning: Representation and process*. Springdale: Erlbaum. p. 7-54.
- Johnson Laird, P. 1983. *Mental models*. Cambridge: Cambridge University Press.
- Johnson-Laird, P. 1999. « Deductive Reasoning ». *Annual Review of Psychology*, vol. 50, p. 109-135.
- Johnson-Laird, P. 2006. *How we reason*. New York: Oxford University Press.
- Johnson-Laird, P. & R. M. J. Byrne. 1991. *Deduction*. Hillsdale. N.J.: Lawrence Erlbaum Associates.
- Johnson-Laird, P. & Byrne, R.M.J. 2002. « Conditionals: A theory of meaning, pragmatics and inference », *Psychological Review*, 109, p. 646 – 678.

- Klir, J.G. & Bo Yuan. 1995. *Fuzzy sets and fuzzy logic: Theory and applications*. Upper Saddle River: Prentice Hall PTR.
- Kuhn, D. 1977. « Conditional reasoning in children ». *Developmental Psychology*, vol 13, n° 4, p. 342-353.
- Manktelow, K. & Man Cheung Chung. Dir. 2004. *Psychology of reasoning: Theoretical and historical perspectives*, New York: Psychology Press.
- Markman A.B. & D. Gentner. 2001. « Thinking ». *Annual Review of Psychology*. vol. 52, p. 223-247
- Markovits, H. & F. Potvin. 2001. « Suppression of valid inferences and knowledge structures: The curious effect of producing alternative antecedents on reasoning with causal conditionals ». *Memory and Cognition*, vol 29, n° 5, p. 736-744.
- Markovits, H., Lortie Forgues, H., & Brunet, M.-L. 2010. « Conditional reasoning, frequency of counterexamples, and the effect of response modality ». *Memory and Cognition*, 38(4), 485-492.
- Markovits, H. & H. Lortie Forgues. 2011. « Conditional reasoning under time constraint: information retrieval and inhibition ». *Thinking and Reasoning*, vol. 16, n° 3, p. 221-232.
- Markovits, H., Brunet, M.-L., Thompson, V., & Brisson, J. 2013. « Direct evidence for a dual process model of deductive inference ». *Journal of Experimental Psychology: LMC*.
- McCarthy, J.M. 1980. « Circumscription : a form of non-monotonic reasoning ». *Artificial Intelligence*, vol. 13, p. 27-39.
- McDermott, D. 1982. « Non-monotonic logic II : Non-monotonic model theories ». *Journal of the Association for Computing Machinery*, vol. 29, p. 33-57.
- McDermott, D. and Doyle, J. 1980. « Non-monotonic logic I ». *Artificial intelligence*, vol. 13 : p. 41-72.
- Miller, G.A. 1956. « The magical number seven, plus or minus two: Some limits on our capacity for processing information ». *Psychological Review*, vol. 63, n° 2, p. 81-97
- Oaksford, M. & N. Chater. 1991. « Against logicist cognitive science ». *Mind and Language*, vol. 6, p. 1-38.
- Oaksford, M. & N. Chater, N. 1996. « Rational explanation on the selection task ». *Psychological Review*, vol. 101, p. 108-131.
- Oaksford, M. & N. Chater. 2001. « The probabilistic approach to human reasoning ». *Trends in Cognitive Sciences*, vol. 5, p. 349-357.

- Oaksford, M. & N. Chater. 2003. « Probabilities and pragmatics in conditional inferences: Suppression and order effects ». In D. Hardman et L. Macchi, dir. *Thinking: psychological perspectives on reasoning, judgement and decision making* édité par John Wiley & Son, p. 95-122.
- Oaksford, M. & N. Chater. 2007. *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford. Oxford University Press.
- Oaksford, M. & N. Chater. 2009. « Precis of bayesian rationality », *Behavioral and Brain Science*, vol. 32, n° 1, p. 69-120.
- Oaksford, M. & N. Chater dir., 2010. *Cognition and conditionals: probability and logic in human thinking*. Oxford: Oxford University Press.
- Oaksford M. & N. Chater. 2011. « The « is-ought fallacy » fallacy ». *Behavioral and Brain Sciences*, vol 34, n° 5, p. 262-263
- Politzer, G. 2005. « Uncertainty and the suppression of inferences ». *Thinking and Reasoning*, vol. 11, n°1, p. 5-33.
- Politzer, G. 2007. « Reasoning with conditionals ». *Topoi*, vol. 26. p. 79-95.
- Politzer, G., & J. F. Bonnefon. 2006. « Two varieties of conditionals and two kinds of defeaters help reveal two fundamental types of reasoning ». *Mind & Language*, vol 21, p. 484-503.
- Popper, K. 2002. *The logic of scientific discovery*. New York : Routledge. 513 p.
- Reiter, R. 1980. « A logic for default reasoning », *Artificial Intelligence*, vol. 13, p. 81-132.
- Reiter, R. 1985. « On reasoning by default », In R. Brachan and H. Levesque (dir.), *Readings in knowledge representation*. Los Altos, CA: Morgan Kaufman, p. 401-410.
- Rescher, N. 1969. *Many-valued logic*. New York: McGraw Hill.
- Rescher, N. 2005. *Epistemic logic*. Pittsburgh: University of Pittsburgh Press.
- Rips, L. J. 1983. « Cognitive processes in propositional reasoning ». *Psychological Review*, vol 90, n° 1, p. 38-71
- Rips, L.J. 1990. « Reasoning ». *Annual Review of Psychology*. vol. 41, p. 321-353.
- Rips L.J. 1994. *The psychology of proof: deductive reasoning in human thinking*. Cambridge: MIT Press.

- Robert, S. « Logique de la découverte et naturalisation de la connaissance : L'épistémologie historique d'Imre Lakatos » In *Philosophie de la Connaissance*. R. Nadeau dir, Presses de l'Université Laval, Vrin, 2009.
- Shafir, E. & R.A. LeBoeuf. 2002. « Rationality ». *Annual Review of Psychology*. vol. 53, p. 191-517.
- Stalnaker, R. 1994. « Non-monotonic consequence relations », *Fundamenta Informaticae*, vol. 21, p. 7-21.
- Stanovich, K.E. 2004. *The robot's rebellion: Finding meaning in the age of Darwin*. Chicago: University of Chicago Press.
- Stanovich KE & R.F. West. 1998a. « Cognitive ability and variation in selection task performance ». *Thinking and Reasoning*, vol. 4, p. 193-230
- Stanovich KE & R.F. West. 1998b. « Individual differences in framing and conjunction effects ». *Thinking and Reasoning*, vol. 4, p. 289-317
- Stanovich KE & R.F. West. 2000. « Individual differences in reasoning: implications for the rationality debate ». *Behavioral and Brain Science*, vol. 23, p. 645-726.
- Stanovich, K.E. & R.F. West. 2008. « On the relative independence of thinking biases and cognitive ability ». *Journal of Personality and Social Psychology*, vol. 94, n° 4, p. 672-695.
- Stenning, K. & M. van Lambalgen. « The logical response to a noisy world » in Oaksford, M. et N. Chater. Ed. 2010. *Cognition and Conditionals: Probability and Logic in Human Thinking*. Oxford: Oxford University Press.
- Stenning, K. & M. van Lambalgen. 2008. *Human reasoning and cognitive science*. Cambridge (MA): MIT Press.
- Stevenson, R. & D.E. Over, D.E. 1995. « Deduction from uncertain premises ». *Quarterly Journal of Experimental Psychology*, vol. 48A, p. 613-643.
- Stevenson R. J. & D. E. Over. 2001. « Reasoning from uncertain premises: Effects of expertise and conversational context ». *Thinking and Reasoning*, vol. 7, n° 4, p. 367-390.
- Tooby, J. & L. Cosmides. 2005. « Conceptual foundations of evolutionary psychology ». *The Handbook of Evolutionary Psychology*, D. Buss dir., Hoboken, NJ: Wiley, p. 5-67.
- Wason, P. C. 1966. « Reasoning ». In B. M. Foss dir., *New horizons in psychology I*. Harmondsworth: Penguin.
- Wason, P. C. 1968. « Reasoning about a rule ». *Quarterly Journal of Experimental Psychology*, vol. 20, p. 273-281.

- Wason, P. & P. Johnson-Laird. 1972. *Psychology of reasoning*. London: Batsford.
- Wason, P.C. et J.St.B.T. Evans. 1975. «Dual processes in reasoning? ». *Cognition*, vol. 3, p. 141-154.
- Wittgenstein, L. 1993. *Tractatus logico-philosophicus*. Paris : Gallimard. 121 p.
- Zadeh, L. 1965. «Fuzzy Sets», *Information and control*, vol. 8, p. 338-353.
- Zadeh, L.A., K.S. Fu, K. Tanaka & M. Shimura. 1975. *Fuzzy sets and their applications to cognitive and decision processes*. New York: Academic Press.
- Zadeh, L.A. & J. Kacprzyk. 1992. *Fuzzy logic for the management of uncertainty*. New York: John Wiley.