

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

SI, ALORS ET PLUS ENCORE: IDENTIFICATION DES COMPÉTENCES
VISÉES ET ÉVALUATION AUTOMATIQUE DU RAISONNEMENT
DÉDUCTIF CHEZ LES ADULTES

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN INFORMATIQUE COGNITIVE

PAR

JANIE BRISSON

OCTOBRE 2019

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de cette thèse 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.10-2015). 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.»

« Doute de tout, même de moi. »
Papy, tu ne t'attendais sûrement pas à ce que
ces mots engendrent une myriade de réflexions,
dont ce travail est l'aboutissement.

TABLE DES MATIÈRES

LISTE DES TABLEAUX.....	VII
LISTE DES FIGURES	IX
LISTE DES ABBRÉVIATIONS	X
RÉSUMÉ	XI
INTRODUCTION	1
0.1 LE PROJET MUSE-LOGIQUE.....	1
0.2 OBJECTIFS DE LA THÈSE	2
0.3 STRUCTURE DE LA THÈSE	3
CHAPITRE I	
CADRE THÉORIQUE.....	5
1.1. LOGIQUE CLASSIQUE ET RAISONNEMENT DÉDUCTIF	5
1.2. EFFETS DU CONTENU : CONTRE-EXEMPLES ET PERSPECTIVE ADOPTÉE	11
1.2.1 <i>Le raisonnement conditionnel chez les humains.....</i>	<i>12</i>
1.2.2 <i>Effets de contenu et débat sur la nature du raisonnement humain.....</i>	<i>12</i>
1.2.3 <i>Perspective adoptée : Le cadre de la mémoire sémantique.....</i>	<i>18</i>
1.3. STRUCTURES LOGIQUES ET COMPÉTENCES VISÉES	19
1.4. UN RÉSEAU BAYÉSIEN COMME OUTIL D'ÉVALUATION AUTOMATIQUE DU RAISONNEMENT.....	21
1.5. EFFETS DE LA FORME : INTUITION LOGIQUE ET COMPÉTENCES VISÉES	22
1.6. LOGIQUE CLASSIQUE ET NORME DU RAISONNEMENT VALIDE : CLARIFICATIONS	24
CHAPITRE II	
UN RÉSEAU BAYÉSIEN COMME MODÈLE DE L'APPRENANT.....	25
2.1 CALCUL DES PROBABILITÉS	25
2.1.1 <i>Loi de Bayes.....</i>	<i>25</i>

2.1.2	<i>Calcul des probabilités postérieures</i>	27
2.1.3	<i>Inférences possibles avec la théorie des probabilités conditionnelles</i> ...	29
2.2	LES RÉSEAUX BAYÉSIENS	30
2.2.1	<i>Muse : Un réseau bayésien à validation experte</i>	32
2.3	ABSTRACT	40
2.4	INTRODUCTION	41
2.5	LOGIC-MUSE'S LEARNER MODEL	41
2.6	ITEM BANK FOR CONDITIONAL REASONING	42
2.7	VALIDATION AND RB INITIALIZATION.	44
2.7.1	<i>Data collection</i>	45
2.7.2	<i>Results</i>	46
2.7.3	<i>BN initialization through the CDM model</i>	47
2.7.4	<i>Validation for BN and underlying psychological model</i>	49
2.8	CONCLUSION	52
2.9	TRAVAUX FUTURS	53
2.9.1	<i>Apprentissage des distributions de probabilités : Nœuds observables</i> ...	53
2.9.2	<i>Apprentissage des distributions de probabilité : Nœuds cachés</i>	55
2.9.3	<i>Apprentissage des distributions de probabilité : Muse-Logique</i>	56
2.10	CONCLUSION	56
CHAPITRE III		
LE RAISONNEMENT À PARTIR D'UNE INCOMPATIBILITÉ		57
3.1	ABSTRACT	59
3.2	INTRODUCTION	60
3.3	DEDUCTIVE REASONING FROM AN INCOMPATIBILITY STATEMENT	62
3.4	CONTENT EFFECTS AND CONDITIONAL REASONING	63
3.5	STUDY 1	65
3.5.1	<i>Pretest</i>	66
3.5.2	<i>Method</i>	68
3.5.3	<i>Results and Discussion</i>	70

3.6	STUDY 2A.....	72
3.6.1	<i>Pretest</i>	73
3.6.2	<i>Method</i>	75
3.6.3	<i>Results and discussion</i>	77
3.7	STUDY 2B.....	80
3.7.1	<i>Method</i>	81
3.7.2	<i>Results and discussion</i>	83
3.8	GENERAL DISCUSSION.....	85
3.9	TRAVAUX FUTURS : EXTENSION DU RB AU RAISONNEMENT À PARTIR D'UNE INCOMPATIBILITÉ.....	89
3.10	TRAVAUX FUTURS : EXTENSION DU RB AU RAISONNEMENT À PARTIR D'UNE DISJONCTION INCLUSIVE.....	91
3.11	CONCLUSION.....	93
CHAPITRE IV		
L'INTUITION LOGIQUE.....		
4.1	ABSTRACT.....	96
4.2	INTRODUCTION.....	96
4.3	EVIDENCE FOR LOGICAL INTUITION: THE "LOGICAL BIAS".....	98
4.4	ADDITIONAL SUPPORTIVE FINDING: THE TWO-RESPONSE PARADIGM.....	99
4.5	CRITIQUES AND BOUNDARY CONDITIONS OF LOGICAL INTUITION.....	99
4.6	PRETEST.....	102
4.7	STUDY 1.....	104
4.7.1	<i>Method</i>	104
4.7.2	<i>Results and discussion</i>	105
4.8	STUDY 2.....	108
4.8.1	<i>Method</i>	108
4.8.2	<i>Results and discussion</i>	108
4.9	GENERAL DISCUSSION.....	114
4.10	INTUITION LOGIQUE ET MODÈLE DE L'APPRENANT.....	117
CONCLUSION.....		119

APPENDICE A	121
APPENDICE B	122
APPENDICE C	123
APPENDICE D	127
APPENDICE E	128
APPENDICE F.....	129
APPENDICE G	133
BIBLIOGRAPHIE.....	135

LISTE DES TABLEAUX

TABLEAU 1.1. TABLE DE VÉRITÉ DE L'OPÉRATEUR D'IMPLICATION	6
TABLEAU 1.2. QUATRE PRÉMISSES MINEURES POSSIBLES	6
TABLEAU 1.3. QUATRE INFÉRENCES À PARTIR D'UNE IMPLICATION	8
TABLEAU 1.4. TABLES DE VÉRITÉ DE LA DISJONCTION INCLUSIVE ET DE L'INCOMPATIBILITÉ	9
TABLEAU 1.5. QUATRE INFÉRENCES DISJONCTIVES	10
TABLEAU 1.6. QUATRE INFÉRENCES À PARTIR D'UNE INCOMPATIBILITÉ	11
TABLEAU 2.1. DISTRIBUTION DE PROBABILITÉS (NEGNEVITSKY, 2005)	28
TABLEAU 2.2. LÉGENDE POUR LES NŒUDS ITEMS FAMILIERS.....	35
TABLEAU 2.3. LÉGENDE POUR ITEMS CAUSAUX CONTREFACTUELS	37
TABLEAU 2.4. LÉGENDE POUR ITEMS ABSTRAITS.....	38
TABLE 2.5. SKILL CLASS CHUNK PROBABILITIES.....	48
TABLE 2.6. GUESS, SLIP AND IDI PARAMETERS	49
TABLE 2.7. SKILL CORRELATIONS.....	50
TABLE 2.8. SKILL DISTRIBUTION	51
TABLEAU 2.9. JEU DE DONNÉES : LANCERS DE PIÈCES DE MONNAIE	54
TABLEAU 2.10. DISTRIBUTION DE PROBABILITÉS : VARIABLE LANCER	55
TABLE 3.1. MEAN PERCENTAGE OF LOGICALLY CORRECT RESPONSES FOR THE FOUR LOGICAL FORMS (DENY FIRST, DENY SECOND, AFFIRM FIRST, AFFIRM SECOND) BY TYPE (MANY, FEW) IN STUDY 1 (STANDARD DEVIATIONS IN PARENTHESES)..	71
TABLE 3.2. MEAN PERCENTAGE OF LOGICALLY CORRECT RESPONSES FOR THE FOUR LOGICAL FORMS (DENY FIRST, DENY SECOND, AFFIRM FIRST, AFFIRM SECOND) BY TYPE (BROAD, REDUCED, CLOSE TO BINARY) IN STUDY 2 (STANDARD DEVIATIONS IN PARENTHESES)	79
TABLE 3.3. INTERPRETATIONS OF STATEMENTS FOR EACH PREMISE TYPE (BROAD, REDUCED, CLOSE TO BINARY) AS A FUNCTION OF FORMULATION (INCOMPATIBLE, CONJUNCTIVE).....	84

TABLE 4.1. PERCENTAGE OF LOGICALLY CORRECT RESPONSES ON OVERALL, CONFLICT AND NO-CONFLICT PROBLEMS FOR STUDY 1, STUDY 2, AND POOLED DATA (STANDARD DEVIATIONS IN PARENTHESES).....	111
TABLE 4.2. RESPONSE LATENCY (IN SECONDS) FOR INCORRECT CONFLICT AND CORRECT NO-CONFLICT PROBLEMS (STANDARD DEVIATIONS IN PARENTHESES).....	111

LISTE DES FIGURES

FIGURE 1.1 GROUPE DE KLEIN POUR L'IMPLICATION ET L'IMPLICATION INVERSE	20
FIGURE 1.2 GROUPE DE KLEIN POUR LA DISJONCTION INCLUSIVE ET L'INCOMPATIBILITÉ	21
FIGURE 2.1 RÉSEAU BAYÉSIEN ET DISTRIBUTION DE PROBABILITÉS (MITCHELL, 1997) 32	
FIGURE 2.2 TÊTE DU RB POUR LE RAISONNEMENT CONDITIONNEL.....	33
FIGURE 2.3 PORTION DU RB POUR LE RAISONNEMENT FAMILIER	34
FIGURE 2.4 PORTION DU RB POUR LE RAISONNEMENT CAUSAL CONTREFACTUEL	37
FIGURE 2.5 PORTION DU RB POUR LE RAISONNEMENT ABSTRAIT	38
FIGURE 2.6 RÉSEAU BAYÉSIEN ET DISTRIBUTION DE PROBABILITÉS : PILE OU FACE ...	54
FIGURE 3.1 TÊTE PROJÉTÉE DU RB POUR LE RAISONNEMENT AVEC INCOMPATIBILITÉ	89
FIGURE 3.2 PORTION PROJÉTÉE DU RB POUR LE RAISONNEMENT AVEC INCOMPATIBILITÉ, CONTENU FAMILIER	90
FIGURE 3.3 PORTION PROJÉTÉE DU RB POUR LE RAISONNEMENT AVEC INCOMPATIBILITÉ, CONTENU ABSTRAIT	91
FIGURE 3.4 TÊTE PROJÉTÉE DU RB POUR LE RAISONNEMENT DISJONCTIF	91
FIGURE 3.5 PORTION PROJÉTÉE DU RB POUR LE RAISONNEMENT DISJONCTIF, CONTENU FAMILIER.....	92
FIGURE 3.6 PORTION PROJÉTÉE DU RB POUR LE RAISONNEMENT DISJONCTIF, CONTENU ABSTRAIT	93
FIGURE 4.1 PORTION DU RB PORTANT SUR L'INTUITION LOGIQUE	118

LISTE DES ABBRÉVIATIONS

RB (BN)	Réseau bayésien (Bayesian network)
STI (ITS)	Système tutoriel intelligent (Intelligent tutoring system)
MPP	Modus Ponendo Ponens
MTT	Modus Tollendo Tollens
AC	Affirmation du Aonséquent
DA	Déni de l'Antécédent
Affirm First	Affirmation of the First clause
Affirm Second	Affirmation of the Second clause
Deny First	Denial of the First clause
Deny Second	Denial of the Second clause
FPCC (FFD)	Familier Peu de Conditions Contraignantes (Familiar Few Disablers)
FBCC (FMD)	Familier Beaucoup de Conditions Contraignantes (Familiar Many Disablers)
FBA (FMA)	Familier Beaucoup d'Alternatives (Familiar Many Alternatives)
FPA (FFA)	Familier Peu d'Alternatives (Familiar Few Alternatives)
CCF	Causal Contrefactuel (Causal counterfactual)
A	Abstrait (Abstract)
IDI	Indice de Discrimination de l'Item (Item Discrimination Index)

RÉSUMÉ

Plusieurs études démontrent la difficulté que les humains éprouvent à raisonner logiquement à partir d'une prémisse conditionnelle (si antécédent (P), alors conséquent (Q)) et les performances observées varient en fonction du contenu de la prémisse. Markovits & Barouillet (2002) expliquent ces effets du contenu par la prégnance des contre-exemples récupérables à partir de la mémoire à long terme. Selon ce cadre théorique, le défi principal du raisonnement conditionnel est de reconnaître la certitude des inférences valides ainsi que l'incertitude des inférences invalides par la génération de contre-exemples. Combiné à ces travaux, un examen formel des opérateurs logiques permet de faire des prédictions sur la manière dont les humains raisonneront à partir d'autres connecteurs. Selon l'analyse de Robert et Brisson (2016), générer des contre-exemples aux inférences invalides équivaut, d'un point de vue formel, à générer des instances du dual de l'opérateur de la prémisse majeure. Cette analyse suggère que l'implication, la disjonction inclusive et l'incompatibilité, étant les trois opérateurs classiques qui ont un dual distinct, représentent le défi principal en matière de raisonnement logique. Dans cette thèse, nous présenterons un outil de diagnostic cognitif du raisonnement conditionnel à partir ces deux approches. Cet outil jouera le rôle d'un modèle de l'apprenant au sein d'un Système Tutoriel Intelligent (STI) pour l'apprentissage de la logique. Les Réseaux Bayésiens (RB) offrent un formalisme tout à fait adéquat pour cette tâche de modélisation: ils permettent d'inférer la probabilité de maîtrise d'un ensemble de compétences étant donné un patron de réponse spécifique (Conati, 2010; Conati & al., 2002). Nous présenterons la conception et la validation d'un RB en tant qu'outil d'évaluation automatique du raisonnement conditionnel. Ce modèle étant conçu de manière à ce qu'il soit extensible, nous projeterons les branches du RB pouvant prendre en compte le raisonnement à partir d'une disjonction inclusive et d'une incompatibilité. Nous présenterons aussi une étude expérimentale sur les effets du contenu au sein du raisonnement à partir d'une incompatibilité. La littérature en psychologie du raisonnement ne s'étant pas encore penchée sur cette question, notre étude sert d'appui empirique essentiel à notre outil d'évaluation. Finalement, nous aborderons un phénomène récemment observé par la psychologie du raisonnement, soit la capacité de développer une compréhension intuitive de certains principes logiques. Nous présenterons une étude qui se penche sur la portée et les limites d'un tel phénomène dans le raisonnement conditionnel. Une prise en compte de ce phénomène par notre outil d'évaluation automatique sera discutée.

***Mots-clés :** Raisonnement, Logique, Effets de contenu, Systèmes tutoriels intelligents, Réseaux bayésiens, Intuition logique.*

INTRODUCTION

0.1 Le projet Muse-Logique

Les systèmes tutoriels intelligents (STI) sont des logiciels offrant un environnement d'apprentissage virtuel qui permet de soutenir des activités d'apprentissage sans impliquer un tuteur humain. Le projet Muse-Logique vise la création d'un STI pour l'apprentissage du raisonnement logique. Les objectifs de ce projet sont d'étudier les bases de l'acquisition des compétences en matière de raisonnement logique, de comprendre les difficultés liées à cet apprentissage et de créer un STI capable de détecter, de diagnostiquer et de corriger les erreurs du raisonnement humain dans diverses situations. Les Systèmes Tutoriels Intelligents ont fait avancer la compréhension de l'apprentissage en général, mais ont aussi fait leur preuve quant à leur capacité de soutenir l'apprentissage du raisonnement en sciences, mathématiques, et même en logique (Barnes & Stamper, 2010; Lesta & Yacef, 2002; Tchetagni, Nkambou & Bourdeau, 2007). Cependant, nous pouvons formuler deux limites aux STI proposés pour l'apprentissage de la logique : 1) Ces systèmes ne se fondent pas sur une élicitation des connaissances et structures logiques ; 2) Ils ne se fondent pas sur des théories explicatives du raisonnement humain.

Le projet Muse-logique a permis de combler ces lacunes en rassemblant les développements récents dans les domaines des Systèmes Tutoriels Intelligents, de la psychologie du raisonnement et de l'Intelligence Artificielle pour proposer un environnement d'apprentissage capable d'aider les apprenants à améliorer leurs compétences en matière de raisonnement. Ce STI permet d'adapter et de personnaliser les activités d'apprentissage en fonction des besoins de l'apprenant. Son architecture ressemble à celle d'un STI classique, comprenant les trois composantes habituelles: le module de l'expert, le module de l'apprenant ainsi que le module du tuteur. L'approche participative adoptée au sein de ce projet a mené à l'élaboration de composantes essentielles pour la construction du STI. Ces composantes incluent un catalogue des erreurs, une banque d'items et des stratégies pédagogiques. Muse-Logique, dans sa version

actuelle, implémente ces composantes pour la logique classique des propositions et offre une panoplie d'activités permettant à un apprenant de développer sa compétence en raisonnement logique dans plusieurs classes de situations. Ce projet a déjà mené à plusieurs publications scientifiques décrivant la conception et la mise en œuvre des trois composantes principales du STI (apprenant, tuteur, expert), développées et validées en s'appuyant sur l'aide d'experts et sur la littérature dans les domaines du raisonnement et de l'informatique (Nkambou, Brisson, Kenfack, Robert, Kissok & Tato, 2015 ; Nkambou, Kenfack, Robert & Brisson, 2015 ; Kenfack, Nkambou, Robert, Tato, Brisson & Kissok, 2016). De plus, des premières évaluations empiriques des composantes expert et apprenant ont été effectuées et publiées (Nkambou, Tato, Brisson, Kenfack, Robert & Kissok, 2016 ; Tato, Nkambou, Brisson, Kenfack, Robert & Kissok, 2016; Tato, Nkambou, Brisson & Robert, 2017).

0.2 Objectifs de la thèse

La présente thèse vise à présenter et valider la conception d'un outil de diagnostic automatique du raisonnement déductif dans le cadre de la logique classique à la lumière des connaissances sur les processus de raisonnement humain en psychologie cognitive. Cet outil jouera le rôle d'un modèle de l'apprenant au sein du système Muse-Logique. Dans notre approche, la récupération d'information pertinente ainsi que l'automatisation de certains principes logiques de base sont au cœur des compétences ciblées pour le raisonnement déductif.

L'objectif de notre problématique cognitive est double. D'une part, nous allons déterminer les compétences visées dans le cadre de l'apprentissage du raisonnement déductif en logique classique. D'autre part, nous allons explorer davantage la sensibilité des humains au contenu et à la forme logique dans le cadre du raisonnement déductif par le biais d'études expérimentales. Ces études jetteront les bases empiriques nécessaires au prolongement du modèle de l'apprenant de Muse-Logique.

Notre problématique informatique vise l'évaluation automatique des compétences en raisonnement, et ce, en tenant compte de la littérature sur les processus de raisonnement

humain. L'objectif est donc de concevoir un programme efficace pouvant tenir lieu de modèle de l'apprenant dans le STI Muse-Logique.

0.3 Structure de la thèse

Au premier chapitre, nous présenterons le contexte théorique nécessaire à l'élaboration de notre outil d'évaluation automatique. Nous présenterons le raisonnement déductif en logique classique, les phénomènes largement observés par les psychologues du raisonnement ainsi que la perspective adoptée au sein de cette thèse. Nous présenterons aussi une analyse des phénomènes recensés via les structures de la logique classique, ce qui nous permettra d'identifier les compétences visées pour Muse-Logique.

Au second chapitre, nous présenterons une structure à validation experte d'un Réseau Bayésien (RB) en tant qu'évaluateur automatique du raisonnement conditionnel. Nous présenterons également un *Cognitive Diagnosis Model* (CDM) basé sur une collecte de données empirique. Ce dernier permettra l'initialisation ainsi qu'une validation empirique du RB. Finalement, nous projeterons nos travaux futurs concernant notre problématique informatique : l'apprentissage automatique des distributions de probabilité du RB pour le raisonnement conditionnel.

Au troisième chapitre, nous présenterons nos travaux effectués au niveau de l'identification des compétences logiques et de l'expérimentation sur les processus de raisonnement humain à partir du connecteur d'incompatibilité. À partir du cadre de la récupération sémantique ainsi que de l'analyse formelle via les structures logiques, nous proposerons d'étendre les observations sur le raisonnement conditionnel au raisonnement à partir d'une incompatibilité. Nous avons conduit des expérimentations afin de tester cette prédiction et nous en ferons rapport dans ce chapitre. À notre connaissance, ces études sont les premières dans la littérature à se pencher sur ce type de raisonnement. Nous projeterons finalement nos travaux futurs au sein du RB: une structure à validation experte pour le raisonnement à partir d'une incompatibilité et d'une disjonction inclusive.

Enfin, au quatrième chapitre, nous présenterons nos travaux sur l'intuition logique dans le raisonnement conditionnel ainsi qu'une possible intégration de ce phénomène dans le modèle de l'apprenant de Muse-Logique

CHAPITRE I

CADRE THÉORIQUE

1.1. Logique classique et raisonnement déductif

L'inférence déductive, c'est-à-dire la capacité de tirer des conclusions nécessaires à partir de prémisses vraies, est un élément clé de la cognition humaine avancée. Pour être logiquement valide, le raisonnement doit satisfaire trois critères. D'abord, les prémisses doivent être considérées comme vraies, et ce, indépendamment de leur statut empirique. Ainsi, un raisonnement logique peut être fait à partir de prémisses fausses, c'est-à-dire non conformes à la réalité, ou imaginaires. En ce sens, une affirmation non conforme à la réalité, telle que « les éléphants sont roses », doit être considérée comme vraie. Ensuite, une conclusion logiquement valide ne fournit pas plus d'information que les prémisses peuvent en fournir. Autrement dit, dans une conclusion logiquement valide, on explicite l'information déjà présente implicitement dans les prémisses. Finalement, à partir de prémisses considérées comme vraies, le raisonnement déductif permet de générer une conclusion qui découle nécessairement. Lorsque plusieurs conclusions sont possibles, la valeur de vérité de chacune d'entre elles est incertaine. Ainsi, la capacité de distinguer entre des conclusions possibles et nécessaires est une habileté cruciale du raisonnement déductif. En ce sens, plusieurs erreurs logiques, appelées des sophismes, consistent à être certain à propos d'une conclusion qui est possible, mais non nécessaire.

Un opérateur (ou connecteur) de la logique classique peut être perçu comme un foncteur de vérité ou, autrement dit, un convertisseur de valeur de vérité. Une manière de définir un tel opérateur est en faisant sa table de vérité, c'est-à-dire en dressant une liste exhaustive des valeurs de vérité résultantes pour toute combinaison des valeurs de vérité des propositions qu'il connecte. Ainsi, une table de vérité d'un opérateur binaire qui connecte deux propositions (P et Q) aura quatre rangées: une pour chaque combinaison possible du

vrai et du faux pour P et Q. Par exemple, l'opérateur d'implication n'est faux que dans une seule rangée : lorsque P est vrai et que Q est faux. Autrement dit, lorsqu'un conséquent (Q) est faux même quand l'antécédent (P) est vrai, on ne peut dire que l'antécédent implique le conséquent. Les trois autres situations possibles (P vrai, Q vrai ; P faux, Q vrai ; P faux, Q faux) sont conformes à l'implication entre P et Q.

Tableau 1.1 Table de vérité de l'opérateur d'implication

	P	Q	(P \supset Q)
1.	V	V	V
2.	V	F	F
3.	F	V	V
4.	F	V	V

Le raisonnement conditionnel que l'implication engendre se présente sous forme d'inférences faites à partir d'une prémisses majeure (P \supset Q) et d'une des quatre prémisses mineures possibles, soit l'affirmation ou la négation de l'antécédent ou du conséquent de l'implication : « P est vrai », « P est faux », « Q est vrai » ou « Q est faux ». Prenons par exemple la prémisses majeure « Si un chien a des puces, alors il se gratte constamment ». Les quatre prémisses mineures possibles sont exposées dans le Tableau 1.2:

Tableau 1.2 Quatre prémisses mineures possibles

P est vrai	Un chien a des puces
P est faux	Un chien n'a pas de puces
Q est vrai	Un chien se gratte constamment
Q est faux	Un chien ne se gratte pas constamment

En combinant la prémisse majeure avec chacune des quatre prémisses mineures possibles, quatre inférences peuvent être effectuées : le *Modus Ponendo Ponens* (*MPP*), le *Modus Tollendo Tollens* (*MTT*), l'affirmation du conséquent (*AC*) et la négation de l'antécédent (*DA*). De ces quatre inférences possibles, seulement deux mènent à des conclusions logiquement valides. Le *MPP* contient les prémisses « P implique Q, P est vrai », et permet la conclusion logiquement valide « Q est vrai ». Par exemple, le raisonnement suivant est un *MPP* : « Si un chien a des puces, alors il se gratte constamment, Fido a des puces, donc Fido se gratte constamment ». Le *MTT* contient les prémisses « P implique Q, Q est faux » et permet la conclusion logiquement valide « P est faux ». Par exemple, le raisonnement « Si un chien a des puces, alors il se gratte constamment, Fido ne se gratte pas constamment, donc Fido n'a pas de puces » est un exemple de *MTT*. Les deux autres formes d'inférence conditionnelle mènent à des conclusions possibles, mais non nécessaires. Elles ne sont donc pas logiquement valides. L'Affirmation du Conséquent (*AC*) contient les prémisses « P implique Q, Q est vrai », et mène à la conclusion invitée « P est vrai ». Un raisonnement de forme *AC* serait, par exemple, « Si un chien a des puces, alors il se gratte constamment. Fido se gratte constamment, donc il a des puces ». Cette conclusion est incertaine puisque les prémisses permettent une possibilité où le conséquent est vrai même si l'antécédent est faux (Par exemple, Fido pourrait se gratter constamment parce qu'il a une maladie de la peau). Le déni de l'antécédent (*DA*) contient les prémisses « P implique Q, P est faux » et la conclusion « Q est faux ». L'inférence « Si un chien a des puces, alors il se gratte constamment. Fido n'a pas de puces, donc il ne se gratte pas constamment » est un exemple de *DA*. Encore une fois, les prémisses n'éliminent pas la possibilité non-P et Q, où Fido se gratte constamment étant donné une maladie de peau, de la peau sèche, etc. Les inférences *AC* et *DA* sont donc invalides puisque les prémisses laissent place à d'autres conclusions possibles, qui sont des contre-exemples à la conclusion invitée.

Les quatre inférences présentées ci-haut sont résumées dans le Tableau 1.3. Nous y présentons, pour chacune des inférences, la prémisse majeure, la prémisse mineure, la conclusion valide ainsi que la conclusion invitée pour les deux inférences invalides. Nous nous référons aussi aux rangées vraies du connecteur d'implication dans lesquelles la prémisse mineure se trouve. Cette référence à la table de vérité expose les raisons de la certitude (ou de l'incertitude) de la conclusion. Par un examen du Tableau 1.3, on observe

que dans le cas des inférences valides, une seule rangée vraie contient la prémisse mineure, ce qui garantit la certitude de la conclusion. Par exemple, dans le cas du *MPP*, l'affirmation de l'antécédent (P) est contenue dans une seule des rangées vraies de la table (lorsque P et Q sont vrais), si bien que Q est la seule conclusion possible sachant que P. Cependant, dans le cas du AC, deux rangées vraies (lorsque P est vrai et Q est vrai ainsi que lorsque P est faux et Q est vrai) contiennent l'affirmation du conséquent (Q) si bien que sur la base de ce dernier, il est impossible de conclure avec certitude à propos de P.

Tableau 1.3 Quatre inférences à partir d'une implication

	<i>Modus Ponendo Ponens (MPP)</i>	<i>Modus Tollendo Tollens (MTT)</i>	Affirmation du conséquent (AC)	Déni l'antécédent (DA)
Majeure	$P \supset Q$	$P \supset Q$	$P \supset Q$	$P \supset Q$
Mineure	P	$\sim Q$	Q	$\sim P$
Conclusion	Valide: Q	Valide: $\sim P$	Invitée: P Valide: Incertitude	Invitée: $\sim Q$ Valide: Incertitude
Rangée(s) vraie(s)	(P & Q)	($\sim P$ & $\sim Q$)	(P & Q) ($\sim P$ & Q)	($\sim P$ & $\sim Q$) ($\sim P$ & Q)

Le fait que l'opérateur d'implication ait trois rangées vraies dans sa table de vérité a des conséquences importantes sur le raisonnement déductif qu'il peut engendrer. En effet, lorsqu'un opérateur n'a que deux rangées vraies dans sa table de vérité (comme la disjonction exclusive ou l'équivalence) les quatre inférences qu'il génère seront valides, puisqu'elles mènent toutes à une seule conclusion possible. Or, lorsque trois rangées sont vraies, seulement deux inférences mèneront à une conclusion nécessaire et les deux autres mèneront à une conclusion possible. En logique classique, quatre opérateurs sont dans cette

situation : l'implication (et l'implication inverse¹), la disjonction inclusive et l'incompatibilité. Une disjonction inclusive est fausse lorsque P et Q sont faux et vraie autrement, alors qu'une incompatibilité est fausse lorsque P et Q sont vrais et vraie autrement. Le Tableau 1.4 expose les tables de vérité complètes de ces deux opérateurs.

Tableau 1.4 Tables de vérité de la disjonction inclusive et de l'incompatibilité

	P	Q	$(P \vee Q)$	$(P Q)$
1.	V	V	V	F
2.	V	F	V	V
3.	F	V	V	V
4.	F	F	F	V

Le raisonnement déductif à partir d'une disjonction inclusive consiste à inférer à partir d'une prémisse majeure disjonctive, combinée avec l'une des quatre prémisses mineures possibles, i.e. l'affirmation ou la négation de l'une des propositions. Les deux formes valides, les *Modus Tollendo Ponens* (*MTP*), nient l'une des propositions dans la prémisse mineure. Un premier *MTP* consiste à raisonner à partir de la prémisse "P ou Q, P est faux" et mène à la conclusion valide, "Q est vrai". (e.g. "Jean joue du violon ou du piano, Jean ne joue pas de violon. Donc, Jean joue du piano ». À partir des prémisses "P ou Q, Q est faux" le *MTP* mène à la conclusion valide "P est vrai". Par exemple, ce serait un raisonnement de la forme "Jean joue du violon ou du piano, Jean ne joue pas de piano. Donc, Jean joue du violon. » Les deux formes invalides nient l'une des propositions de la prémisse majeure. Nous les appellerons les *Modus Ponendo Tollens* (*MPT*). Un premier *MPT* consiste à raisonner à partir des prémisses "P ou Q, P est vrai" et mène à la conclusion

¹ Notons que nous ne nous attarderons pas sur le raisonnement à partir d'une implication inverse étant donné la grande similitude entre ce connecteur et l'implication. Il s'agit en fait d'une implication où l'antécédent et le conséquent sont inversés. Nous nous restreindrons donc au raisonnement conditionnel tel qu'il est étudié dans la littérature. Nous le nommons ici par seul souci d'exactitude par rapport aux opérateurs de la logique classique.

invitée “Q est faux”. Par exemple, à partir des prémisses “Jean joue du violon ou du piano, Jean joue du violon”, un *MPT* consisterait à conclure « Jean ne joue pas de piano ». De même, à partir des prémisses “P ou Q, Q est vrai.” un second *MPT* mène à la conclusion invitée “P est faux” (e.g. “ Jean joue du violon ou du piano, Jean joue du piano. Donc, Jean ne joue pas de violon.”). Dans les deux cas, la conclusion invitée est l’une de deux conclusions possibles, si bien que la réponse logiquement correcte est l’incertitude à propos de cette dernière. Le tableau 1.5 résume les quatre inférences disjonctives présentées.

Tableau 1.5 Quatre inférences disjonctives

	<i>Modus Tollendo Ponens (MTP)</i>	<i>Modus Tollendo Ponens (MTP)</i>	<i>Modus Ponendo Tollens (MPT)</i>	<i>Modus Ponendo Tollens (MPT)</i>
Majeure	$P \vee Q$	$P \vee Q$	$P \vee Q$	$P \vee Q$
Mineure	$\sim P$	$\sim Q$	P	Q
Conclusion	Valide: Q	Valide: P	Invitée: $\sim Q$	Invitée: $\sim P$
			Valide: Incertitude	Valide: Incertitude
Rangée(s) vraie(s)	$(\sim P \ \& \ Q)$	$(P \ \& \ \sim Q)$	$(P \ \& \ \sim Q)$ $(P \ \& \ Q)$	$(\sim P \ \& \ Q)$ $(P \ \& \ Q)$

Le raisonnement déductif à partir d’une incompatibilité consiste à inférer à partir d’une prémisse majeure dans laquelle deux propositions sont incompatibles, combinée avec l’une des quatre prémisses mineures possibles, i.e. l’affirmation ou la négation de l’une des propositions. Les deux formes valides affirment l’une des propositions dans la prémisse mineure. Nous appellerons ces deux inférences le *Affirm First* et le *Affirm Second*. Le *Affirm First* consiste à raisonner à partir de la prémisse « P est incompatible avec Q, P est vrai » et mène à la conclusion valide, « Q est faux » (i.e. « Être à Montréal est incompatible avec être à Paris, John est à Montréal. Donc, John n’est pas à Paris »). À partir des prémisses « P est incompatible avec Q, Q est vrai » le *Affirm Second* mène à la conclusion valide « P est faux ». Par exemple, ce serait un raisonnement de la forme « Être à Montréal

est incompatible avec être à Paris, John est à Paris. Donc, John n'est pas à Montréal »). Les deux formes invalides nient l'une des propositions dans la prémisse mineure. Nous les appellerons le *Deny First* et le *Deny Second*. Le *Deny First* consiste à raisonner à partir des prémisses « P est incompatible avec Q, P est faux » et mène à la conclusion invitée « Q est vrai ». Par exemple, à partir des prémisses « Être à Montréal est incompatible avec être à Paris, John n'est pas à Montréal. Donc, John est à Paris ». À partir des prémisses « P est incompatible avec Q, Q est faux » le *Deny Second* mène à la conclusion invitée « P est vrai ». (i.e. « Être à Montréal est incompatible avec être à Paris, John n'est pas à Paris. Donc, John est à Montréal ». Dans les deux cas, la conclusion invitée est l'une de deux conclusions possibles, si bien que la réponse logiquement correcte est l'incertitude à propos de cette dernière.

Tableau 1.6 Quatre inférences à partir d'une incompatibilité

	<i>Affirm First</i>	<i>Affirm Second</i>	<i>Deny First</i>	<i>Deny Second</i>
Majeure	$P \mid Q$	$P \mid Q$	$P \mid Q$	$P \mid Q$
Mineure	P	Q	$\sim P$	$\sim Q$
Conclusion	Valide: $\sim Q$	Valide: $\sim P$	Invitée: Q Valide: Incertain	Invitée: P Valide: Incertain
Rangée(s) vraie(s)	$(P \ \& \ \sim Q)$	$(\sim P \ \& \ Q)$	$(\sim P \ \& \ Q)$ $(\sim P \ \& \ \sim Q)$	$(P \ \& \ \sim Q)$ $(\sim P \ \& \ \sim Q)$

Dans la perspective où la compétence fondamentale en raisonnement déductif est de savoir distinguer entre les conclusions nécessaires et celles qui sont possibles, l'implication, la disjonction inclusive et l'incompatibilité nécessitent une attention particulière dans l'identification des compétences en logique classique. C'est dans cette perspective que nous avons conçu notre outil d'évaluation automatique du raisonnement.

1.2. Effets du contenu : contre-exemples et perspective adoptée

Dans cette section, nous présenterons le volet sémantique de notre problématique cognitive dans deux perspectives : la perspective normative de la logique classique et la perspective descriptive du raisonnement humain en psychologie expérimentale. Nous proposerons une analyse de nos objectifs qui intègre ces deux perspectives ainsi qu'une identification de certaines compétences fondamentales en matière de raisonnement logique.

1.0.1 Le raisonnement conditionnel chez les humains

Il est possible de faire la distinction entre plusieurs types de raisonnement déductif au sein de la littérature actuelle en psychologie du raisonnement. Néanmoins, le raisonnement conditionnel est sans doute la forme la plus étudiée. L'implication permet de s'engager dans une pensée hypothético-déductive et est une part importante du raisonnement ainsi que de la prise de décision au quotidien (Evans & Over, 2004). Elle constitue une expression très importante de la langue française et permet d'exprimer des relations causales, des menaces, des obligations, des promesses ou des engagements.

1.0.2 Effets de contenu et débat sur la nature du raisonnement humain

Selon les lois de la logique, la validité d'une inférence ne dépend que de la structure formelle des prémisses. Autrement dit, quel que soit le contenu des prémisses, les conclusions des inférences valides devraient être considérées comme certaines à condition que les prémisses soient vraies. Les conclusions des inférences invalides devraient être considérées comme incertaines. Cependant, la littérature sur le raisonnement humain montre clairement que les raisonneurs n'évaluent pas les inférences uniquement sur la base de leurs propriétés formelles. En effet, de nombreuses études ont montré la grande variabilité du raisonnement humain avec des inférences de même forme logique, mais de contenu différent (c.g., Cummins, Lubart, Alksnis et Rist, 1991; Markovits et Vachon, 1990; Thompson, 1994).

Le type de raisonnement le plus étudié est le raisonnement conditionnel et, plus spécifiquement, sur les quatre formes d'inférences tirées d'une prémisse conditionnelle

(*MPP*; *MTT*; AC; DA). L'effet du contenu avec ces quatre formes d'inférences a été abondamment documenté. Une part importante des effets dus au contenu est causée par les antécédents alternatifs et les conditions contraignantes de la prémisse majeure. Les études montrent que plus une prémisse majeure contient d'antécédents alternatifs, plus les personnes ont tendance à reconnaître l'incertitude des inférences AC et DA. Les antécédents alternatifs sont des contre-exemples aux conclusions des inférences AC et DA alors que les conditions contraignantes sont des contre-exemples aux conclusions des inférences *MPP* et *MTT*. Un antécédent alternatif est un antécédent autre que P qui permet d'impliquer l'effet (Q). Par exemple un antécédent alternatif à la prémisse « Si un chien a des puces, alors il se gratte constamment » serait une autre raison pour laquelle un chien se gratte, comme une maladie de peau. D'un point de vue formel, il s'agit de la troisième rangée de la table de vérité de l'implication ($\sim P \ \& \ Q$). Les mêmes études montrent également que plus une prémisse majeure contient de conditions contraignantes, plus les personnes tendent à refuser la certitude des inférences *MPP* et *MTT*. Une condition contraignante est quant à elle une condition particulière qui permet de douter de la règle conditionnelle. Une condition contraignante à la prémisse précédente serait, par exemple, une condition dans laquelle un chien a des puces sans se gratter constamment, comme avoir une médication appropriée, une faible sensibilité, etc. Formellement, il s'agit de la deuxième rangée de l'opérateur, soit sa seule rangée fautive ($P \ \& \ \sim Q$).

Cette variabilité due au contenu au sein du raisonnement conditionnel est à la base de l'un des principaux débats entre deux grandes catégories de théories sur la nature du raisonnement déductif.

1.0.2.1 Les approches probabilistes

Les théories probabilistes considèrent que les raisonneurs estiment la probabilité de la conclusion sur la base de leurs connaissances statistiques à propos des prémisses (e.g. Evans, Over et Handley, 2005; Oaksford et Chater, 2007). La variabilité liée au contenu peut s'expliquer par l'effet des connaissances sur ces estimations. De plus, ces modèles décrivent un processus dans lequel des informations supplémentaires peuvent être utilisées

pour modifier les estimations via une mise à jour bayésienne. Lorsqu'on leur demande de faire une inférence déductive, qui nécessite un jugement de validité, les raisonneurs transformeraient leur estimation de probabilité en un jugement dichotomique. Ces modèles expliquent donc la variabilité du raisonnement humain liée au contenu ainsi que son caractère non monotone, c'est-à-dire qu'une conclusion préalablement tirée peut être modifiée suite à l'ajout d'information nouvelle.

1.0.2.2 La théorie des modèles mentaux

Une autre classe de théories se concentre particulièrement sur l'utilisation de l'information fournie dans les prémisses afin de générer des contre-exemples potentiels à la conclusion. La plus influente de ces théories est la théorie des modèles mentaux (Johnson-Laird, 2001; Johnson-Laird & Byrne, 2002). Selon cette théorie, les raisonneurs tirent des inférences déductives en se représentant des situations correspondant aux possibilités vraies inhérentes à la prémisse majeure. Ces possibilités correspondent aux rangées vraies de la table de vérité du connecteur logique de cette prémisse. Un raisonneur générera une conclusion si elle fait partie d'au moins un modèle. Il ira ensuite à la recherche de contre-exemples, c'est-à-dire des situations possibles qui constituent des alternatives à la conclusion. Si aucun contre-exemple n'est trouvé, la conclusion est acceptée. Dans des conditions idéales, la mémoire de travail du raisonneur permet une représentation exhaustive de toutes les possibilités vraies inhérentes à la prémisse majeure. Cependant, ce processus peut être influencé par plusieurs facteurs, dont l'interprétation sémantique de la prémisse majeure, certains effets pragmatiques ainsi que les limites de la mémoire de travail du raisonneur.

1.0.2.3 La théorie de la récupération sémantique

Bien que la théorie de modèles mentaux explique plusieurs phénomènes liés au raisonnement conditionnel, elle ne rend pas compte de l'impact du nombre d'antécédents alternatifs d'une prémisse sur les inférences invalides. Dans une variante de la théorie des modèles mentaux, appelée théorie de la récupération sémantique, Markovits et Barouillet

(2002) expliquent la variabilité liée au contenu dans le raisonnement conditionnel. Selon ces auteurs, le raisonnement au quotidien dépend largement de la récupération de l'information au sein d'un espace sémantique. La théorie suppose que les modèles sont produits par un processus de récupération en mémoire des informations sur les prémisses majeures. Prenons à titre d'exemple une prémisses majeure telle que « Si un animal est un chien, alors il a quatre pattes ». La théorie de la récupération sémantique suggère qu'un individu construit avant tout une représentation mentale de la prémisses majeure appelée le modèle initial :

P Q Ex : Un animal est un chien Il a quatre pattes

Lorsque ce même individu traite ensuite une prémisses mineure de forme AC (« un animal a quatre pattes »), ce dernier doit générer un antécédent alternatif afin de rejeter la conclusion invalide (cette conclusion étant « l'animal est un chien »). Dans ce cas particulier, l'affirmation du conséquent active en mémoire la catégorie d'animaux qui ont quatre pattes. Cette catégorie, en plus de comporter beaucoup d'éléments, contient l'antécédent (« un animal est un chien »). La récupération de l'information au sein d'une même catégorie n'est pas exigeante pour la mémoire de travail. Conséquemment, ce type de prémisses permet à de jeunes enfants, pour lesquels le raisonnement n'est pas aisé, de raisonner logiquement en générant facilement un antécédent alternatif tel qu'un chat (Markovits, 2000; Markovits, Venet, Janveau-Brennan, Malfait, Pion, & Vadeboncoeur, 1996). Considéré dans le cadre de la théorie de la récupération sémantique, on obtient les modèles suivants :

P Q Ex : Un animal est un chien Il a quatre pattes

B Q Un animal est un chat Il a quatre pattes

Le contenu de la prémisses permet de récupérer facilement en mémoire à long terme un antécédent alternatif (*b*) à l'implication « si P, alors Q ». Ce processus de récupération ne nécessite pas une inhibition importante de l'information et est très peu exigeant pour la mémoire de travail.

Comme dans la théorie des modèles mentaux, la ligne « $P \rightarrow Q$ » représente le cas d'occurrence de l'antécédent et du conséquent de la prémisse majeure. Toutefois, les antécédents alternatifs sont représentés par un cas spécifique (B) dans la seconde ligne du modèle. Ainsi, les contre-exemples générés durant le processus de *fleshing-out*, plutôt qu'être des représentations abstraites d'une possibilité, sont des instances de cette dernière récupérées à partir de la mémoire à long terme. La seconde ligne de l'exemple précédent sert de contre-exemple à une conclusion pour l'inférence AC (« Un animal a quatre pattes, donc c'est un chien»). Cette dernière est donc considérée comme incertaine et les participants fournissent alors une réponse correcte sur le plan logique. Maintenant, un modèle qui inclut une condition contraignante est donné par l'exemple suivant :

P	Q	Ex :	Je mange beaucoup de bonbons	J'ai des caries
	non-		Je mange beaucoup de bonbons &	
P & D	Q		Je me brosse souvent les dents	Je n'ai pas de caries

Dans la seconde ligne de ce modèle, « $P \& D$ » représente une condition contraignante possible. Cette ligne constitue un contre-exemple à la conclusion d'un *MMP* (« je mange beaucoup de bonbons donc j'ai des caries »). Dans ce cas, les participants répondent, à tort, que la conclusion n'est pas valide.

1.0.2.4 Volet développemental de la récupération sémantique

Dans le cadre de la théorie de la récupération sémantique, l'efficacité du processus de récupération et le nombre de contre-exemples présents dans la mémoire à long terme du raisonneur déterminent la probabilité qu'au moins un puisse être récupéré. Pour être conforme à la logique classique, le raisonnement exige, d'une part, la récupération d'antécédents alternatifs contenus dans la mémoire sémantique à long terme et, d'autre part, l'inhibition des conditions contraignantes liées à la prémisse.

Jusqu'à maintenant, nous nous sommes concentrés sur les processus de raisonnement avec du contenu qui réfère aux connaissances empiriques du raisonneur à propos des prémisses,

mais qu'en est-il de la pensée abstraite où le raisonneur ne peut avoir recours à son bagage de connaissances ? Markovits (2013) suggère que plus une prémisse présente un contenu familier, plus la récupération d'un contre-exemple aux conclusions invalides est rapide et aisée. Conséquemment, plus une prémisse est familière, plus les raisonneurs ont tendance à éviter les erreurs de raisonnement. En ce sens, Markovits a élaboré une échelle graduée des niveaux de contenu pour une prémisse qui croît en difficulté et en niveau d'abstraction. La transition entre deux niveaux de contenu est graduelle et faite par étapes : lorsqu'un niveau est maîtrisé, la compétence dans le niveau supérieur débute.

Le premier niveau de raisonnement permet de générer une catégorie d'antécédents alternatifs dans un contenu familier. Par exemple, avec la prémisse causale familière « Si on lance une roche dans la vitre, alors la vitre brisera. » le raisonneur est en mesure de générer des antécédents au sein de la catégorie familière « différentes manières de briser une vitre. » Ce niveau est le premier échelon de l'échelle et le plus aisé à maîtriser par les adultes. On distingue, au sein de cette échelle, les prémisses qui ont beaucoup d'antécédents alternatifs et celles qui en ont peu. Les premières permettent la construction des items de raisonnement les plus simples et les secondes augmentent légèrement le niveau de difficulté.

Le second niveau de compétence (le niveau contrefactuel) permet la génération d'une catégorie irréaliste (plus abstraite) d'antécédents alternatifs. Par exemple, avec la prémisse contrefactuelle « Si on lance un mouchoir dans la vitre, alors la vitre brisera. », le raisonneur peut générer un antécédent alternatif contrefactuel comme « lancer une plume dans la vitre ».

Le troisième niveau de compétence est le niveau abstrait. Il demande une représentation abstraite des prémisses qui permet de générer des antécédents alternatifs abstraits. Lorsque le raisonneur atteint ce niveau, il a une compréhension complète du lien d'implication : il comprend que pour une prémisse conditionnelle (inconnue ou abstraite), un antécédent alternatif peut être généré, et ce, indépendamment de son bagage de connaissances. Par exemple, avec la prémisse abstraite « Si une personne blop, alors elle devient brandup. »,

le raisonneur doit comprendre que le lien conditionnel entre l'antécédent et le conséquent n'exclut pas la possibilité d'antécédents alternatifs.

1.0.3 Perspective adoptée : Le cadre de la mémoire sémantique

Plutôt que de mettre à l'épreuve les théories rivales sur la nature du raisonnement humain, le présent projet s'inscrit dans une perspective plus générale : il vise à identifier les phénomènes largement observés ainsi qu' à élargir la portée de ces débats. D'abord, chacun des modèles recensés ci-haut repose sur une hypothèse fondamentale : les mécanismes centraux du raisonnement humain sont très sensibles aux propriétés sémantiques du problème. Les processus de raisonnement sous-jacents étant une estimation de probabilité de la conclusion (approches probabilistes) ou une recherche approfondie de contre-exemples (récupération sémantique), les deux perspectives prédisent une tendance à rejeter la conclusion qui grandit avec le nombre de contre-exemples permis par les prémisses. Nous adoptons donc une perspective décrite par De Neys, Schaeken & d'Ydewalle (2002) comme étant le cadre de la mémoire sémantique. Selon cette perspective, l'impact de la récupération d'information sémantique sur le raisonnement déductif constitue une base commune pour le débat actuel sur la nature du raisonnement. Cet impact a été observé de différentes manières, mais principalement par le nombre d'antécédents alternatifs et de conditions contraignantes (Cummins, 1995; Cummins et al. 1991; Thompson, 1994) ou par la force de l'association entre l'antécédent et le conséquent, soit un contre-exemple plus fortement associé à la prémisse majeure, ce qui facilite sa récupération en mémoire à long terme (Quinn & Markovits, 1998; De Neys, Schaeken & d'Ydewalle, 2003a). Cependant, la conception du modèle de l'apprenant de Muse-Logique demande d'adopter une théorie psychologique sous-jacente. La théorie de la récupération sémantique est la plus près de cette perspective et fournit l'explication la plus fine des effets du contenu dans le raisonnement. De plus, son volet développemental constitue un avantage majeur pour modéliser la progression des compétences au raisonnement dans un contexte d'apprentissage. En effet, à notre connaissance, l'échelle des niveaux de contenu de Markovits est le modèle le plus explicite et développé quant aux effets du contenu dans

le raisonnement logique chez les humains. Nous nous baserons donc sur cette théorie pour la conception du modèle de l'apprenant de Muse-Logique.

1.3. Structures logiques et compétences visées

Précédemment, nous avons soutenu qu'en plus de l'implication, la disjonction inclusive et l'incompatibilité sont des connecteurs particulièrement importants dans le cadre du raisonnement logique. Bien que la littérature en psychologie du raisonnement nous informe principalement sur les processus de raisonnement conditionnel, une analyse qui combine ces travaux à un examen formel des opérateurs logiques permet de faire des prédictions sur la manière dont les humains raisonneront à partir des autres connecteurs. En effet, les contre-exemples aux conclusions invitées des inférences AC et DA sont tous représentés par la même expression logique, c'est-à-dire, lorsque P est faux et que Q est vrai ($\sim P \ \& \ Q$). Or, cette expression détient une propriété logique particulière: il s'agit du dual de l'opérateur d'implication. Autrement dit, la capacité à générer des antécédents alternatifs est, d'un point de vue formel, la capacité à générer des instances du dual de l'opérateur de la prémisse majeure, ce dernier jouant le rôle de contre-exemple aux conclusions invitées (Robert & Brisson, 2016).

La dualité est une propriété logique qui combine deux autres propriétés : la réciprocity et l'inversion. Plus précisément, deux opérateurs sont réciproques si les expressions qu'ils forment sont équivalentes lorsque les propositions connectées par l'un des deux sont niées. C'est le cas de l'implication et de l'implication inverse. Autrement dit, l'expression « Si P, alors Q » est équivalente à « Si non-Q, alors non-P ». Par exemple, la règle « S'il pleut, alors il y a des nuages » est équivalente à la règle « S'il n'y a pas de nuages, alors il ne pleut pas ». Ainsi, $(P \supset Q) \equiv (\sim P \subset \sim Q)$ ou $(\sim P \supset \sim Q) \equiv (P \subset Q)$ est une loi de la logique classique. De plus, deux opérateurs sont inverses si les expressions qu'ils forment sont équivalentes lorsque l'une d'elles est niée. C'est le cas, par exemple, de la tautologie $((P \supset Q) \supset (P \supset Q))$ qui est toujours vraie) et de la contradiction $((P \supset Q) \supset (P \supset \sim Q))$ qui est toujours fausse). En effet, lorsqu'on nie une expression toujours vraie, on obtient une expression qui est toujours fausse. Ainsi, $(P \supset Q) \equiv \sim(P \supset \sim Q)$ ou $(P \supset \sim Q) \equiv \sim(P \supset Q)$, sont des lois de la logique

classique. Finalement, un opérateur est le dual d'un autre s'il est l'inverse de sa réciproque ou la réciproque de son inverse. L'expression $(\sim P \ \& \ Q)$ est le dual de l'implication. Ainsi $(P \supset Q) \equiv \sim(P \ \& \ \sim Q)$ est aussi une loi logique. Cette dernière exprime que l'implication est équivalente à une expression qui dit "Ce n'est pas le cas que P et non-Q", P et non-Q étant le seul cas faux de l'implication.

Lorsqu'un groupe d'opérateurs est tel que chaque opérateur est l'inverse, le réciproque et le dual d'autres opérateurs, les quatre opérateurs réunis forment un groupe de Klein. De plus, lorsque les opérateurs inverse, réciproque et dual sont distincts les uns des autres, ces derniers forment un groupe de Klein complet. Il existe deux groupes de Klein complets en logique classique. Le premier, illustré dans la Figure 1.1, est composé de l'implication et l'implication inverse (en haut de la figure) ainsi que leurs duals respectifs (au bas de la figure).

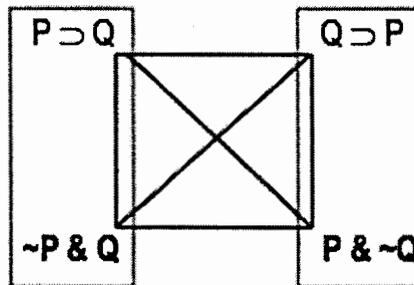


Figure 1.1 Groupe de Klein pour l'implication et l'implication inverse

Le second est illustré dans la Figure 1.2 et est composé de la disjonction inclusive et de l'incompatibilité (haut de la figure) et de leurs duals (bas de la figure)

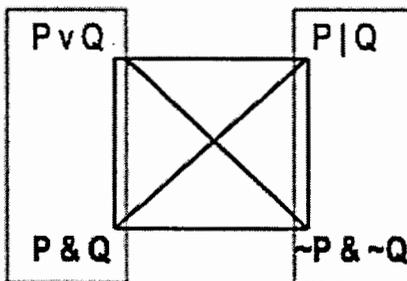


Figure 1.2 Groupe de Klein pour la disjonction inclusive et l'incompatibilité

Tel qu'exposé dans cette figure, le dual de la disjonction inclusive est l'expression $(P \& Q)$. En revenant au Tableau 1.5 et aux descriptions des inférences disjonctives, on remarque que cette expression représente les contre-exemples aux conclusions des inférences invalides. En effet, à partir de la prémisse majeure P ou Q , conclure non- Q sur la base de P (ou conclure non- P sur la base de Q) est invalide puisqu'une autre situation est possible : celle où P et Q sont le cas. De même, à partir de la prémisse majeure P est incompatible avec Q , conclure Q sur la base de non- P (ou conclure P sur la base de non- Q) est invalide puisque le dual de l'incompatibilité (non- P et non- Q) est aussi possible (Robert, & Brisson, 2016).

Cette analyse suggère donc que l'implication, la disjonction inclusive et l'incompatibilité, étant les trois opérateurs classiques qui ont un dual distinct, représentent le défi principal en matière de raisonnement logique. Notre outil d'évaluation automatique pour Muse-Logique se concentrera donc sur le raisonnement à partir de ces trois opérateurs. (Pour une analyse complète du lien entre ces structures logiques et le raisonnement humain, voir Robert & Brisson, 2016.)

1.4. Un réseau bayésien comme outil d'évaluation automatique du raisonnement

Au sein d'un STI, la modélisation de l'état des compétences de l'apprenant est fondamentale pour le bon fonctionnement du système, tant au niveau de son évaluation qu'au niveau du choix de la stratégie pédagogique à utiliser. Le modèle de l'apprenant est essentiel pour faire en sorte que le STI s'adapte au niveau d'apprentissage approprié pour l'apprenant.

Dans le cas de Muse-Logique, une évaluation qui ferait uniquement la somme des bonnes réponses aux problèmes d'inférence ne suffit pas à conclure que l'apprenant possède l'une des compétences visées. En effet, plusieurs adultes utilisent des stratégies heuristiques de raisonnement qui mènent à l'acceptation des inférences valides sans que les inférences invalides ne soient évitées, alors que d'autres heuristiques mènent à refuser les quatre

inférences de l'implication (Markovits, 1985). Ainsi, une simple sommation des bonnes réponses classerait ces raisonneurs comme étant compétents à 50% alors qu'aucun d'entre eux ne maîtrise le raisonnement conditionnel. Pour pallier à ce problème, le meilleur critère d'évaluation de la compétence logique est la capacité de considérer une inférence valide comme étant certaine et une invalide comme étant incertaine (Markovits, 2013).

De plus, au cours de son progrès dans le système, les compétences de l'apprenant sont appelées à changer et les interventions pédagogiques du tuteur sont adaptées à l'état actuel de ses compétences. Il est donc nécessaire que l'évaluation de l'apprenant puisse être faite et modifiée en temps réel, et ce, tout au long de son utilisation du système. De plus, une compétence doit être attribuée avec un certain degré de certitude, si bien que la modélisation d'un apprenant ne peut être qu'une approximation de son état réel. Il est donc primordial que le STI puisse s'appuyer sur un formalisme qui permette de générer des inférences incertaines à propos de l'apprenant (Mayo & Mitrovic, 2001). La théorie bayésienne des probabilités répond tout à fait à ces exigences: elle permet d'inférer la probabilité de bien réussir à un exercice (B), étant donné la maîtrise d'une certaine compétence (A). De plus, les réseaux bayésiens permettent une représentation structurale de l'inférence bayésienne à propos de différentes compétences étant donné un certain patron de réponses. Ce type de structure s'avère donc parfaitement adaptée pour représenter les compétences quant aux inférences à partir des trois opérateurs logiques identifiés et leurs différents niveaux de contenu.

1.5. Effets de la forme : intuition logique et compétences visées

Jusqu'à maintenant, nous avons exposé notre perspective sur les processus de raisonnement déductif en nous concentrant sur la large littérature concernant les effets du contenu sur le raisonnement humain. Cependant, des études récentes ont montré que la forme logique d'une inférence peut elle aussi influencer le raisonnement au quotidien. En effet, plusieurs études récentes ont montré que les raisonneurs semblent détecter le conflit entre leurs réponses intuitives et certaines règles très simples provenant de la logique, de

l'arithmétique et de la théorie des probabilités (e.g. Bago & De Neys, 2017; De Neys, Cromheeke, & Osman, 2011; Pennycook, Trippas, Handley, & Thompson, 2014).

À partir de cette littérature, De Neys (2012, 2014) a proposé que certains raisonneurs ont une compréhension intuitive des principes logiques de base. Autrement dit, lorsqu'un raisonneur fournit une réponse biaisée à un problème logique simple, il peut détecter son erreur intuitivement sans toutefois être en mesure de l'inhiber. Selon lui, l'intuition logique serait le résultat d'une automatisation de certains principes logiques de base provenant d'une exposition répétée à ces derniers.

Des études récentes apportent un soutien empirique à cette proposition. Par exemple, certains chercheurs ont observé ce que nous pourrions appeler un « biais logique ». Ce biais est lié au biais de croyance, c'est-à-dire la tendance à émettre un jugement de validité en se basant sur concordance entre ses croyances et une conclusion plutôt que sur la structure logique de l'inférence. Cette heuristique rend les raisonneurs plus enclins à endosser une inférence invalide lorsque sa conclusion est croyable et à rejeter une inférence valide lorsque sa conclusion est incroyable. Des études ont exploré la possibilité d'un phénomène inversé, c'est-à-dire la possibilité que la validité logique d'un argument puisse biaiser les jugements de croyance en la conclusion. Ces études ont présenté des inférences où validité et croyance étaient en accord ou en conflit et ont demandé aux participants les conclusions dans un court laps de temps (Handley & Trippas, 2015; Trippas, Handley, Verde et Morsanyi, 2016 ; Trippas, Thompson et Handley, 2017). Les chercheurs ont constaté que les participants avaient besoin de plus de temps et étaient plus sujets aux erreurs lorsque la validité logique des problèmes entraînait en conflit avec la crédibilité de la conclusion. D'autres études ont révélé que même dans des conditions rapides et difficiles, les participants acceptent les inférences valides plus souvent que les invalides, suggérant ainsi qu'une réponse logique puisse provenir de processus rapides et intuitifs (Bago & De Neys, 2017; Newman, Gibb, & Thompson, 2017).

Étant donné, d'une part, le nombre grandissant d'études appuyant cette théorie et d'autre part, l'exposition répétée à des inférences logiques lors de l'utilisation de Muse-Logique, il apparaît très pertinent d'inclure ce concept dans le modèle de l'apprenant du STI.

Cependant, étant donné la récence de ces travaux, la portée et les limites de l'intuition logique doivent encore être déterminées, et ce, même au niveau du raisonnement conditionnel.

1.6. Logique classique et norme du raisonnement valide : clarifications

L'utilisation de la logique classique comme norme du raisonnement au sein de cette thèse pourrait suggérer que nous adoptons ce système formel comme norme absolue du raisonnement valide. La perspective adoptée au sein de ce travail est toutefois bien différente et mérite clarifications. D'abord, le travail effectué au sein du STI Muse-Logique peut être fait en utilisant toute logique non classique comme système de référence. La perspective adoptée au sein de cette thèse n'a donc aucune prétention quant au caractère absolu de la logique classique comme norme du raisonnement valide. La logique classique a donc été adoptée comme point de départ et ce, pour les raisons suivantes. D'abord, afin de remplir le mandat de cette thèse, nous avons consulté la littérature en psychologie du raisonnement afin d'en extraire les phénomènes largement observés, dont l'existence fait consensus au sein de la communauté. Nous avons alors identifié les effets du contenu (observés depuis plusieurs décennies), ainsi que les effets de la forme logique (observés plus récemment) dans le raisonnement conditionnel. Il appert que ces phénomènes sont observés par la biais de tâche de raisonnement syllogistiques, toutes définies dans le cadre de la logique classique. Ainsi, par soucis de continuité avec la littérature actuelle ainsi que le matériel qu'elle utilise, nous avons dû nous concentrer sur la logique classique comme système formel de choix. De plus, dans un souci pédagogique, les logiques non classiques étant définies comme des extensions de la logique classique, cette dernière constitue la première étape à franchir pour un apprenant de Muse-Logique. Le travail effectué au sein de cette thèse est donc mené relativement à ce choix pragmatique et pédagogique. Nous adoptons cependant un point de vue agnostique quant à savoir si les humains adoptent, ou devraient adopter, la logique classique comme norme de leurs raisonnements au quotidien.

CHAPITRE II

UN RÉSEAU BAYÉSIEN COMME MODÈLE DE L'APPRENANT

2.1 Calcul des probabilités

Les réseaux bayésiens étant intrinsèquement liés au calcul des probabilités, nous commencerons ce chapitre par quelques notions de base de ce calcul. D'abord, la loi générale de la probabilité conditionnelle, généralement utilisée pour calculer la probabilité qu'un événement (E) advienne étant donnée une hypothèse (H) est la suivante:

$$\Pr(E|H) = \frac{\Pr(E \& H)}{\Pr(H)} \quad (2.1)$$

Cette loi énonce que la probabilité conditionnelle qu'un événement (E) advienne étant donné une hypothèse (H) est égale à la probabilité que l'événement et l'hypothèse adviennent tous les deux divisée par la probabilité de l'hypothèse. Le tout n'est possible que si la probabilité de H est supérieure à zéro, puisqu'une division par zéro est impossible (c'est-à-dire que l'hypothèse H est probable). Par exemple, si on lance un dé non pipé, la probabilité que le dé tombe sur 6 est de 1/6. Toutefois, la probabilité conditionnelle que le dé tombe sur 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 (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. Pr (6 & pair) étant de 1/6 et Pr (pair) étant de 1/2, la loi permet de faire le calcul suivant (Hacking, 2001):

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

2.1.1 Loi de Bayes

Alors que la loi générale de la probabilité conditionnelle permet de calculer la probabilité d'un événement étant donné une hypothèse, la loi de Bayes permet de calculer la probabilité d'une hypothèse étant donné l'observation d'un événement ($\Pr(H|E)$) en se basant sur la probabilité que l'hypothèse produise l'effet ($\Pr(E|H)$), qu'on appelle aussi la mesure de vraisemblance de l'hypothèse, ainsi que la probabilité de l'hypothèse et la probabilité de l'événement :

$$\Pr(H|E) = \frac{\Pr(E|H) \times \Pr(H)}{\Pr(E)} \quad (2.2)$$

La loi de Bayes peut être dérivée à partir de la loi générale de probabilité de la manière suivante :

- | | | |
|----|--|--|
| 1. | $\Pr(E H) = \frac{\Pr(E \& H)}{\Pr(H)}$ | Loi de probabilité générale |
| 2. | $\Pr(H E) = \frac{\Pr(H \& E)}{\Pr(E)}$ | 1. Substitution E/H et H/E |
| 3. | $\Pr(E \& H) = \Pr(E H) \times \Pr(H)$ | 1. Produit croisé |
| 4. | $\Pr(H \& E) = \Pr(E H) \times \Pr(H)$ | 3. Commutativité de la conjonction |
| 5. | $\Pr(H E) = \frac{\Pr(E H) \times \Pr(H)}{\Pr(E)}$ | 2, 4. Remplacement de $\Pr(H \& E)$ par $\Pr(E H) \times \Pr(H)$ |

Lorsqu'un événement peut être expliqué par un ensemble d'hypothèses exhaustives et mutuellement exclusives, la probabilité de cet événement équivaut à la somme des probabilités conjointes de cet événement et de chacune des hypothèses :

$$\Pr(E) = \sum_{k=1}^n \Pr(E \& H_k) \quad (2.3)$$

Or, tel qu'exposé dans la preuve ci-haut, un produit croisé effectué à partir de la loi de probabilité générale permet d'exprimer une probabilité conjointe de la manière suivante :

$$\Pr(E) = \sum_{k=1}^n \Pr(E \& H_k) = \sum_{k=1}^n \Pr(E|H_k) \times \Pr(H_k) \quad (2.4)$$

En remplaçant $\Pr(E)$ dans la loi de Bayes par l'équation 2.4, on obtient une version de cette loi qui permet de calculer la probabilité d'une hypothèse, choisie parmi un ensemble d'hypothèses, à partir de plusieurs événements. Il est aussi possible de calculer cette probabilité à partir d'un ensemble d'évènements :

$$\Pr(H_i|E_1 E_2 \dots E_n) = \frac{\Pr(E_1 E_2 \dots E_n|H_i) \times \Pr(H_i)}{\sum_{k=1}^n \Pr(E_1 E_2 \dots E_n|H_k) \times \Pr(H_k)} \quad (2.5)$$

Notons toutefois que ce calcul n'est possible que si les hypothèses et les événements sont exhaustifs et mutuellement exclusifs.

Cependant, l'équation 2.5 exige le calcul de toutes les combinaisons possibles d'évènements pour chaque hypothèse, ce qui alourdit considérablement la tâche du système. Il est donc nécessaire de présumer le postulat d'indépendance conditionnelle des événements selon lequel l'occurrence d'un événement n'a pas d'incidence sur la probabilité d'occurrence des autres événements (Negnevitsky, 2005). En appliquant ce postulat, on obtient

$$\Pr(H_i|E_1 E_2 \dots E_n) = \frac{\Pr(E_1|H_i) \times \Pr(E_2|H_i) \times \dots \times \Pr(E_n|H_i) \times \Pr(H_i)}{\sum_{k=1}^n \Pr(E_1|H_k) \times \Pr(E_2|H_k) \times \dots \times \Pr(E_n|H_k) \times \Pr(H_k)} \quad (2.6)$$

où la probabilité de chaque événement étant donné chaque hypothèse est calculée indépendamment des autres événements.

2.1.2 Calcul des probabilités postérieures

L'équation 2.6 permet à un système de calculer la probabilité postérieure de chaque hypothèse (parmi un ensemble d'hypothèses) étant donné l'observation d'un ou de plusieurs événements parmi un ensemble d'événements. La distribution *a priori* des probabilités, c'est-à-dire la probabilité antérieure de chaque hypothèse ($\Pr(H_i)$) ainsi que la vraisemblance de chaque hypothèse pour chaque événement ($\Pr(H_i|E_k)$), peut être préalablement fournie par un expert du domaine dans lequel le système raisonne. Nous verrons plus loin que ces probabilités peuvent aussi être apprises à partir de données.

À titre d'exemple, le Tableau 2.1 présente la distribution de probabilités pour trois hypothèses mutuellement exclusives ainsi que les probabilités conditionnelles pour chacune de ces hypothèses étant donné trois événements indépendants telle qu'exposée dans le Tableau 2.1.

Tableau 2.1 Distribution de probabilités (Negnevitsky, 2005)

Probability	Hypothesis		
	$i = 1$	$i = 2$	$i = 3$
$p(H_i)$	0.40	0.35	0.25
$p(E_1 H_i)$	0.3	0.8	0.5
$p(E_2 H_i)$	0.9	0.0	0.7
$p(E_3 H_i)$	0.6	0.7	0.9

Supposons qu'on désire obtenir la probabilité conditionnelle de l'hypothèse 1 (H_1) étant donné l'évènement 3 (E_3). En plaçant les variables du problème dans l'équation 2.6, on obtient l'équation suivante :

$$\Pr(H_1|E_3) = \frac{\Pr(E_3|H_1) \times \Pr(H_1)}{\sum_{k=1}^3 \Pr(E_3|H_k) \times \Pr(H_k)}$$

En remplaçant les variables par leur valeur associée, on obtient :

$$\Pr(H_1|E_3) = \frac{0,6 \times 0,4}{0,6 \times 0,4 + 0,7 \times 0,35 + 0,9 \times 0,25} = 0,34$$

La probabilité conditionnelle de l'hypothèse 1 étant donné l'évènement 3 est donc de 0,34. Cette probabilité peut changer si on ajoute de l'information dans le système. Par exemple, si on observe aussi l'évènement 1 (E_1), l'équation X devient :

$$\Pr(H_1|E_1E_3) = \frac{\Pr(E_1|H_1) \times \Pr(E_3|H_1) \times \Pr(H_1)}{\sum_{k=1}^3 \Pr(E_1|H_k) \times \Pr(E_3|H_k) \times \Pr(H_k)}$$

En remplaçant les variables par leur valeur associée, on obtient :

$$\Pr(H_1|E_1E_3) = \frac{0,3 \times 0,6 \times 0,4}{0,3 \times 0,6 \times 0,4 + 0,8 \times 0,7 \times 0,35 + 0,5 \times 0,9 \times 0,25} = 0,19$$

Ainsi, après l'observation de l'évènement 1, la probabilité conditionnelle de l'hypothèse 1 diminue alors à 0,19. (Negnevitsky, 2005)

2.1.3 Inférences possibles avec la théorie des probabilités conditionnelles

La probabilité conditionnelle peut être vue comme étant analogue à l'implication logique. En effet, la probabilité conditionnelle permet d'exprimer la relation de causalité, c'est-à-dire la probabilité d'un lien causal entre une hypothèse et un évènement. Cette relation est analogue au lien implicatif, entre un antécédent (cause) et un conséquent (effet). Dans cette perspective, l'implication en logique classique devient un cas limite de la probabilité conditionnelle, où les propositions ne prennent que des valeurs classiques (0 ou 1). Il est donc possible de comparer les raisonnements cautionnés par le calcul des probabilités avec les raisonnements de la logique classique, ce qui permet de voir les avantages du calcul des probabilités.

D'abord, la probabilité postérieure ($\Pr(H|E)$) expose la probabilité d'une cause possible étant donné un effet. Si on entre l'observation d'un évènement E dans le système, il pourra inférer la probabilité de l'hypothèse H et effectuer un raisonnement abductif, aussi appelé inférence diagnostique, analogue à l'inférence de l'affirmation du conséquent en logique. Un exemple serait la probabilité d'un profil de compétences pour un apprenant étant donné son patron de réponse aux questions qui lui ont été posées.

Ensuite, l'hypothèse de vraisemblance ($Pr(E|H)$) expose la probabilité d'un effet étant donné une cause. Ainsi, si on entre dans le système qu'une hypothèse H est le cas, il pourra inférer la probabilité de l'événement E et ainsi effectuer un raisonnement déductif, aussi appelé inférence causale, analogue à l'inférence par *Modus Ponendo Ponens*. Par exemple, il inférera la probabilité d'une réponse à une question sachant le profil de l'apprenant. Des variantes de ce calcul permettent aussi de calculer la probabilité d'une hypothèse après l'observation d'un événement et d'une hypothèse alternative. Par exemple, la probabilité qu'un patient soit atteint d'une pneumonie étant donné qu'il a de la toux peut être recalculée après qu'on ait déterminé que le patient est atteint d'une grippe. On calculera alors la probabilité $Pr(H_1|E \& H_2)$, où H_2 est une cause alternative pour E . Ce type de raisonnement est appelé une inférence inter-causale et permet d'expliquer (*explain away*) une observation. Ce type d'inférence est un avantage particulier du calcul des probabilités comparativement à la logique puisqu'il cautionne un raisonnement non-monotone: une conclusion préalablement tirée peut être modifiée suite à l'ajout d'information nouvelle. Il est aussi possible de combiner les inférences énumérées ci-haut et faire une inférence mixte. Par exemple, on peut calculer la probabilité d'une cause étant donné l'observation d'un effet et d'une cause d'ordre supérieur. On fera alors une inférence diagnostique combinée à une inférence causale. Il pourrait alors s'agir de calculer la probabilité qu'un patient soit atteint d'une pneumonie étant donné qu'il a une toux et qu'il est infecté par un streptocoque (Russell & Norvig, 2009).

Ainsi, ces quatre formes d'inférences effectuées au sein du calcul des probabilités permettent au système d'effectuer des inférences déductives et abductives tout en quantifiant l'incertitude des connaissances dans le domaine sur lequel il raisonne.

2.2 Les réseaux bayésiens

Un Réseau Bayésien (RB) est constitué de deux éléments fondamentaux : une structure, exprimée de manière graphique, et des paramètres, exprimés par des distributions de probabilité. La structure d'un RB est un graphe fini, acyclique et orienté, c'est-à-dire un graphe qui a un nombre fini de nœuds, dont aucun chemin orienté ne renvoie au nœud de

départ et dont tous les liens ont une orientation allant d'un nœud vers un autre. Chaque lien représente une dépendance bayésienne (souvent interprété comme une causalité) entre deux nœuds du graphe, ce qui définit des nœuds parents et des nœuds enfants. Cette particularité est fondamentale dans la détermination de la structure d'un RB : il ne peut y avoir de lien de dépendance entre des nœuds qui n'ont pas de lien de parenté.

Les probabilités conditionnelles associées à chaque nœud peuvent représenter une distribution locale de probabilités et la distribution jointe de probabilités. La distribution locale de probabilité représente la distribution de probabilité de chaque nœud considérant toutes les valeurs de ses nœuds parents. Autrement dit, la distribution locale de probabilité représente la probabilité qu'une variable (X_i) ait la valeur x_i étant donné l'état de l'ensemble de ses nœuds parents ($pa(X_i)$). Lorsque les probabilités conditionnelles de tous les nœuds étant données toutes les valeurs de leurs parents sont représentées, on obtient une distribution jointe de probabilités. Cette manière de représenter les distributions de probabilité expose le présupposé selon lequel un nœud n'est dépendant que de ses nœuds parents, ce qui permet de ne considérer que les nœuds pertinents pour le calcul d'une probabilité (Conati, 2010).

Prenons, par exemple, la variable *Campfire* de la Figure 2.1. La structure (à gauche) exprime le fait que le nœud *Campfire* est dépendant des nœuds *Storm* et *BusTourGroup* (qui sont ses nœuds parents), mais indépendant des autres nœuds du réseau. Le tableau (à droite) est la distribution locale de probabilité pour le nœud *Campfire* étant donné l'état de ses nœuds parents. Par exemple, l'entrée au coin supérieur gauche du tableau qui y figure exprime que la probabilité que *Campfire* soit vraie étant donné que *Storm* et *BusTourGroup* sont vrais est de 0.4 (Mitchell, 1997).

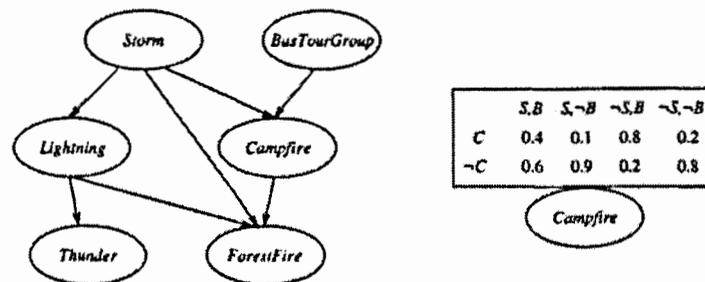


Figure 2.1 Réseau bayésien et distribution de probabilités (Mitchell, 1997)

Un avantage des réseaux bayésiens est qu'ils permettent de représenter les connaissances *a priori* concernant la dépendance causale entre certaines variables du domaine modélisé. Il est donc possible de construire la structure d'un RB à partir des connaissances des experts du domaine.

Lorsqu'un RB est utilisé pour la modélisation d'un apprenant, des nœuds observables représentent un ensemble d'items et des nœuds cachés (ou latents) représentent une compétence du domaine d'apprentissage. Les liens entre les nœuds représentent la dépendance bayésienne entre la réussite d'un item et une compétence du domaine, entre deux compétences ou entre deux items. Le modèle de l'apprenant généré par cette structure contient donc deux parties : le modèle de la compétence et le modèle de la performance (Almond & al., 2007). Ce formalisme est tout à fait adéquat pour représenter les compétences pertinentes au raisonnement déductif ainsi que les performances aux différents items. Pour le modèle de l'apprenant de Muse-Logique, nous avons donc construit la structure d'un RB en tant qu'outil d'évaluation automatique du raisonnement conditionnel.

2.2.1 Muse : Un réseau bayésien à validation experte

La structure à validation experte que nous avons construite et l'identification des compétences visées pour une maîtrise complète du raisonnement conditionnel sont basées sur les effets du contenu dans le raisonnement, l'approche développementale de Markovits (2013) ainsi que l'approche formelle de Robert et Brisson (2016) exposées précédemment. Les compétences visées sont l'inhibition des conditions contraignantes, la génération des antécédents alternatifs ainsi que la bonne gestion mentale des trois situations pertinentes à l'implication (i.e. les trois rangées vraies de la table de vérité du connecteur). Ces trois compétences de base sont déclinées dans trois situations différentes, soit le raisonnement avec du contenu familier, contrefactuel et abstrait. Les nœuds au bas de la structure représentent les types d'items qui seront sélectionnés par le système, soit les quatre formes logiques du raisonnement conditionnel. Nous exposons ci-bas les différentes

portions de la structure du RB pour le raisonnement conditionnel. La Figure 2.2 montre le haut de la structure. Le premier nœud représente la compétence globale. La structure montre des liens de dépendance entre ce dernier et les trois nœuds qui le succèdent.

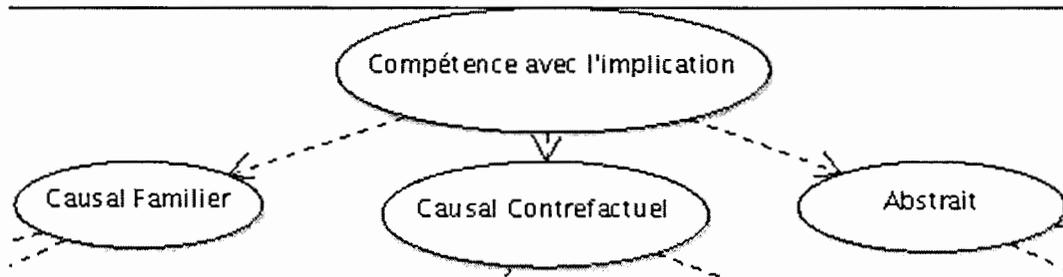


Figure 2.2 Tête du RB pour le raisonnement conditionnel

En le lisant de gauche à droite, la première branche du RB expose les compétences pertinentes au raisonnement conditionnel avec du contenu de plus en plus difficile. Le premier nœud enfant du nœud racine « Compétence avec l'implication » est donc « Causal Familier », suivi de « Causal contrefactuel » et de « Abstrait ». Nous exposons ci-bas chacune des trois structures qui concernent ces niveaux de contenu. La Figure 2.3 expose les compétences pertinentes au raisonnement avec du contenu familial.

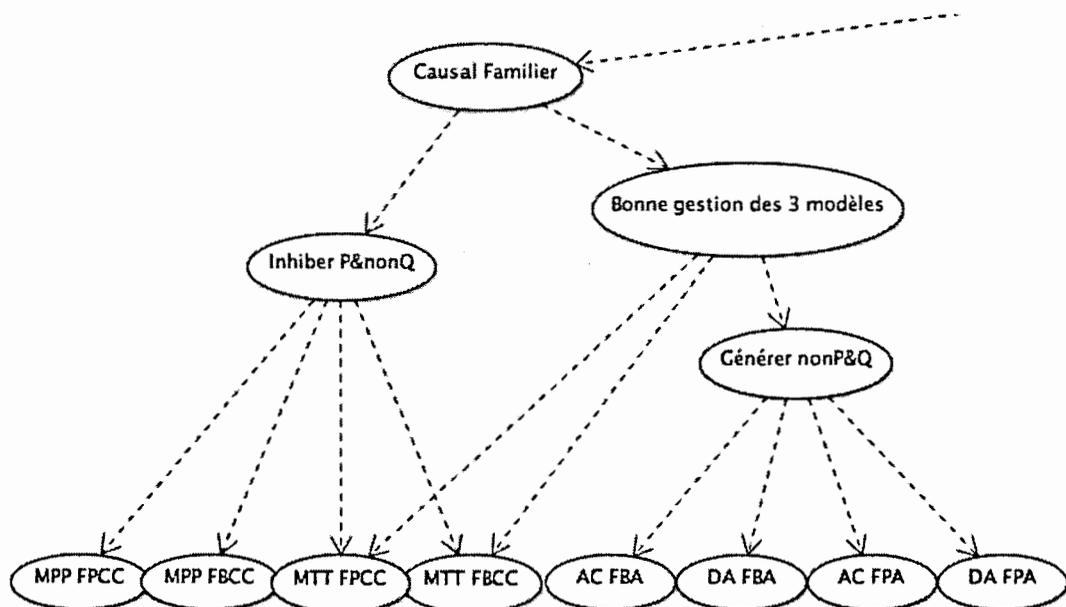


Figure 2.3 Portion du RB pour le raisonnement familial

Dans cette structure, le nœud « Inhiber P&nonQ » représente la capacité d'inhiber les conditions contraignantes de la prémisse conditionnelle, nécessaire pour reconnaître la certitude des inférences valides *MPP* et *MTT*. Les nœuds enfants de ce dernier sont donc les nœuds qui représentent le *MPP* avec peu (*MPP FPCC*) et beaucoup (*MPP FBCC*) de conditions contraignantes ainsi que le *MTT* avec peu (*MTT FPCC*) et beaucoup (*MTT FBCC*) de conditions contraignantes.

De plus, le nœud « Bonne gestion des 3 modèles » représente une compétence pertinente aux raisonnements *MTT*, AC et DA. Cette compétence consiste à garder en mémoire de travail les trois modèles pertinents à une prémisse conditionnelle ou, autrement dit, les trois rangées vraies du connecteur d'implication. Le *MPP*, de par sa simplicité et de par le fait qu'il peut être réussi même avec une compréhension simplifiée de l'implication, n'est pas un bon indicateur de cette compétence avancée. Nous avons donc déterminé qu'une capacité à répondre correctement aux trois inférences plus complexes (*MTT*, AC et DA) était le meilleur indicateur de cette compétence.

Le nœud « Générer nonP&Q » représente la capacité de générer les antécédents alternatifs qui serviront de contre-exemples aux conclusions invitées des inférences AC et DA. On retrouve donc parmi ses nœuds enfants ceux qui représentent la compétence avec le AC et le DA avec beaucoup (AC FBA et DA FBA respectivement) et peu (AC FPA et DA FBA respectivement) d'antécédents alternatifs. Le Tableau 2.2 résume la signification de chacun de ces nœuds et donne un exemple pour chacun d'eux.

Finalement, les relations de dépendance entre les nœuds de compétences exposés permettront au système de déterminer les probabilités de maîtrise des compétences à partir du patron de réponse de l'apprenant. Plus précisément, un ensemble de bonnes réponses aux items *MPP* et *MTT* augmentera la probabilité de maîtriser l'inhibition des conditions contraignantes. Des bonnes réponses aux items AC et DA augmenteront la probabilité de maîtriser la génération des antécédents alternatifs. De plus, un ensemble de bonnes réponses aux *MTT*, AC et DA augmentera la probabilité d'une bonne gestion des trois modèles pertinents à l'implication. L'évaluation de toutes ces compétences aura un impact

sur la probabilité de maîtrise du raisonnement conditionnel avec du contenu familier, ce qui, finalement, aura un impact sur la probabilité de maîtriser le raisonnement conditionnel dans son ensemble.

Tableau 2.2 Légende pour les nœuds items familiaux

Noeud	Signification	Exemple
<i>MPP</i> FPCC	<i>Modus Ponendo Ponens</i> Familier Peu de Conditions Contraignantes	Si Sébastien saute dans la piscine, alors il sera mouillé. Sébastien saute dans la piscine. Donc, Sébastien est mouillé.
<i>MPP</i> FBCC	<i>Modus Ponendo Ponens</i> Familier Beaucoup de Conditions Contraignantes	Si la clé de contact a été tournée, alors la voiture démarrera. La clé de contact a été tournée. Donc, la voiture démarrera.
<i>MTT</i> FPCC	<i>Modus Tollendo Tollens</i> Familier Peu de Conditions Contraignantes	Si Sébastien saute dans la piscine, alors il sera mouillé. Sébastien n'est pas mouillé. Donc, Sébastien n'a pas sauté dans la piscine
<i>MTT</i> FBCC	<i>Modus Tollendo Tollens</i> Familier Beaucoup de Conditions Contraignantes	Si la clé de contact a été tournée, alors la voiture démarrera. La voiture n'est pas démarrée. Donc, la clé de contact n'a pas été tournée.
AC FBA	Affirmation du Conséquent Familier Beaucoup d'Alternatives	Si on lance une roche dans une fenêtre, alors la fenêtre brisera. La fenêtre est brisée. Donc, on a lancé une roche dans la fenêtre.
DA FBA	Déni de l'Antécédent Familier Beaucoup d'Alternatives	Si on lance une roche dans une fenêtre, alors la fenêtre brisera. On n'a pas lancé de roche dans la fenêtre. Donc, la fenêtre n'est pas brisée.
AC FPA	Affirmation du Conséquent Familier Peu d'Alternatives	Si un chien a des puces, alors il se gratte constamment. Fido se gratte constamment. Donc, Fido a des puces.

DA	Déni de l'Antécédent Familier	Si un chien a des puces, alors il se gratte constamment.
FPA	Peu d'Alternatives	Fido n'a pas de puces.
		Donc, Fido ne se gratte pas constamment.

La Figure 2.4 expose la portion du RB relative au raisonnement avec du contenu contrefactuel. Les nœuds ainsi que les relations de dépendance entre eux sont les mêmes que dans la structure relative au raisonnement avec du contenu familier : la structure réfère aux mêmes processus de raisonnement et aux mêmes compétences visées. La différence est au niveau des nœuds qui représentent les formes logiques. Notons que la distinction entre peu et beaucoup d'alternatives ou de conditions contraignantes n'est plus présente dans cette structure. En effet, la quantité d'antécédents alternatifs et de conditions contraignantes avec des prémisses contrefactuelles est difficilement mesurable et aucune étude n'investigue cette dimension dans la littérature actuelle. Nous nous sommes donc restreints à des nœuds qui représentent la compétence au raisonnement avec les quatre formes logiques avec du contenu contrefactuel, sans distinction supplémentaire.

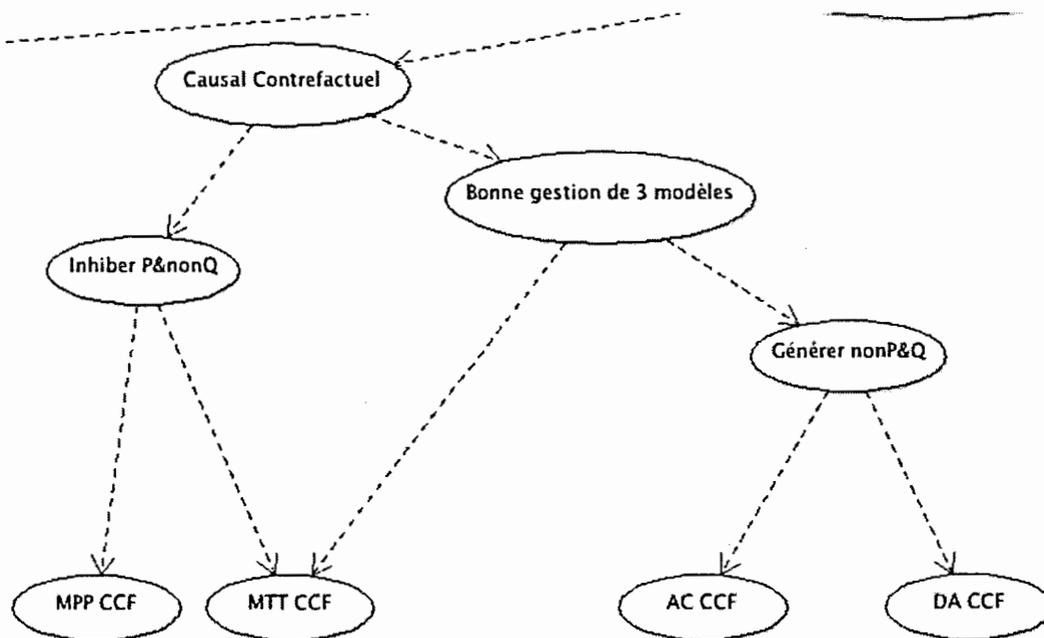


Figure 2.4 Portion du RB pour le raisonnement causal contrefactuel

Le Tableau 2.3 résume la signification de chacun de ces nœuds et est appuyé par des exemples.

Tableau 2.3 Légende pour items causaux contrefactuels

Noeud	Signification		Exemple
<i>MPP</i> CCF	<i>Modus Ponendo</i> Contrefactuel	<i>Ponens</i>	Si on lance du ketchup sur une chemise, alors elle deviendra propre. On a lancé du ketchup sur une chemise. Donc, la chemise est propre.
<i>MTT</i> CCF	<i>Modus Tollendo</i> Contrefactuel	<i>Tollens</i>	Si Louise se brosse les dents régulièrement, alors elle aura des caries. Louise n'a pas de caries. Donc, Louise ne se brosse pas les dents régulièrement.
AC CCF	Affirmation du Conséquent Contrefactuel		Si on lance du ketchup sur une chemise, alors elle deviendra propre. Une chemise est propre. Donc, on a lancé du ketchup sur la chemise.
DA CCF	Déni de l'Antécédent Contrefactuel		Si Louise se brosse les dents régulièrement, alors elle aura des caries. Louise ne se brosse pas les dents régulièrement. Donc, Louise n'a pas de caries.

La Figure 2.5 expose la dernière portion du RB, soit le raisonnement avec du contenu abstrait. Les nœuds « Inhiber P&nonQ », « Bonne gestion des 3 modèles » et « Générer nonP&Q » sont présents, entretiennent les mêmes relations de dépendance et réfèrent à des processus de raisonnement semblables à ceux des structures précédentes. Les nœuds « *MPP A* », « *MTT A* », « *AC A* » et « *DA A* » réfèrent aux quatre formes logiques avec du contenu abstrait. Notons que, comme dans le cas du raisonnement contrefactuel, la distinction quant au nombre d'antécédents alternatifs et de conditions contraignantes n'est

pas présente. La raison est toutefois différente. À ce niveau de difficulté du raisonnement conditionnel, le raisonneur doit comprendre que, pour toute prémisse conditionnelle, un antécédent alternatif peut être généré, et ce, indépendamment de son bagage de connaissances. La génération des alternatives, l'inhibition des conditions contraignantes et la gestion des trois modèles demandent donc, à ce niveau, une compréhension plus abstraite que dans les situations précédentes. Le Tableau X résume la signification des nœuds qui représentent les quatre formes logiques avec du contenu abstrait et est appuyé par des exemples.

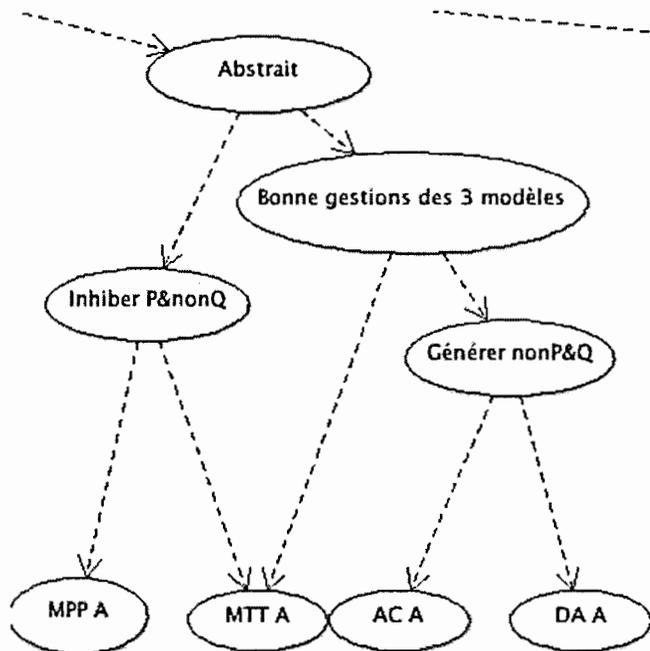


Figure 2.5 Portion du RB pour le raisonnement abstrait

Tableau 2.4 Légende pour items abstraits

Noeud	Signification	Exemple
<i>MPP A</i>	<i>Modus Ponendo Ponens</i> Abstrait	Si une personne morp, alors elle deviendra plede. Pierre morp. Donc, Pierre deviendra plede.
<i>MTT A</i>	<i>Modus Tollendo Tollens</i> Abstrait	Si une personne morp, alors elle deviendra plede. Pierre n'est pas devenu plede.

AC A	Affirmation du Conséquent Abstrait	Donc, Pierre ne morp pas. Si une personne morp, alors elle deviendra plede. Pierre est devenu plede
DA A	Déni de l'Antécédent Abstrait	Donc, Pierre morp. Si une personne morp, alors elle deviendra plede. Pierre ne morp pas. Pierre n'est pas devenu plede.

Enfin, la structure globale du RB prend en compte la hiérarchie entre les compétences appartenant aux trois niveaux de contenu. Ainsi, les compétences relatives au raisonnement abstrait demandent une certaine maîtrise des compétences relatives au raisonnement contrefactuel, qui elles-mêmes demandent une certaine maîtrise des compétences relatives au raisonnement familier.

2.2.1.1 Initialisation du RB, CDM et retour sur le modèle cognitif sous-jacent

Dans l'article qui suit, nous présentons une collecte de données ainsi qu'un *Cognitive Diagnosis Model* (CDM) qui permettront d'initialiser notre RB ainsi que de fournir une première évaluation empirique du modèle cognitif sur lequel il repose. Cet article a été soumis pour publication à la revue *Applied Measurement in Education*.

2.3 Abstract

In our previous works, we have presented Logic-Muse, an ITS that helps improve logical reasoning skills in multiple situations (REFs). In this paper, we will present Logic-Muse's learner model built as a Bayesian Network. We will report data mining techniques from the CDM package that were applied on problem-solving data from human participants. We will show how we will use the results in order to initiate and validate our Bayesian network. Relevant findings regarding our BN's underlying psychological model will be discussed.

2.4 Introduction

Decades of research in cognitive science show that human reasoning does not function accordingly to the rules of formal logic (e.g. Cummins, Lubart, Alksnis, & Rist, 1991; Gilovich, Griffin & Kahneman, 2002; Markovits & Vachon, 1990; Thompson, 1994). When looking for solutions to improve human skills in this area, a number of questions arise: what's important in the assessment of logical competence? What are the phenomena involved in the acquisition of logical reasoning skills? What should be the characteristics of a smart tutoring system (ITS) designed to support this learning? These questions cannot be answered without an appropriate understanding of human reasoning processes and the active participation of relevant experts, including logicians, psychologists, education professionals and IT specialists. By bringing together specialists from these different fields, the Logic-Muse project aims to study the basics of logical reasoning skills acquisition, to understand the difficulties associated with this learning and to create an ITS that can detect, diagnose and correct reasoning errors in various situations. The architecture of logic-Muse looks like a classical ITS with its three usual model components: the logical reasoning expert, the learner model and the tutor. The participatory approach adopted led to the development of basic components useful for building the Logic-Muse ITS. These components include the catalog of errors, structures and meta-structures of logical reasoning, and remedial strategies that promote learning. These have been previously elicited, validated and argued by the experts. Logic-Muse in its current version implements these components for propositional logic and offers a numerous activities allowing a learner to develop his logical skills in several situations established by the experts.

This paper aims at evaluating the learner's model of Logic-Muse. More precisely, it uses various data mining techniques on problem-solving data from human participants in order to initiate and validate the ITS learner model.

2.5 Logic-Muse's learner model

Logic Muse's learner model has several dimensions. The episodic memory keeps track of the exercises performed by the learner as well as all related performance elements. The cognitive model is a bayesian network (BN) whose nodes are the 96 units of knowledge related to reasoning, as identified by the knowledge domain experts. The prior probabilities as well as the relations of influence between the different units of knowledge were also established beforehand by the experts. The nodes are directly connected to the reasoning activities (items which correspond to the leaves nodes of the BN). The skills involved in the Bayesian network includes the inhibition of exceptions to the premises, the generation of counterexamples to the conclusion and the ability to manage all the relevant models for familiar, counterfactual and abstract situations (Markovits, 2013). The system's estimate of student skill acquisition is continually updated every time the student responds to a problem. That answer is then used as evidence by the system to re-compute the probability that this student knew the skill before the answer. Exercises are then chosen by the tutor based on these probabilities. The episodic memory keeps track of all the exercises performed by the learner. A Q-Matrix (see Appendix A) and a Bayesian network structure (See Appendix B) highlight the relations between the items (48 grouped into 16 classes in our case) and the knowledge units or cognitive skills (9 classes) (See appendix C for the complete item set).

2.6 Item bank for conditional reasoning

The item used for conditional reasoning in Logic-Muse are the four logical forms of conditional reasoning: the *Modus Ponendo Ponens* (MPP), the *Modus Tollendo Tollens* (MTT), the Affirmation of the Consequent (AC) and the Denial of the Antecedent (DA).

The reason why we chose conditional of reasoning as a starting point for the ITS is that many experimental studies in psychology have shown the great variability of human reasoning with inferences of the same logical form but differing in content (e.g. Cummins, Lubart, Alksnis, & Rist, 1991; Markovits & Vachon, 1990; Thompson, 1994). In related studies, much focus has been placed on the effects of content on conditional (if-then) reasoning. According to the logical definition of the conditional connective, two of these

inferences are valid and two are invalid. The *Modus Ponens* inference (“If P then Q, P is true, therefore Q is true”), referred to as MP and the *Modus Tollens* inference (“If P then Q, Q is false, therefore P is false”) referred to as MT are both valid and lead to necessary conclusions. By contrast, the Affirmation of the Consequent inference (“If P then Q, Q is true, therefore P is true”) referred to as AC, and the Denial of the Antecedent inference (“If P then Q, P is false, therefore Q is false”) referred to as DA) are both invalid since their putative conclusion doesn’t necessarily follow from the premises.

Each of these four logical forms are declined in 3 levels of content based on a developmental model of conditional reasoning. Markovits (2013) suggests that the more a premise has familiar content, the more rapid and easy the retrieval of a counter-example to invalid conclusions will be. Consequently, the more familiar a premise is, the more reasoners will avoid reasoning errors. In this sense, Markovits has developed a graduated scale of content levels for a premise that grows in difficulty and level of abstraction. The transition between two levels of content is gradual and done in stages: when a level is mastered, the skill in the next level begins.

The familiar level contains meaningful premises representing a plausible rule linking two known entities or categories. Content effects with these types of inferences have been heavily documented. Investigations of content related variability can fall under a general perspective called the “semantic memory framework” (De Neys, Schaeken & d’Ydewalle, 2002) where the retrieval of stored knowledge largely impacts reasoning with meaningful premises. The impact of information retrieval on conditional reasoning has been mostly observed through the effect of potential counterexamples to a putative conclusion. For the AC and DA inferences, such counterexamples are alternative antecedents, i.e. antecedents that differ from P but imply the consequent Q. For the MPP and MTT inferences, counterexamples are disabling conditions, i.e. a condition that prevents the antecedent P from implying the consequent Q. Many studies have shown that the number of potential counterexamples (Cummins, 1995; Cummins et al. 1991; Thompson, 1994) or the strength of association between them and the premise (De Neys, Schaeken & d’Ydewalle, 2003; Quinn & Markovits, 1998) determines approval rate of the four forms of conditional inference. For example, with the premise “If a rock is thrown at a window, then the window

will break”, reasoners will tend to accept the AC inference (a window is broken, therefore a rock was thrown at it) less often than with the premise “If a finger is cut, then it will bleed” (a finger bleeds, therefore it has been cut). The reason is that the former premise contains many alternative antecedents, like throwing a chair, a car accident, a tropical storm, etc., that are counterexamples to the putative conclusion, while the latter contains fewer of such antecedents (a finger is crushed, etc.).

The counterfactual level contains a reversed causal rule known to be false. It allows the generation of an unrealistic category of alternative antecedents. For example, with the counterfactual premise “If a feather is thrown at a window, then the window will break”, one could generate a counterfactual alternative antecedent like “throwing tissue on a window” or a disabling condition like “The window was strong enough to stay intact” .

The abstract level contains if-then rules linking made-up words, e.g. “If one blops, then one will become plede”. This level requires an abstract representation of the premises that can generate abstract alternative antecedents. According to Markovits (2013), when a reasoner reaches this level, he has a complete understanding of the implicative link: he understands that for a conditional premise (unknown or abstract), an alternative antecedent can be generated regardless of background knowledge.

Logic-Muse’s item bank for conditional reasoning thus consists of a total of 16 item classes (item nodes in the BN).

2.7 Validation and RB initialization.

Logic-Muse’s learner model aims to represent the user’s skills as accurately as possible using the Bayesian network we created. It allows for diagnosis and modelling of the learner’s current state of mastery for each identified skill. Validity and reliability are thus very important features of this model. In this paper, we will provide data to initiate and evaluate our model’s validity. To this end, we will use the R library CDM (George et al., 2016; Ravand & Robitzsch, 2015) to run a Cognitive Diagnosis Model on input data previously collected. This CDM model is built using the item bank, a Q-Matrix

(items/skills) designed by experts, as well as data collected from all student responses to items (De La Torre, 2009). The Q-Matrix (See Appendix A) connects items categories to the involved skills. The resulting model is part of the learner model as well and allows for initial predictions of the learner's state of mastery given his response vector.

2.7.1 Data collection.

2.7.1.1 Participants, procedure and materials.

A total of 294 participants were recruited online via the Prolific Academic platform. For each of the 16 items classes, three items were constructed. This was done in order to obtain a more reliable measure for each competence class. We then created three items blocks according to our three levels of content (Familiar, Counterfactual, Abstract). These were presented to participants in the following order: A block of familiar items was presented first, followed by the counterfactual and abstract blocks that were presented randomly. Within each block, major premises and item subcategories (few and many alternatives; few and many disablers) were presented randomly.

At the beginning of the questionnaire, participants first received the following instructions:

“In the next pages, we will present rules that you **must suppose to be true**. Following the rule, an observation will be presented to you. Your task is to select the conclusion that logically follows from the information provided. Here is an example of a question that you will have to answer:

Suppose it is true that: If we touch something that is very hot, then we burn ourselves.
Observation: Marco touched something that is very hot.

One can conclude that:

1. Marco burned himself.
2. Marco did not burn himself.

3. One cannot conclude whether Marco burned himself or not.

Please indicate below that you have read the instructions. Then press the "Next" button at the bottom right to start the questionnaire.”

They were then presented with the 48 items just like the example above.

2.7.2 Results.

2.7.2.1 Data Preparation

In order to use the CDM library, we prepared the raw data obtained. For each of the 48 items, participants had to choose between 3 answers (the valid one, the invalid typical one and the invalid atypical one). Answers were encoded as “1” for valid and “0 otherwise. We then had to choose between three possible response matrices according to the number of valid responses for the three repeated measures for each of the 16 categories. Each category could have been encoded as 1 if the participant was successful for at least 1 out of 3, at least 2 out of 3 or 3 out of 3 items. We opted for the response matrix with at least 2 successful responses out of 3. This threshold was chosen out of consistency with previous modelling choices: a majority of successful responses is the criteria to activate a competence node in our Bayesian Network.

2.7.2.2 Choice of model: DINA

We then conducted a CDM model on the 294 response patterns, which allowed us to estimate various parameters such as posterior probabilities, goodness of fit indicator, guess (the probability that a learner could correctly answer an exercise without having the necessary skills), slip (the probability of an erroneous response while the learner has the necessary skills), tetrachoric correlations, and marginal skill probabilities. We selected the Deterministic Input, Noisy “And” gate (DINA) CDM model since it makes the same assumption we made in our modelling of skills: the learner must have mastered all the

related skills in order to be successful to an item. The closely related Deterministic Input, Noisy “Or” gate (DINO) model rather assumes that at least one of the related skills needs to be mastered in order to succeed to an item, and was thus deemed inappropriate for our study.

2.7.3 BN initialization through the CDM model.

2.7.3.1 A posteriori probabilities.

The CDM allows us to predict a user’s probability of mastering the overall competence (root node) via its pre-test results. For this, we use the *a posteriori* probabilities obtained. To do so, from a learner’s vector of competence, we seek the line of the *a posteriori* matrix containing the same vector or a similar one. The joint probability matching this pattern, calculated based on the probabilities associated with each skill, is then used as the a priori likelihood (prior probability) of mastering the root node.

The first matrix obtained indicated that the most probable vector was the “111 000 000”, which means a learner masters the 3 competences for the familiar level of content, but no competence for the counterfactual and the abstract levels. This seems to show a separation between the familiar content and the other two levels. However, given the very high number of possible combinations for these 9 competences, the other vectors were very numerous, showed very small probabilities and a lot of them were equiprobable. We thus decided to use vector classes based on the types of competences identified in our model. The *a posteriori* matrix based on our classification is shown in the following Table 2.5. The left column represents the 6 vector classes we created. Each level of content (familiar, counterfactual, abstract) represents a triplet of skills (inhibit, generate, and manage) for this level, while each skill (inhibit, generate, manage) refers to the same skill in all three levels (familiar, counterfactual and abstract). In the first row, “successful” means all three skills for the corresponding level (or all three levels for the corresponding skill) are mastered, regardless of the performance in the other two levels or skills. For example, being successful in the familiar class may refer to vectors such as “111 111 111”, “111 011 111”,

“111 011 011” or even “111 000 000”. In the second row, “failed” means the opposite: all three skills for the corresponding level (or all three levels for the corresponding skill) are not mastered, irrespective of a learner’s performance in the other two levels or skills. The “only successful” row means that, of all three skills or levels, only the corresponding one is fully (111) mastered. For example, being only successful in the familiar class may come from the vector “111 011 011” or even “111 000 000”, but not “111 111 010” or “111 000 111”. The “Only failed” row refers to the opposite situation: only the corresponding skill or level triple is failed (at least one zero in the triplet), while the two other triplets are fully mastered (111).

Table 2.5 Skill class chunk probabilities

		Skill Class Chunk Probabilities			
		Successful	Failed	OnlySuccessful	OnlyFailed
	Familiar	0,95	0,05	0,55	0,00
Causal	CounterFactual	0,30	0,70	0,00	0,10
	Abstract	0,37	0,63	0,01	0,03
	Inhibit	0,39	0,61	0,07	0,01
	Generate	0,40	0,60	0,10	0,03
	Manage	0,41	0,59	0,10	0,02

Results for the different difficulty levels point to a graduation in performance. As can be expected, results show that the “Familiar” level stands out in both “Successful” and “Only Successful” categories: this level is seldom failed and is usually the only one mastered by learners. Contrary to what was predicted by our psychological model, the ‘Counterfactual’ level, not the Abstract level, is the most often failed and exclusively failed (‘Only failed’) by learners. While the difference in mastery level is not as pronounced as in the ‘Familiar’ case, it is nonetheless noteworthy, as both establish clear upper and lower boundaries for learning and performance. For instance, while the probability that learners exclusively fail the ‘Familiar’ case is close to zero, the same goes for the probability that learners exclusively master the ‘CounterFactual’ level. As for the in-between ‘Abstract’ level, results show that it is nonetheless a difficult level to master, closer to the ‘CounterFactual’ level. Finally, results for the different skills (‘Inhibit’, ‘Generate’, ‘Manage’) are inconclusive, as no clear skill pattern can be found across the different difficulty levels.

2.7.4 Validation for BN and underlying psychological model.

As an absolute indicator of goodness of fit for the CDM model, we opted for the item pairwise χ^2 measure (Chen, de la Torre & Zhang, 2013). This measure indicates that the model is inadequate if the p-value of the maximal item pairwise χ^2 measure is above the 0.05 significance level (Groß, Robitzsch, & George, 2015). In the present case, the χ^2 test results were $\chi^2 = 33.95$ with $p = 6.793316 \times 10^{-7}$, which is clearly below any significance threshold and thus indicates that the present model cannot be ruled out as inadequate.

2.7.4.1 Guess, slip & Item Discrimination Index (IDI).

Results for guess, slip, and item discrimination index parameters are shown in Table 2.6. First, we noted a high guess and low IDI for both MTT with familiar content. This can be explained by the fact that a biconditional interpretation of the major premise leads to the correct answer to the MTT, regardless of conditional skills (Markovits, 2013). It is also interesting to note that all items involving few alternatives have a lower IDI than their many alternatives counterparts. This goes accordingly with our psychological model: invalid reasoning problems with few alternatives are harder to reject, and thus require a better mastery of conditional reasoning. Beyond these remarks, however, all items have good IDI values. This is especially true for the MPP counterfactual, which suggests that successful completion of this task is the best way to ensure that conditional reasoning has been fully mastered. As it will be discussed below, this finding is consistent with results obtained for marginal skill probabilities.

Table 2.6 Guess, slip and IDI parameters

	Guess		Slip		Item Easin. (G+(1-S))/2	Item Discrim. 1-G-S
	Est.	SE	Est.	SE		
MPPFewDis	0,27	0,04	0,02	0,01	0,06	0,72
MPPManyDis	0,19	0,03	0,05	0,02	0,05	0,76
MPPCCF	0,00	0,00	0,05	0,02	0,04	0,95
MPPAbs	0,17	0,02	0,00	0,00	0,02	0,83
MTTFewDis	0,51	0,06	0,09	0,02	0,15	0,40
MTTManyDis	0,45	0,05	0,14	0,03	0,14	0,41
MTTCCF	0,07	0,01	0,19	0,03	0,14	0,74
MTTAbs	0,22	0,03	0,09	0,02	0,15	0,69
ACManyAlt	0,08	0,02	0,03	0,01	0,03	0,89
ACFewAlt	0,02	0,01	0,25	0,05	0,10	0,73
ACCCF	0,02	0,01	0,07	0,02	0,05	0,91
ACAbs	0,01	0,00	0,11	0,03	0,07	0,87
DAManyAlt	0,10	0,02	0,04	0,02	0,05	0,86
DAFewAlt	0,03	0,01	0,20	0,05	0,03	0,77
DACCF	0,11	0,02	0,04	0,02	0,15	0,85
DAAbs	0,05	0,02	0,04	0,02	0,12	0,91
α	0,14	0,02	0,09	0,02	0,08	0,77

2.7.4.2 Tetrachoric correlations.

Tetrachoric correlations between skills are shown in Table 2.7. Based on our sample size, correlation scores over 0.33 are considered significant with $\alpha = 0.05$ (Guilford & Lyons, 1942). Using this criterion for this analysis, one pattern that stands out is that skills with familiar content correlate highly with other skills of the same content level, but not with counterfactual and abstract level skills. However, counterfactual level skills correlate well with themselves and abstract level skills, and vice versa. Overall, these findings suggest a clear separation between the familiar level of content and other two levels.

Table 2.7 Skill correlations

Skill Correlations										
	InhibitFam	GenerateFam	ManageFam	InhibitCF	GenerateCF	ManageCF	InhibitAbs	GenerateAbs	ManageAbs	α
InhibitFam	1	0,85	0,9	-0,06	0,05	0,04	0,19	0,15	0,15	0,36
GenerateFam	0,85	1	0,78	0,05	0,06	0,16	0,2	0,16	0,17	0,38
ManageFam	0,9	0,78	1	0,22	0,06	0,04	0,2	0,15	0,16	0,39
InhibitCF	-0,06	0,05	0,22	1	0,64	0,6	0,67	0,5	0,52	0,46
GenerateCF	0,05	0,06	0,06	0,64	1	0,15	0,31	0,4	0,35	0,34
ManageCF	0,04	0,16	0,04	0,6	0,15	1	0,35	0,41	0,43	0,35
InhibitAbs	0,19	0,2	0,2	0,67	0,31	0,35	1	0,3	0,31	0,39
GenerateAbs	0,15	0,16	0,15	0,5	0,4	0,41	0,3	1	0,34	0,38
ManageAbs	0,15	0,17	0,16	0,52	0,35	0,43	0,31	0,34	1	0,38
α	0,36	0,38	0,39	0,46	0,34	0,35	0,39	0,38	0,38	0,38

2.7.4.3 Marginal skill probabilities.

Mean mastery percentage for each skill is shown in Table 2.8. The hardest skill to master seems to be « Inhibit counterfactual » (44.3%), which requires learners to inhibit disabling conditions to counterfactual conditional statements. This observation is consistent with the fact that the MPP counterfactual has the highest item discrimination index. Moreover, these findings seem at odds with our initial psychological model. Indeed, the latter considers abstract skills to be the hardest and least mastered ones; consequently, abstract skills should be better indicators of mastery for conditional reasoning than counterfactual skills, not the other way around. With hindsight, one could however argue that inhibiting exceptions to a rule known to be false, such as “If I throw ketchup on a shirt, then the shirt will be clean”, might prove harder than simply inhibiting an imaginary one. This might also be the reason why performances with counterfactual content are similar, or even sometimes worse, than performances with abstract content. Finally, here as in the previous analyses, the much higher probability of familiar level skills clearly separates them from the other two levels’ skills, as the difference between abstract level skills probabilities and their lower-scoring counterfactual counterpart is not nearly as pronounced.

Table 2.8 Skill Distribution

Estimate	Skill Distribution									
	InhibitFam	GenerateFam	ManageFam	InhibitCF	GenerateCF	ManageCF	InhibitAbs	GenerateAbs	ManageAbs	α
	0,072	0,056	0,076	0,4427	0,5512	0,5393	0,6762	0,6365	0,6428	0,71

2.8 Conclusion

In summary, analysis of CDM model data provided two major findings regarding the underlying developmental theory of our Bayesian Network. First, there seems to be a clear separation between the familiar level of content and the other two levels. Moreover, the hardest skill is not an abstract one, like our psychological model assumes, but a counterfactual one, i.e. the “inhibit counterfactual” skill. This surprising result, evidenced both by item discrimination indices and marginal skill probabilities, can be explained by the difficulty to inhibit exceptions, which are in fact realistic situations, to a false rule. Consequently, counterfactual level skills may very well prove harder to master than Abstract level counterparts.

2.9 Travaux futurs

Bien que le RB ait été validé par des experts du domaine d'apprentissage, ce dernier pourrait aussi être appris automatiquement à partir de données. Notamment, les connaissances expertes sur le raisonnement humain ne permettent pas de fournir une approximation exacte des tables de probabilités. Il serait donc avantageux d'avoir recours à des techniques d'apprentissage automatique pour la détermination de ces paramètres. Ce travail allant au-delà des limites de la présente thèse, il y est présenté comme faisant partie des travaux futurs pour le projet Muse-Logique. Nous exposerons les bases de l'apprentissage automatique pour les distributions de probabilités.

2.9.1 Apprentissage des distributions de probabilités : Nœuds observables

Lorsque la structure du RB est déterminée *a priori* et que tous ses nœuds ont des valeurs observables, l'apprentissage des distributions de probabilités se fait directement à partir des données, via le calcul du maximum de vraisemblance, du maximum *a posteriori* ou de l'espérance *a posteriori*. Le calcul du maximum de vraisemblance (MV) est une estimation statistique qui calcule la probabilité qu'un nœud X_i ait la valeur x_k étant donné que ses nœuds parents $pa(X_i)$ ont la valeur x_j . Cette probabilité est déterminée en calculant le nombre d'apparitions ($N_{i,j,k}$) d'une valeur pour un nœud lorsque ses nœuds parents ont une valeur quelconque, divisée par la somme des apparitions des valeurs que ce nœud peut avoir:

$$Pr(X_i = x_k | pa(X_i) = x_j) = \theta_{i,j,k} = \frac{N_{i,j,k}}{\sum_k N_{i,j,k}} \quad (2.7)$$

Le calcul du maximum *a posteriori* ou de l'espérance *a posteriori* sont des variantes de ce calcul (Naïm, Willemin, Lcray, Pourret & Bccker, 2007). Afin d'exemplifier ce calcul, imaginons un cas simple à deux variables : Un jeu de pile ou face avec deux pièces de monnaie pipées. Le lancer d'une pièce de monnaie (la variable *Lancer*) peut prendre deux valeurs (Pile, Face) et la pièce utilisée (la variable *Pièce*) peut aussi prendre deux valeurs (A, B). La Figure 2.6 expose un RB représentant la dépendance causale entre la pièce

utilisée et le résultat du lancer ainsi que la distribution de probabilité pour la variable *Lancer* :

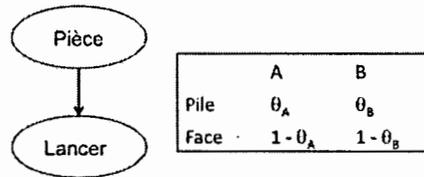


Figure 2.6 Réseau bayésien et distribution de probabilités : Pile ou face

Le but d'un tel réseau serait de prédire quelle pièce a été lancée (l'hypothèse) étant donné le côté sur lequel elle est tombée (l'évidence). Tel qu'exposé précédemment, la loi de Bayes requiert de connaître l'hypothèse de vraisemblance, soit la probabilité qu'une pièce tombe sur un côté donné (pile ou face) étant donné la pièce utilisée (A ou B). Cette probabilité est représentée dans la distribution jointe de probabilité conditionnelle et peut être apprise à partir de données. Si on détient une base de données d'apprentissage complète concernant le résultat du lancer et la pièce utilisée pour un ensemble de lancers, le calcul du maximum de vraisemblance est simple : on divise le nombre de fois où la pièce A est tombée sur pile par le nombre de lancers pour la pièce A, ce qui nous permet d'obtenir le paramètre θ_A , et on fait de même pour la pièce B. Supposons qu'on dispose de la base d'apprentissage exposée dans le Tableau 2.9, où 2 suites de 5 lancers ont été effectuées pour chaque pièce :

Tableau 2.9 Jeu de données : lancers de pièces de monnaie

A	Pile	Face	Pile	Face	Face
A	Pile	Pile	Face	Pile	Pile
B	Pile	Face	Pile	Face	Face
B	Face	Pile	Pile	Pile	Face

En utilisant la formule du maximum de vraisemblance, on obtient :

$$Pr(Lancer = pile | pa(Lancer) = A) = \theta_A = \frac{6}{6+4} = 0,6$$

pour l'estimation du paramètre θ_A , et on obtient :

$$Pr(Lancer = pile | pa(Lancer) = B) = \theta_B = \frac{5}{5+5} = 0,5$$

pour l'estimation du paramètre θ_B . Il serait aussi possible d'obtenir les paramètres pour la valeur face de la variable *Lancer* de la même manière, mais comme ces deux valeurs sont exhaustives et mutuellement exclusives, nous pouvons les obtenir en calculant $1 - \theta_A$ pour la pièce A et $1 - \theta_B$ pour la pièce B. Ainsi, la distribution locale de probabilité pour la variable *Lancer* devient la suivante :

Tableau 2.10 Distribution de probabilités : Variable Lancer

	A	B
Pile	0,6	0,5
Face	0,4	0,5

2.9.2 Apprentissage des distributions de probabilité : Nœuds cachés

Lorsque la structure d'un RB est déterminée, mais que certains de ses nœuds ne sont pas directement observables, c'est-à-dire qu'ils représentent une variable latente, l'apprentissage des paramètres est plus complexe. Supposons que, dans l'exemple précédent, la variable *Pièce* soit latente. Nous disposerions alors d'une base d'apprentissage où seuls les résultats des lancers sont connus. Pour obtenir une distribution de probabilité conditionnelle, il faudra estimer le maximum de vraisemblance à défaut de pouvoir l'obtenir directement et l'algorithme EM permet de faire une telle estimation. Après avoir fait une initialisation aléatoire des paramètres, l'algorithme EM permet de les modifier afin d'obtenir une estimation optimale. La première étape (l'étape E) consiste à calculer la probabilité de chaque complétion possible des données manquantes à partir des paramètres du réseau à l'itération t . Dans l'exemple précédent, il s'agirait de calculer la

probabilité que chaque série de lancers ait été faite avec la pièce A et de même pour la pièce B. Le système génère ensuite une nouvelle distribution de données, pondérée par les probabilités obtenues. Par exemple, supposons que la probabilité que la première série de lancers soit générée par A soit de 0.8 et de 0.2 pour B, il s'agirait de générer deux séries de lancers fictives en attribuant 80% des résultats à la pièce A et 20% des résultats à la pièce B. En répétant cette opération pour chaque série de lancers, on obtient la nouvelle distribution de données, où toutes les variables ont maintenant une valeur attribuée. La seconde étape (l'étape M) consiste à calculer de nouveaux paramètres à partir de cette distribution. Le calcul du maximum de vraisemblance peut alors être utilisé directement. Les étapes E et M sont répétées de manière itérative jusqu'à la convergence de l'estimation des paramètres. (Naïm & al., 2007 ; Do & Batzoglou, 2008).

2.9.3 Apprentissage des distributions de probabilité : Muse-Logique

Dans le RB que nous avons construit pour Muse-Logique, plusieurs nœuds représentent des compétences liées au raisonnement et non les problèmes de raisonnement qui seront présentés aux apprenants. Il s'agit donc de nœuds cachés. Il serait donc de mise d'entreprendre l'apprentissage des distributions de probabilités à partir de l'algorithme EM.

2.10 Conclusion

Dans ce chapitre, nous avons présenté le modèle de l'apprenant de Muse-Logique via un Réseau Bayésien basé sur les effets du contenu dans le raisonnement conditionnel. Ce type de raisonnement constitue l'une des trois grandes branches de la structure du réseau (conditionnel, incompatibilité et disjonction inclusive). Au prochain chapitre, nous présenterons les bases expérimentales de la construction de la branche du réseau qui concerne le raisonnement à partir de l'opérateur d'incompatibilité.

CHAPITRE III

LE RAISONNEMENT À PARTIR D'UNE INCOMPATIBILITÉ

Au chapitre précédent, nous avons exposé la partie de notre outil d'évaluation automatique qui concerne le raisonnement conditionnel. Comme mentionné en introduction, cet outil devra être prolongé afin d'évaluer le raisonnement à partir des connecteurs d'incompatibilité et de disjonction inclusive (Robert & Brisson, 2016). Cependant, alors que la littérature sur le raisonnement conditionnel est très abondante, à notre connaissance, aucune étude n'a exploré le raisonnement à partir d'une incompatibilité. Le but du présent chapitre est donc de combler ce manque afin de fournir une base empirique aux extensions du RB de Muse-Logique. Nous y exposerons donc nos études empiriques sur les inférences déductives à partir de l'opérateur d'incompatibilité. Bien que la pertinence du raisonnement conditionnel dans la vie de tous les jours soit bien établie dans la littérature, le raisonnement à partir d'une incompatibilité peut sembler artificiel et sa pertinence peut sembler restreinte aux classes de logique classique. Cependant, ces inférences sont pertinentes pour le raisonnement humain puisqu'elles constituent l'équivalent formel d'une manipulation du discours appelée sophisme du faux dilemme. La capacité de bien saisir ce qui est engendré par une situation d'incompatibilité permet donc de suspendre son jugement face à une fausse dichotomie. Cette compétence est donc très pertinente pour le raisonnement humain de la vie de tous les jours.

D'un point de vue formel, le sophisme du faux dilemme consiste à présenter une prémisse en tant que disjonction exclusive entre deux propositions alors qu'en réalité, les prémisses sont liées par une relation d'incompatibilité. Tel qu'énoncé précédemment, l'effet des contre-exemples sur le raisonnement est un effet de contenu largement rapporté dans la littérature sur le raisonnement conditionnel. Le but de nos expérimentations est d'investiguer de possibles effets similaires au sein du raisonnement à partir d'une incompatibilité, particulièrement au sein des inférences invalides *Deny First* et *Deny Second*.

Notre hypothèse générale est que plus la prémisse majeure permettra la récupération de contre-exemples/instances du dual, plus les participants reconnaîtront l'incertitude des inférences invalides. Dans une première étude, nous avons utilisé une loi logique de traduction entre opérateurs afin de transformer des inférences conditionnelles avec deux niveaux d'antécédents alternatifs en prémisses d'incompatibilité. Nous avons ensuite identifié trois classes de prémisses qui varient en fonction du contenu et les avons utilisées afin de construire trois classes de problèmes d'inférences qui contiennent de plus en plus de contre-exemples aux inférences invalides.

Les résultats de ces études nous ont permis de conclure que, comme avec le raisonnement conditionnel, le dual de l'opérateur d'incompatibilité ($\sim P \ \& \ \sim Q$) n'est pas toujours considéré par les raisonneurs et que cette tendance est directement liée au contenu des prémisses. Reasonner logiquement à partir de l'opérateur d'incompatibilité nous apparaît donc être un défi pour les raisonneurs humains, et ce, pour des raisons similaires au raisonnement conditionnel.

L'article qui suit expose les études que nous avons menées sur ce type de raisonnement. Cet article a été publié dans la revue *Memory and Cognition*.

3.1 Abstract

In the present studies, we investigated inferences from an incompatibility statement. Starting with two propositions that cannot be true at the same time, these inferences consist in deducing the falsity of one from the truth of the other or deducing the truth of one from the falsity of the other. Inferences of this latter form are relevant to human reasoning since they are the formal equivalent of a discourse manipulation called the false dilemma fallacy, often used in politics and advertising in order to force a choice between two selected options. Based on research on content related variability in conditional reasoning, we predicted that content would have an impact on how reasoners treat incompatibility inferences. Like conditional inferences, they present two invalid forms for which the logical response is one of uncertainty. We predicted that participants would endorse a smaller proportion of the invalid incompatibility inferences when more counterexamples are available. In study 1, we found the predicted pattern using causal premises translated into incompatibility statements with many and few counterexamples. In study 2, we replicated the content effects found in study 1, but with premises for which the incompatibility statement is a non-causal relation between classes. These results suggest that the tendency to fall into the false dilemma fallacy is modulated by the background knowledge of the reasoner. They also provide additional evidence on the link between semantic information retrieval and deduction.

3.2 Introduction

Deductive inference, that is, the ability to draw conclusions from given premises, is a key element of advanced human cognition. It requires generating a conclusion that follows “logically” from the premises, i.e. a conclusion that is necessarily true, provided that the premises were true. When a conclusion follows possibly but not necessarily from the premises, its truth value is uncertain even if the premises are true. Distinguishing between certain and uncertain conclusions is thus a key ability of deductive reasoning. Many logical mistakes consist in being certain about a conclusion that is in fact uncertain. Certain types of discourse manipulation, called informal fallacies, can lead an interlocutor to overlook possible conclusions so that she accepts a favored one as certain. For example, shortly after 9/11, when President Bush declared war on terror, he addressed the Congress by saying, “Either you are with us or you are with the terrorists.” This is an example of an informal fallacy, often found in politics or advertising, called the false dilemma or false dichotomy. This fallacy consists in reducing many options to only two dichotomous ones, thus forcing a choice between them (Hurley, 2014). One version of a false dilemma highlights the two ends of a continuum and discards all of the “in between” situations. Another version that will be the major focus of this paper, discards alternative options, i.e. situations where a third option is true and the two presented options are false. In the example above, the fallacy is of this latter version and is used as a persuasion technique. That is, it presents an option that one wants to be chosen by the interlocutor (to support the United States in their invasion of Iraq) and another one known as undesirable (to sympathize with terrorism), and lets the interlocutor reason, “I don’t support terrorism, therefore I must support this war.” Of course, alternative options which are neither one nor the other are possible: supporting economic, legal or educational interventions are all peaceful ways to stand against terrorism.

From a logical point of view, the false dilemma consists in falsely presenting a premise as an exclusive disjunction between two propositions. This can happen in two types of situation. On one hand, when the fallacy discards “in between” situations, the propositions are presented as being bound by an exclusive disjunction (Either P or Q), which allows

only for one proposition to be true and the other one false, while they are bound by an inclusive one (P or Q, or both), which allows for both propositions to be true (Van Eemeren & Grootendorst, 2016). Whether the basic interpretation of *or* is inclusive by default (Barrett & Stenner, 1971; Crain & Khlentzos, 2007; Grice, 1989; Kamp & Reyle, 1993; Newstead & Griggs, 1983) or depends on the semantic interpretation of the premise (Bauer & Johnson-Laird, 1993; Johnson-Laird & Byrne, 1991; Johnson-Laird, Byrne & Schaeken, 1992) is an open question. However, previous research has shown that disjunctive reasoning can be modulated by the content of the premises, with performances sometimes showing an exclusive interpretation of the disjunction (Quelhas & Johnson-Laird, 2017; Newstead, Griggs & Chrostowski, 1984 ; Roberge, 1977). These results thus suggest that reasoners might fall into this version of the fallacy and that this propensity might be affected by the content of the premises, but further studies are needed to investigate this question.

On the other hand, when the fallacy discards alternative options, the two propositions are in fact bound by the incompatibility connective (P is incompatible with Q), also called the NAND connective (Not both P and Q) or the Sheffer stroke. The incompatibility does not allow for both premises to be true at the same time, but is consistent with any other situation. Therefore, both the exclusive disjunction and the incompatibility connective are true when one proposition is false and the other one is true, but the incompatibility connective is also true when both propositions are false (Rautenberg, 2006). A logically equivalent way of describing this version of the false dilemma fallacy would be to say that the two propositions are presented as contradictories, i.e. an opposition where the propositions necessarily have opposite truth values, while in fact they are contraries, i.e. where the propositions could also be simultaneously false. Therefore, when two propositions are incompatible, the truth of one necessarily leads to the falsity of the other, but critically, the falsity of one does not necessarily lead to the truth of the other. In the example above, it would mean that if one supports terrorism, one is against the US war, but if one is against the war, one may or may not be on the terrorists' side. Being able to fully understand what is entailed by a situation of incompatibility may help to suspend judgment when a situation may seem like a dichotomy, but in reality is not. This ability is

an important part of human reasoning in daily life, and will accordingly be the major focus of the following studies.

3.3 Deductive reasoning from an incompatibility statement

While many inferential patterns can be generated from the same logical connective, deductive reasoning has most often been investigated through four particular ones. These inferences present two premises and a conclusion. The major premise bears a binary connective between two propositions, the minor premise affirms or denies one of these propositions and the conclusion affirms or denies the other proposition involved in the major premise. Following this pattern, deductive reasoning from an incompatibility statement can be investigated through inferences where a major premise presents two incompatible propositions, combined with one of four possible minor premises that affirms or denies one of the incompatible propositions. Within the four inferential modes generated, two are logically valid and two are invalid. The two valid ones affirm the first or the second proposition in the minor premise. We will refer to them as the Affirm First and the Affirm Second inference forms. The Affirm First inference involves reasoning with the premises “P is incompatible with Q, P is true” and leads to the logically correct conclusion, “Q is false.” (e.g. “Being in Montreal is incompatible with being in Paris, John is in Montreal. Therefore, John is not in Paris”). The Affirm Second inference involves the premises “P is incompatible with Q, Q is true” and leads to the logically correct conclusion, “P is false” (“Being in Montreal is incompatible with being in Paris, John is in Paris. Therefore, John is not in Montreal.”) The two invalid forms deny the first or the second proposition in the minor premise. We will refer to them as the Deny First and the Deny Second inferences. The Deny First inference involves reasoning with the premises “P is incompatible with Q, P is false.” Although these premises suggest the putative conclusion “Q is true”, this is a fallacy since the premises do not lead to a necessary conclusion. For example, the premises “Being in Montreal is incompatible with being in Paris, John is not in Montreal.” suggests that “John is in Paris.” But since John could be elsewhere, e.g. in Rome, one cannot be certain about the invited conclusion. The Deny Second inference

involves the premises “P is incompatible with Q, Q is false”, and also has no certain conclusion.

Thus, according to standard logic, the invited conclusions of the Affirm First and Affirm Second inferences should be endorsed as certain provided that the premises were true, while the invited conclusions of the Deny First and Deny Second inferences should be rejected. The false dilemma fallacy occurs when the invited conclusion for the Deny First and Deny Second inferences are incorrectly accepted as certain. What then determines when this fallacy will be produced? The key concept is the idea that for every pair of incompatible situations (P, Q), there exist others for which both P and Q are false. These situations are thus counterexamples to the false dilemma fallacy and are sometimes referred to as *Tertium Quid* or Third options, that is, an option that exposes a dichotomy as false. When an incompatibility is presented as an exclusive disjunction in a discourse manipulation, the goal is for these Third options to be discarded out of hand. However, the aim of the present studies is to investigate under what conditions they can be overlooked even when the major premise is presented explicitly as an incompatibility statement. Our basic hypothesis is that the effects of Third options to the false dilemma fallacy can be understood as a variant of the effect of background knowledge on reasoning that has been documented with respect to conditional reasoning. More specifically, we hypothesized that the more readily Third options can come to mind, the more reasoners will tend to avoid the false dilemma fallacy (for an exposition of the formal links between counterexamples on incompatibility and conditional inferences, see Robert & Brisson, 2016).

3.4 Content effects and conditional reasoning

Many studies have shown the great variability of human reasoning with inferences of the same logical form but differing in content (e.g. Cummins, Lubart, Alksnis, & Rist, 1991; Markovits & Vachon, 1990; Thompson, 1994) and a large part of the literature on human reasoning is an attempt to explain such variability. Much focus has been placed on the effects of content on conditional (if-then) reasoning. Such reasoning has been mostly investigated through the four inferential modes explained above. According to the logical

definition of the conditional connective, also called the material conditional, two of these inferences are valid and two are invalid. The *Modus Ponens* inference (If P then Q, P is true, therefore Q is true), referred to as MP and the *Modus Tollens* inference (If P then Q, Q is false, therefore P is false) referred to as MT are both valid and lead to necessary conclusions. By contrast, the Affirmation of the Consequent inference (If P then Q, Q is true, therefore P is true) referred to as AC, and the Denial of the Antecedent inference (If P then Q, P is false, therefore Q is false) referred to as DA, are both invalid since their putative conclusion doesn't follow necessarily from the premises.

Content effects with these four forms of inferences have been well documented. Content related variability can be understood under a general perspective called the "semantic memory framework" (De Neys, Schaeken & d'Ydewalle, 2002) where the retrieval of stored knowledge impacts reasoning with meaningful premises. The impact of information retrieval on conditional reasoning has been mostly observed through the effect of potential counterexamples to a putative conclusion. For the AC and DA inferences, such counterexamples are alternative antecedents, i.e. antecedents that differ from P but imply the consequent Q. For the MP and MT inferences, counterexamples are disabling conditions, i.e. a condition that prevents the antecedent P from implying the consequent Q. Many studies have shown that the number of potential counterexamples (Cummins, 1995; Cummins et al. 1991; Thompson, 1994) or the strength of association between them and the premise (De Neys, Schaeken & d'Ydewalle, 2003a; Quinn & Markovits, 1998) determines the rate of approval of the four forms of conditional inferences. For example, with the premise "If a rock is thrown at a window, then the window will break", reasoners will tend to accept the AC inference (a window is broken, therefore a rock was thrown at it) less often than with the premise "If a finger is cut, then it will bleed" (a finger bleeds, therefore it has been cut). The reason is that the former premise contains many alternative antecedents, such as throwing a chair, a car accident, a tropical storm, etc., that are counterexamples to the putative conclusion, while the latter contains fewer of such antecedents (a finger is crushed, etc.).

As explained above, the Deny First and Deny Second inferences are formal equivalents of the false dilemma fallacy. Just like the AC and DA inferences, these can have a variable

number of potential counterexamples, depending on the content of the major premise involved. Thus, based on the effects of content on conditional reasoning, we make the hypothesis that the more Third options are available in memory, the more reasoners will tend to be uncertain about the Deny First and Deny Second inferences. In other words, the more reasoners can think of situations outside of a dichotomy, the less likely they will be to fall into the false dilemma fallacy.

3.5 Study 1

In this study, we used premises based on a causal relation expressed as an incompatibility statement. We started with well-known causal conditional premises that have many and few alternative antecedents and translated them into incompatibility statements. For example, we started with the conditional premise “If one has a car accident, then one will be late for a meeting.” This premise has many alternative antecedents (getting up late, being stuck in traffic, helping a friend, etc.) which are counterexamples to the conclusion of the AC and DA inferences. We constructed the following incompatibility statement: “Having a car accident is incompatible with being on time at a meeting.” While presented in a different way, the possibilities that were alternative antecedents for the conditional are now Third options for the incompatibility statement, i.e. a situation where neither P nor Q are the case. This transformation was based on a law of classical logic that states the equivalence between a conditional statement and an incompatibility between the antecedent and the denied consequent, i.e. $(P \rightarrow Q) \equiv (P \uparrow \sim Q)$ (Sheffer, 1913). Simply put in terms of causality, this law expresses the idea that a causal relation is equivalent to one where you cannot have both the cause and the absence of its effect. From this law, we can also translate alternative antecedents into Third options and disablers into Exceptions to an incompatibility statement. Provided that an alternative antecedent is a situation where P is false and Q is true, the law indicates that this same situation would be one where both premises, i.e. P and Not-Q, are false. This is precisely the definition of a Third option to a false dilemma. Moreover, provided that a disabler is a situation where P is true and Q is false, the same reasoning leads to indicating a situation where both premises, i.e. P and Not-Q, are true as an Exception, which would indeed undermine the incompatibility

statement between these premises.

Additionally, to avoid the use of a negation in the incompatibility statements, we identified conditionals for which the consequent is a dichotomous term that can be denied implicitly. In the example above, it means that the “not being late for a meeting” was translated into “being on time for a meeting.” The same procedure was used for causal relations with few alternative antecedents.

The causal relations that were transformed into incompatibilities were chosen to have a variable number of alternative antecedents, and relatively few disablers. We first pretested a set of such incompatibility statements to ensure that the classification into many and few alternatives was maintained.

3.5.1 Pretest

A total of 22 participants took part in the pretest (8 men, 14 women, average age = 33 years, 6 months), were recruited via the online platform Crowdfunder and were native English speakers. Participants were imposed no time limit and took part in the experiment individually.

We constructed 12 incompatibility statements based on causal conditional premises, of which 6 were chosen to have few alternatives and 6 to have many alternatives, which was expected to translate into few or many Third options. Moreover, all statements were chosen to have relatively few disabling conditions, which translated into few exceptions to the incompatibility. Half of the participants generated Third options for each of 12 statements, presented in a random sequence. Statements were chosen to have few exceptions, and in order to verify this, the other half generated potential exceptions for the statements.

The Third option generation task was constructed based on the methodology used by Verschueren, Schaeken, De Neys & d'Ydewalle (2004). We presented an incompatibility statement followed by a situation where P and Q are false. We then asked participants to give as many explanations as possible for this situation, thus making them draw up a list of possible alternatives to the pair of incompatible propositions. Participants were given a

limit of five explanations to prevent artificial explanations and fatigue while allowing variation between premises. Below is an example of such a task:

“Rule: Having a car accident is incompatible with being on time at a meeting.

Situation: Paul didn’t have a car accident BUT he was not on time for his meeting.

List as many explanations for this situation as possible (5 maximum).”

The same procedure was used to construct the generation task for exceptions:

“Rule: Having a car accident is incompatible with being on time at a meeting.

Situation: Paul had a car accident BUT he was on time for his meeting.

List as many explanations for this situation as possible (5 maximum).”

Responses for each premise were rated by two judges who were instructed to ignore any paraphrase of an existing response. We then calculated the mean number of Third options and exceptions generated for each of the 12 incompatibility premises. From this, we selected three premises with the highest mean number of Third options and three with the lowest mean number of Third options generated, with all having an equally low level of exceptions. The former were put in the “Many” group and the latter were put in the “Few” group. The mean number of Third options generated in the Many group was significantly higher ($M = 3.78$; $SD = 1.36$) than the mean number of Third options generated in the Few group ($M = 2.47$; $SD = 1.24$), $t(11) = 6.306$, $p < 0.001$, $d = 1.00$. See Appendix D for the selected premises, the conditional version they were initially translated from and the mean numbers of Third options and exceptions for each of them.

Our main hypothesis is that the tendency to make the false dilemma fallacy is related to the number of readily available Third options. In this initial study, we used the causal premises identified in the pretest to construct inference problems with many and few Third options. We predicted that participants would produce more uncertainty responses to the Deny First and Deny Second forms for problems with many alternatives than for those with few alternatives.

3.5.2 Method

3.5.2.1 Participants and procedure

A total of 98 students from a college in Montreal took part in the experiment (60 women, 37 men (one participant failed to indicate gender), average age: 18 years, 0 months, age range: 16-21 years). Each participant was randomly allocated one of the booklets, was told to take as much time as needed and took part in the experiment individually.

3.5.2.2 Design and materials

Four paper and pencil booklets were constructed. The first one presented 3 sets of inferential problems based on the major premises with many counterexamples. For each major premise, participants received problems corresponding to the Affirm First, Affirm Second, Deny First and Deny Second inferences. The order of these was randomly determined for each major premise, which were, in order:

- “1. Having a car accident is incompatible with being on time at a meeting;
2. Being on a plane is incompatible with sleeping well;
3. Drinking coffee in the evening is incompatible with falling asleep easily.”

A second booklet presented 3 sets of inferential problems based on the premises with few counterexamples. For each major premise, we used the same inference order determined in the first booklet. The premises were, in order:

- “1. Drinking a lot of alcohol is incompatible with being sober;

2. Being in a warm environment is incompatible with feeling cold;

3. A displaced rail is incompatible with the train staying on track.”

For each of these two booklets, an alternative version was constructed with the order of the major premises inverted. On the first page of each booklet, participants were asked for their gender, age and grade level. They were then given the following instructions:

“In the following pages, we are going to show you some rules that you must **suppose to be true**. You have to assume that the rules are always true. For each rule, we are also going to show you some observations. Your task is to select the conclusion that follows logically from the rule and the given observation.”

On the top of each of the following pages, an incompatibility statement was presented. On the same page, four logical problems corresponding to the Affirm First, Affirm Second, Deny First and Deny Second inferences were presented. For each inference, participants had to choose amongst three possible conclusions. The following is an example of such a statement and a Deny First inference problem:

“Suppose that it is always true that:

Having a car accident is incompatible with being on time for a meeting.

For each of the following observations, select the conclusion that follows logically from the rule and the given observation:

Helen didn't have a car accident. One can conclude that:

1. Helen was on time for her meeting.
2. Helen was not in time for her meeting.
3. One cannot conclude whether or not Helen was on time for her meeting.”

In summary, Premise-type (*Many, Few*) was a between-subjects variable and Logical form (Affirm First, Affirm Second, Deny First, Deny Second) was a within-subject variable.

3.5.3 Results and Discussion

Preliminary analysis showed that some participants gave unexpected responses to either the valid or invalid forms, that is, they neither endorsed nor suppressed valid or invalid inferences, but rather endorsed the opposite of the invited conclusion, i.e. Affirm First: “Q is true”; Affirm Second: “P is true”; Deny First: “Q is false”; Deny Second: “P is false”. Of these, 3 participants produced close to or more than 3 standard deviations from the mean number of unexpected responses (3 or more responses out of 12). They were thus eliminated from further analysis. Of these 3, 2 were from the Many condition and 1 was from the Few condition. All further analyses were conducted on the 95 remaining participants (46 in the Many, 49 in the Few condition).

We then calculated the percentage of logically correct responses out of three inferences for each of the four logical forms (see Table 3.1). First, a 2 (Validity: Valid, Invalid) X 2 (Inference-type: First, Second) X 2 (Premise-type: Many, Few) X 2 (Order: First, Second) mixed-design ANOVA revealed no significant effect of Order, $F(1, 91) = 0.762, p = 0.39$. We then performed an ANOVA with Validity (Valid, Invalid) and Inference-type (First, Second) as repeated measures and Premise-type (Many, Few) as a between-subjects variable. This gave no main effect of Validity, $F(1, 93) = 0.017, p = 0.9$, Inference-type, $F(1, 93) = 3.598, p = 0.06$, nor Premise-type, $F(1, 93) = 0.218, p = 0.64$. The results also showed a significant interaction between Validity and Premise-type $F(1, 93) = 6.618, p < 0.05, partial\ eta^2 = 0.066$, a significant three-way interaction between Validity, Inference-type and Premise-type $F(1, 93) = 4.964, p < 0.05, partial\ eta^2 = 0.051$ and no significant interaction between Inference-type and Premise-type, $F(1, 93) = 2.694, p = 0.11$, nor between Validity and Inference-type, $F(1, 93) = 0.56, p = 0.46$.

Post-hoc comparisons used the Tukey procedure with $p = 0.05$. We first analyzed the Validity X Inference-type interaction. For the invalid forms, this showed that the number of logically correct responses was greater for the Many, $M = 0.70; SE = 0.06$, than for the Few premises, $M = 0.51; SE = 0.06$. The analysis of the three-way interaction showed that this difference was maintained between the Many, $M = 0.69; SE = 0.06$, and Few premises, $M = 0.52; SE = 0.05$, for the Deny First form as well as between the Many, $M = 0.70; SE$

= 0.06, and Few premises, $M = 0.49$; $SE = 0.06$, for the Deny Second form. For the valid forms, there was no significant difference between the number of logical responses for the Many and the Few premises. However, analysis of the three-way interaction revealed that the difference was significant for the Affirm Second form, $M = 0.48$; $SE = 0.06$; $M = 0.70$; $SE = 0.06$, but did not reach significance for the Affirm First form.

Table 3.1 Mean percentage of logically correct responses for the four logical forms (Deny First, Deny Second, Affirm First, Affirm Second) by Type (Many, Few) in Study 1 (standard deviations in parentheses)

Type	Deny First	Deny Second	Affirm First	Affirm Second
Many	0.69 (0.42)	0.70 (0.41)	0.59 (0.43)	0.48 (0.44)
Few	0.52 (0.33)	0.49 (0.41)	0.67 (0.33)	0.70 (0.36)

The results of this initial study are generally consistent with our hypothesis. They show that the tendency to accept the invited conclusion for the two invalid forms, and thus fall into the false dilemma fallacy, is directly linked to the relative number of Third options. When participants have access to fewer numbers of such options, they show an increased tendency to accept the invited conclusion when this is not merited.

The results also showed an unexpected content effect for the Affirm Second form, which was rejected more often for the Many premises than for the Few premises. This is surprising since both the Many and the Few premises were chosen to have equally low numbers of exceptions. The suppression effect literature on conditional reasoning indeed indicates that such content effects on valid inferences are primarily driven by disabling conditions rather than alternative antecedents (Byrne, 1989; 1991, Cummins, 1995, Cummins et al., 1991, Thompson, 1994). Accordingly, suppression of the valid incompatibility inferences should be driven by exceptions rather than third options. However, previous research on conditional reasoning has observed an indirect impact of alternatives on the suppression of

valid inferences. Some studies have shown that retrieval of alternative antecedents to conditional premises may induce retrieval of disabling conditions. Notably, the generation of alternative antecedents has been associated with more rejection of the valid MP and/or MT inferences in both children (Janveau-Brennan & Markovits, 1999) and adults (Markovits & Potvin, 2001; De Neys & al., 2002). Similarly, the retrieval of third options for the Many incompatibility premises might have facilitated the retrieval of exceptions. This explanation however remains open for debate since these suppression effects on valid inferences have not been replicated when alternative antecedents were provided explicitly to participants (De Neys, Schaeken & d'Ydewalle, 2003b). Nonetheless, the suppression effects found in this study were unclear and minor and did not affect support for our main hypothesis. Possible suppression effects on the valid incompatibility inferences thus fall outside the scope of this paper and should be investigated in further studies.

3.6 Study 2A

The results of the first study suggest that reasoning from incompatibility statements uses the same basic semantic retrieval processes that are responsible for content effects in conditional reasoning. However, the causal premises used in this study, while presented as incompatibilities, might have triggered a causal interpretation which in turn could have generated a retrieval process linked to the basic conditional relation. In that case, these results could be seen as a simple replication of the well-known content effects on conditional reasoning. We thus decided to replicate these results using premises that were not causal.

In order to do this, we started by constructing incompatibility statements using categories designed to produce variable numbers of Third options. We first constructed incompatibility statements of the form “X is A is incompatible with X is B”, where A and B are base-level categories derived from the same parent category. For example, dogs and horses are both animals. The corresponding incompatibility would be that “For an animal, being a dog is incompatible with being a horse.” We constructed an initial set using parent categories which allowed many possible base-level categories, which we refer to as Broad

statements. We then constructed a set of Reduced statements, where the Reduced parent was chosen to have relatively few base-level instances, e.g. polar animals. Finally, we constructed a set of Close to binary statements, for which the parent was an action category that allowed almost no other options, e.g. from the parent voting on a bill, which allows voting for or against or abstaining. (See APPENDIX E for the parent and base-level categories for all of the incompatibilities that were generated by type). It should also be noted that, while minimized, potential exceptions to the rule could be generated from the premises used in study 1. In this study, premises were constructed in such a way that the generation of exceptions would require the construction of imaginary possibilities, like an animal that is both a dog and a horse. These premises thus allowed for no realistic exceptions.

3.6.1 Pretest

Participants were asked to write as many Third options as they could for each of thirteen pairs of base level categories.

A total of 19 students (5 men, 14 women, average age = 21 years, 11 months) were examined for the pretest. All participants were native French speakers and volunteers and were students at the Université du Québec à Montréal. Each participant was randomly allocated one of the booklets, was told to take as much time as needed and took part in the experiment individually.

Two, fourteen pages, paper and pencil booklets were constructed. On the first page of each booklet, participants were asked for their gender, age and grade level. They were then given the following instructions (translated from the original French):

“In the following pages, we will ask you to provide as many elements as possible belonging to a given category. You must answer the questions in the presented order and you cannot go back to the questions you have previously answered.”

On the top of each of the following pages, a short context was presented into which a pair of subcategories belonging to a particular parent category was presented. Participants were

then asked to write down as many Third options as they could. The following is an example of such context for the “animal” category (translated from the original French):

“In an encyclopedia, we learn about animals. An animal can be a dog or a horse. On the following lines, make a list of all the animals that are neither a dog nor a horse. Write down everything that comes to mind.”

The rest of the page was filled with blank lines on which participants had to write down their answers.

On the first version of the booklet, questions were asked in the following order: four questions of the Close to binary content type, five questions of the Reduced type and four questions of the Broad type. A second version of the booklet was constructed in which the order of the questions was inverted.

Responses for each item were rated by two judges who were instructed to ignore any paraphrase of an existing response. We then calculated mean numbers of Third options generated for each item (see APPENDIX E). Inspection of these results showed that the German city question generated a very low mean number of Third options ($M=1.07$; $SD=1.16$), probably due to participants’ lack of knowledge. This left 4 items in each of the three categories, which were all used in Study 2.

We first examined differences in the total numbers of Third options generated as a function of Category type. We performed an ANOVA with the total number of Third options generated as a dependent variable with Category type (Broad, Reduced, Close to binary) as a repeated measure and Order as a between-subjects independent variable. This gave only a significant main effect of Category type, $F(2, 30) = 21.33$, $p < 0.001$, $partial\ eta^2 = 0.587$. Contrast analyses with Bonferroni corrections showed that, as expected, the total mean number of Third options generated for the Broad type ($M=44.76$; $SD=33.68$) was greater than the total number for the Reduced type ($M=11.71$; $SD=6.17$) ($p < 0.01$), which was in turn greater than the total number for the Close to binary type ($M=3.06$; $SD=2.59$) ($p < 0.001$).

The items retained after the pretest can thus be classified as a function of the relative

availability of Third options. If the retrieval model generalizes to these kinds of premises, then we can predict that uncertainty responses to the invalid forms should vary by category. Specifically, we predict that the number of uncertainty responses to the Deny First and Deny Second forms will be greater when premises come from the Broad category than when they come from the Reduced category, which in turn will be greater than for premises from the Close to binary category.

3.6.2 Method

3.6.2.1 Participants and procedure

A total of 237 students (97 men, 140 women, average age: 22 years, 8 months, age range: 17-52) participated in this study. All participants were native French speakers and volunteers and were recruited in colleges or universities in Montreal. Each participant was randomly allocated one of the booklets, was told to take as much time as needed and took part in the experiment individually.

3.6.2.2 Design and materials

Six paper and pencil booklets were constructed. Each booklet presented 4 sets of inferential problems, which consisted of the Affirm First, Affirm Second, Deny First and Deny Second inferences. Note that given that no effect of Order was found in Study 1, we kept this order for each problem set in all 6 booklets. Problem sets for the first booklet were based on major premises that come from the Broad category. The major premises were, in order (translated from the original French):

- “1. For a vegetable, being a broccoli is incompatible with being a pepper;
2. For an animal, being a dog is incompatible with being a horse;
3. For a person, being in Montreal is incompatible with being in Paris;

4. For a fruit, being a grape is incompatible with being a strawberry”

A second booklet presented 4 sets of inferential problems based on the premises from the Reduced category. These were, in order (translated from the original French):

- “1. For a root vegetable, being a potato is incompatible with being a carrot;
2. For a polar animal, being a polar bear is incompatible with being a penguin;
3. For a fruit with pit, being a peach is incompatible with being a cherry;
4. For a dessert of the day in a restaurant; being the chocolate cake is incompatible with being the lemon pie”

Finally, a third booklet presented 4 sets of inferential problems based on the premises from the Close to binary category. These were, in order (translated from the original French):

- “1. For a player at the “heads and tails” game, betting on tails is incompatible with betting on heads;
2. For a driver at a fork on the road, taking the right road is incompatible with taking the left road;
3. For a player at the “even and odd” game, betting on an even number is incompatible with betting on an odd number;
4. For a person who votes on a bill, voting for the bill is incompatible with voting against the bill”

For each of these three booklets, an alternative version was constructed where the order of the major premises was inverted. On the first page of each booklet, participants were asked for their gender, age and grade level. They were then given the following instructions (translated from the original French):

“In the following pages, you will be presented statements that you must **suppose to be true**. Your task is to select the conclusion that follows logically from the given statements.”

On the top of each of the following pages, a short context was presented into which the incompatibility was presented. On the same page, four logical problems were presented in the following order: Affirm First, Affirm Second, Deny First, Deny Second. For each problem, participants had to choose amongst three possible conclusions. The following is an example of such a context and an Affirm First inference problem in the Many Third options condition:

“In an encyclopedia, we learn about animals. An animal cannot be both a dog and a horse at the same time. In other words, being a dog is incompatible with being a horse.

Suppose it is always true that:

Being a dog is incompatible with being a horse.

An animal is a dog. One can conclude that:

1. This animal is a horse.
2. This animal is not a horse.
3. One cannot conclude if this animal is a horse or not.”

In summary, Premise-type (Broad, Reduced, Close to binary) is a between-subjects variable and logical form is a within-subjects variable.

3.6.3 Results and discussion

Preliminary analysis showed that seven participants failed to answer a majority of the problems. They were then eliminated from further analysis. With the remaining 230 participants, we calculated the percentage of logically correct responses (out of four inferences) for the four logical forms (see Table 3.2).

A 2 (Validity: Valid, Invalid) X 2 (Inference-type: First, Second) X 2 (Premise-type: Many, Few) X 2 (Order: First, Second) mixed-design ANOVA revealed no significant effect of Order, $F(1, 223) = 0.033, p = 0.86$. We then performed an ANOVA with Validity (Valid, Invalid) and Inference-type (First, Second) as repeated measures and Premise-type (Broad, Reduced, Close to binary) as a between-subjects variable. This gave main effects of Validity $F(1, 226) = 107.262, p < 0.001, partial\ eta^2 = 0.322$ and Premise-type, $F(2, 226) = 14.566, p < 0.001, partial\ eta^2 = 0.114$, and no main effect of Inference-type, $F(1, 226) = 0.049, p = 0.83$. The results also showed a significant interaction between Validity and Premise-type, $F(1, 226) = 107.262, p < 0.001, partial\ eta^2 = 0.322$ and no significant interaction between Inference-type and Validity, $F(1, 226) = 0.201, p = 0.65$, Inference-type and Premise-type, $F(1, 226) = 1.756, p = 0.18$, nor between Validity, Inference-type and Premise-type, $F(1, 226) = 0.775, p = 0.46$. This showed that participants had greater accuracy on the valid, $M=89.67, SE=1.42$, than on the invalid inferences, $M=60.88, SE=2.52$.

Post hoc comparisons used the Student-Newman-Keuls procedure with $p = 0.05$. Note that, when three means are compared, as is the case here, this procedure holds the familywise error rate to .05 (Howell, 2012). We first analysed the main effect of Premise-type. This showed that the number of logically correct responses was greater for the Broad premises, $M = 0.85; SE = 0.03$, than the Close to Binary premises, $M = 0.65; SE = 0.03, p(q [3, 226] = 0.17) < 0.05$. The difference between the number of logically correct responses between the Broad and Reduced premises, $M = 0.76; SE = 0.03$ and between the Close to binary and Reduced premises did not reach significance, $p(q [2, 226] = 0.12) > 0.05$. We then analyzed the Validity X Premise-type interaction. This showed the predicted pattern. For the invalid inferences, the number of logically correct responses was greater for the Broad premises, $M = 0.80; SE = 0.04$, than for the Reduced premises, $M = 0.67; SE = 0.05$, which was greater than the number for the Close to binary premises, $M = 0.36; SE = 0.27, p(q [2, 226] = 0.12) < 0.05$. For the valid inferences, there was no significant difference in the number of logically correct responses between the Broad, Reduced and Close to binary premises.

These results provide a clear replication of the content effects found in Study 1. That is, the fewer Third options are available, the more reasoners tend to be certain about the Deny First and Deny Second inferences, i.e. the greater the tendency to make the False dilemma fallacy. Moreover, the premises used in study 1 represented a causal relation, which could have facilitated a conditional interpretation of the statements. We tried to minimize this possibility here by using categorical premises bound by a commutative relation, i.e. that involved no directionality nor entailment. Again, this suggests that the retrieval process found in conditional reasoning is also at work in deductive reasoning from an incompatibility statement. Moreover, in contrast to Study 1, we did not find evidence for a content effect with the valid inferences. This was to be expected since the premises we used allowed for no realistic exceptions. This also supports the possibility that the retrieval of such exceptions underlied the suppression of valid inferences observed in study 1. However, the relationship between the invalid inferences and the effects of content on the suppression of valid inferences should be investigated in further studies.

Interestingly, participants performed better overall on the valid, $M = 0.90$, $SE = 0.01$, than on the invalid forms, $M = 0.61$, $SE = 0.03$, while no such difference was found in study 1. A possible explanation for this could be that the invalid inferences contained a negated premise while the valid ones did not. In fact, previous research has found that disjunctive inferences are harder with negative categorical premises. (Johnson-Laird et al. 1992 ; Roberge, 1976). Our results suggest a similar phenomenon in reasoning from an incompatibility.

Table 3.2 Mean percentage of logically correct responses for the four logical forms (Deny First, Deny Second, Affirm First, Affirm Second) by Type (Broad, Reduced, Close to binary) in Study 2 (standard deviations in parentheses)

Premise-type	Deny First	Deny Second	Affirm First	Affirm Second
Broad	0.80 (0.34)	0.80 (0.35)	0.90 (0.22)	0.89 (0.24)
Reduced	0.66 (0.41)	0.67 (0.40)	0.83 (0.28)	0.86 (0.27)
Close to binary	0.37 (0.41)	0.35 (0.40)	0.95 (0.11)	0.94 (0.16)

Finally, it should be noted that pragmatic interpretation of the premises from the Close to binary content types might explain some of the high levels of False dilemma fallacies observed with these items, over and above the number of Third options. These premises being action categories, the situations may have been interpreted as ones where it is better to act rather than to abstain. For example, with the premise “Voting for a bill is incompatible with voting against it”, some participants may have interpreted this situation as one where it is better to play a role in the decision and vote, thus favouring a choice between the two propositions. However, in the Broad and Reduced premises which are very similar object categories, this type of pragmatic effect is unlikely to account for the variation between these content types.

3.7 Study 2B

Studies 1 and 2 showed that the number of Third options available underlies the endorsement of invalid inferences with an incompatibility statement, which supports our hypothesis on the effect of content on the False Dilemma fallacy. However, one potential limitation to these studies might be that the formulation “P is incompatible with Q” might have been ambiguous to some participants and might have generated different interpretations of the major premises. In the presentation of these statements, we tried to explain the meaning of an incompatibility by using the logically equivalent conjunctive formulation “P and Q cannot be true at the same time”. An interesting question that arises from this is the extent to which the Conjunctive and Incompatible formulations might produce different interpretations of the major premise. One task that has been frequently used to examine interpretations of statements is the truth table task (Evans & Over, 2004). This presents subjects with a statement involving two terms, followed by the four combinations of true or false terms. Participants must indicate whether each of the combinations make the statement true, false or neither true nor false. Patterns of responses correspond to different forms of interpretation. We thus constructed truth table tasks with a sample of the premises used in study 2. Following an incompatibility statement, the four

combinations of true or false forms of the premises were presented: P & Q, not-P & Q, P & not-Q, not-P & not-Q. Each task was presented with an Incompatible or Conjunctive formulation.

3.7.1 Method

3.7.1.1 Participants and procedure

A total of 100 students (51 men, 49 women, average age: 21 years, 8 months, age range: 17-40) participated in this study. All participants were native French speakers and volunteers and were recruited in colleges or universities in Montreal. Each participant was randomly allocated one of the booklets, was told to take as much time as needed and took part in the experiment individually.

3.7.1.2 Design and materials

Four paper and pencil booklets were constructed. Two of these presented statements of incompatibility using an Incompatible formulation, while two used the Conjunctive formulation. On the first page of each booklet, participants were asked for basic demographic information and were presented with the following instructions:

“In the following pages, you will be presented with various rules. For each rule, you will be presented with different situations. For each of these, you must indicate whether the situation shows that the rule is true, or that the rule is false, or whether it does not show that the rule is true or false. Indicate **true** if the situation shows that the rule is true, **false** if the situation shows that the rule is false, or **one cannot know** if the situation does not allow knowing if the rule is true or false.”

A first booklet was constructed for the Incompatible condition. 6 statements of incompatibility between two propositions were presented, with each statement on the top

of a new page. These statements were taken from study 2. Two of these were from the Close to binary category, which were, in order:

“1. For a player at the “heads and tails” game, betting on tails is incompatible with betting on heads;

2. For a person who votes on a bill, voting for the bill is incompatible with voting against the bill”

The two following statements came from the Reduced category and were, in order:

“3. For a polar animal, being a polar bear is incompatible with being a penguin;

4. For a fruit with pit, being a peach is incompatible with being a cherry”

These were in turn followed by two statements from the Broad category, which were, in order:

“5. For a vegetable, being a broccoli is incompatible with being a pepper;

6. For a person, being in Montreal is incompatible with being in Paris”

Below each statement, 4 specific cases were presented and corresponded to all four combinations of affirming or negating either the first or the second proposition. The order of these was randomly determined for each premise. Three choices were given for each combination. For example, following the statement “For a player at the “heads and tails” game, betting on tails is incompatible with betting on heads”, the following statement, combining a false and a true proposition, was presented with the three options directly following it:

1. Julien bets on heads and doesn't bet on tails.

This situation shows that the rule is:

1. True
2. One cannot know
3. False

The three other situations were:

2. Louis doesn't bet on heads and bets on tails.
3. Martine doesn't bet on heads and doesn't bet on tails.
4. Anne bets on heads and bets on tails.

A second booklet was constructed for the Conjunctive condition. This booklet was identical to the first one, excepted that the statements were presented with a Conjunctive formulation. The two statements from the Close to Binary category were:

- “1. A player at the “heads and tails” game cannot both bet on tail and bet on head;
2. A person who votes on a bill cannot both vote for the bill and vote against the bill”

The two following statements were from the Reduced category:

- “3. A polar animal cannot be both a polar bear and a penguin;
4. A fruit with pit cannot be both a peach and a cherry”

Which were followed by two statements from the Broad category:

- “5. A vegetable cannot be both a broccoli and a pepper;
6. A person cannot be both in Montreal and in Paris”

A second version of each booklet was constructed where the order of the premises was inverted.

3.7.2 Results and discussion

We first examined patterns of responses for each of the 6 statements. Overall, the most common patterns corresponded to two forms of incompatibility interpretations (these combined for 57% of total responses).

3.7.2.1 Incompatibility

This required judging that the P & Q statements made the rule false while the not-P & Q, P & not-Q and not-P & not-Q statements made the rule true.

3.7.2.2 Defective incompatibility

This required judging that the P & Q statements made the rule false, with the not-P & Q and P & not-Q statements judged as making the rule true or considered to be irrelevant and the not-P & not-Q statements judged as irrelevant.

It should be noted that the same kinds of defective interpretations have been found with truth table tasks using conditional statements. The only other relatively frequent pattern that could be interpreted was the exclusive disjunction, which accounted for less than 8% of total responses.

3.7.2.3 Exclusive disjunction

This required judging that the P & Q and not-P & not-Q statements made the rule false and the not-P & Q and P & not-Q statements made the rule true.

Finally, most of the other patterns were uninterpretable and these were grouped together in an "other" category along with a few other interpretable patterns which corresponded to fewer than 5% of total responses. Table 3.3 shows the relative frequency of these patterns as a function of premise type and formulation.

Table 3.3 Interpretations of statements for each premise type (Broad, Reduced, Close to binary) as a function of formulation (Incompatible, Conjunctive)

Interpretation

Premise type	Formulation	Incompatibility	Defective Incompatibility	Exclusive disjunction	Other
Close to Binary	Incompatible	10	44	15	31
	Conjunctive	28	35	12	25
Reduced	Incompatible	8	36	4	52
	Conjunctive	14	40	9	37
Broad	Incompatible	20	41	2	37
	Conjunctive	28	42	3	27

In order to examine whether the two formulations differed in the extent to which they generated incompatibility interpretations, we counted up the number of incompatibility interpretations generated for the two premises for each of the three premise types. We then performed an ANOVA with the number of incompatibility interpretations as dependent variable, with Premise-type as repeated measure and Formulation as a between-subjects variable. This showed nonsignificant effects for both Formulation, $F(1, 98) = 1.98, p = .16$ and the Formulation X Premise-type interaction, $F(2, 97) = .01, p = .98$. There was however a significant effect of Premise-type, $F(2, 97) = 12.69, p < .001$. Post hoc comparisons were made using paired sample t-tests with a Bonferroni correction. This showed that there was a significant difference in the mean number of incompatibility interpretations between the Broad, $M = 1.31, SE = 0.07$, and the Reduced premise types, $M = 0.98, SE = 0.09, t(99) = 5.06, p < 0.001$, with neither showing a significant difference with the Close to Binary premise types, $M = 1.17, SE = 0.08, t(99) = 1.71, p = 0.09$ and $t(99) = 2.07, p = 0.04$, respectively.

Thus, while there were differences in the extent to which the different premise types generated incompatibility interpretations, the use of the different formulations did not by themselves produce significant differences in this respect.

3.8 General Discussion

The False dilemma fallacy is a common error of reasoning. It can be seen as wrongly interpreting a situation as a dichotomy between two options while in fact they are more loosely bound by a relation of incompatibility. This transformation is a rhetorical device which puts the emphasis on the fact that both options cannot be true at the same time, while placing the possibility that both could be false into the background. The strength of this effect can be seen by the results of both studies, which show overall levels of fallacious thinking as high as 65%, even with the kinds of emotionally neutral items used.

Moreover, the results showed a great deal of variability in the extent to which participants made the fallacy. Our key hypothesis here is that a major component of this variability is the result of a retrieval failure whose probability is related to the amount of information stored in semantic memory. In other words, a key factor in understanding the strength of the False dilemma fallacy is the role of semantic memory processes in reasoning (see Markovits, 2014 for a review). In the case of reasoning from an incompatibility, we have argued that the critical information concerns potential Third options, which are cases where neither of the two presented ones are true. The semantic memory model allows the general prediction that the extent to which people make the False dilemma fallacy will be inversely related to the numbers of potential Third options that are readily available in semantic memory.

The results of these studies are consistent with this prediction. Study 1 and 2 show that the tendency to be uncertain about the invalid forms grows with the number of available Third options in the major premise. In Study 1, we translated causal premises into incompatibility statements with many and few Third options. As predicted, the endorsement of the Deny First and Deny Second inference forms was greater when premises contained few Third options than when they contained many Third options. In Study 2, we replicated these effects using premises with non-causal content, i.e. where no directionality is involved. We constructed premises that stated an incompatibility relation between two base-level categories belonging to a parent category. As regards to the number of base-level categories they contained, each parent category was either broad (many base-level categories), reduced (moderate number of base-level categories) or close to binary (close to two base-level categories). We used these categories to construct incompatibility statements where

each base-level category outside of the statement would be a Third option. We predicted that the tendency to draw the Deny First and Deny Second invalid inferences would be inversely proportional to the number of available Third options. Again, our results showed the predicted pattern: endorsement of these inferences was highest when premises came from the Close to binary type, then decreased when premises came from the Reduced type and decreased even more when premises came from the Broad type.

The presented studies are in line with research on content effects in conditional reasoning. It has been shown that the endorsement of the AC and DA inferences decreases when more counterexamples are available. It should be noted that two major classes of theories provide an account for this type of content related variability in conditional reasoning. Mental models theory (Johnson-Laird, 2001; Johnson-Laird & Byrne, 2002) focuses on the use of information in order to generate potential counterexamples, while probabilistic theories consider that people's inferences generate estimations of the likelihood of a given conclusion (e.g., Evans, Over, & Handley, 2005; Oaksford & Chater, 2007). However, both theories adopt the assumption that the core mechanisms of human reasoning are highly sensitive to the semantic properties of the problem rather than being primarily based on formal rules analogous to those of classical logic (Braine, 1978; Rips, 1983). The present studies were conducted within the general perspective that reasoning is affected by memory processes, and adopt a neutral position as regards to the specific debate on the nature of human reasoning. The present results attempt to extend the scope of the semantic memory framework by examining a new type of inference. To our knowledge, these studies are the first to show the content effects on reasoning from an incompatibility statement. This suggests that the impact of the retrieval process on reasoning with meaningful premises goes beyond the scope of conditional reasoning and is a key component in the tendency to draw the False dilemma fallacy. When few Third options were available, reasoners tend to treat an incompatibility statement as a dichotomous one, which is precisely the desired outcome when one uses the false dilemma fallacy to force a choice between two selected options. Our findings suggest that reasoners are vulnerable to such manipulations, more so when few instances of such options can be retrieved from memory. Consequently, users of the false dilemma fallacy, like politicians or advisers, might take advantage of their

audience's lack of knowledge about a topic to persuade people to agree with them while other options are possible.

3.9 Travaux futurs : Extension du RB au raisonnement à partir d'une incompatibilité

Nous exposerons maintenant comment ces observations empiriques permettent d'étendre notre outil d'évaluation automatique du raisonnement à partir d'une incompatibilité. En effet, ces observations permettent de projeter une branche du RB qui concerne le raisonnement à partir d'une incompatibilité avec du contenu familier. La Figure 3.1 expose la tête d'une telle structure. Nous y avons inclus le raisonnement familier (issu de nos expérimentations) ainsi que le raisonnement abstrait, qui serait constitué d'énoncés d'incompatibilité entre des catégories imaginaires. Nous n'avons pas inclus de nœud contrefactuel puisqu'aucune étude n'a investigué la possibilité d'un tel énoncé d'incompatibilité. Le nœud pourrait toutefois être ajouté s'il paraissait souhaitable d'étudier de tels énoncés.

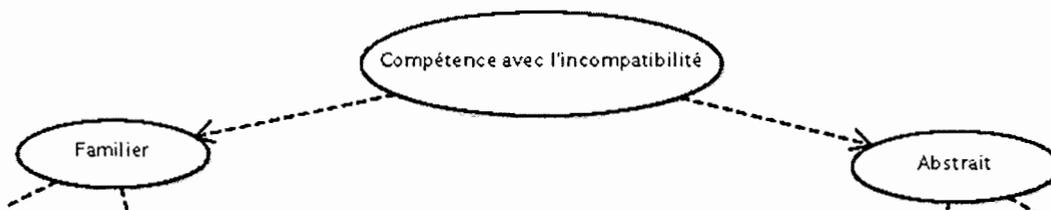


Figure 3.1 Tête projetée du RB pour le raisonnement avec incompatibilité

Considérant que la théorie de la récupération sémantique puisse être étendue au raisonnement avec incompatibilité, nous pouvons supposer que les nœuds d'inhibition des conditions contraignantes, de génération des alternatives et de gestion des trois modèles seront présents dans la structure. De plus, comme nos études ont identifié trois niveaux de contenu familier relativement aux alternatives, nous pouvons supposer l'existence de trois compétences distinctes au niveau des inférences *Deny First* et *Deny Second*. Nous devons toutefois poursuivre les expérimentations, notamment au niveau de possibles conditions contraignantes, avant de valider cette structure.

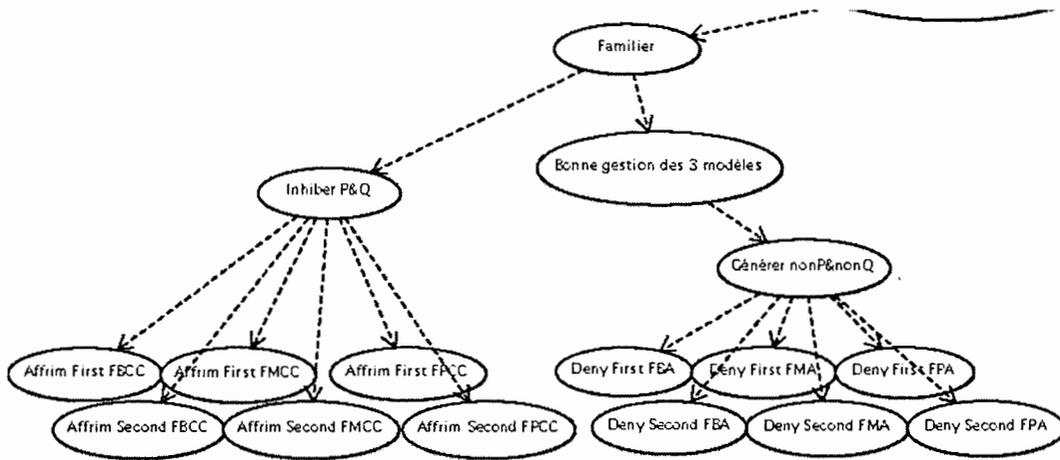


Figure 3.2 Portion projetée du RB pour le raisonnement avec incompatibilité, contenu familial

Il est aussi possible de projeter une portion du RB qui concerne le raisonnement avec incompatibilité avec du contenu abstrait. Partant de cette portion du RB qui concerne le raisonnement conditionnel abstrait, une telle structure pourrait ressembler à l'ébauche exposée dans la figure suivante.

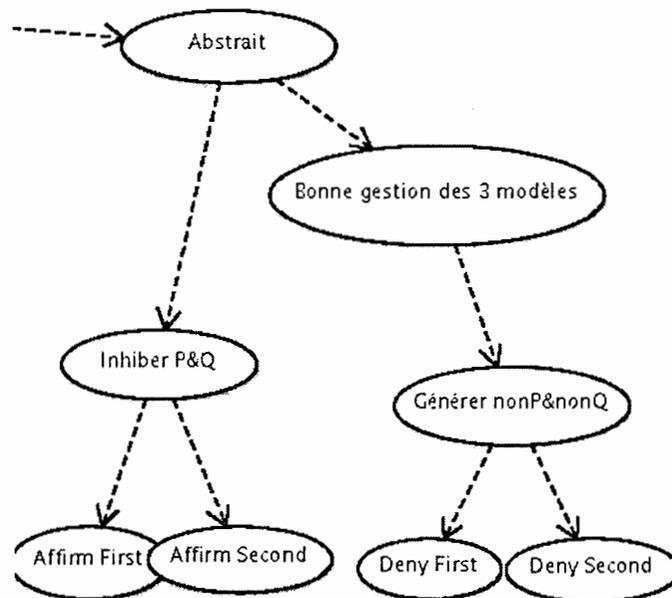


Figure 3.3 Portion projetée du RB pour le raisonnement avec incompatibilité, contenu abstrait

Cependant, afin de valider cette structure, il serait nécessaire de conduire des expérimentations supplémentaires afin de vérifier l'hypothèse selon laquelle ce type de contenu est plus difficile que le contenu familier avec les incompatibilités.

3.10 Travaux futurs : Extension du RB au raisonnement à partir d'une disjonction inclusive

En plus d'une suite des investigations sur le raisonnement à partir d'une incompatibilité, la présente thèse permet de mener des expérimentations semblables à propos du raisonnement disjonctif. La Figure 3.4 expose la tête de la structure. Comme avec l'incompatibilité, nous avons omis le raisonnement contrefactuel puisqu'aucune étude n'a encore exploré une telle possibilité.

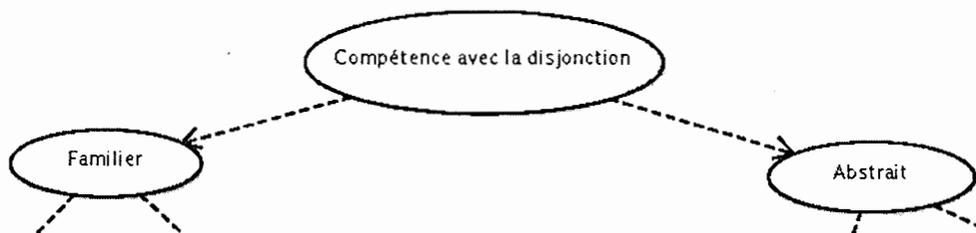


Figure 3.4 Tête projetée du RB pour le raisonnement disjonctif

La Figure 3.5 expose les compétences avec le niveau de contenu familier. Ce dernier peut contenir des énoncés tels que ceux présentés en introduction, p. ex. « Jean joue de la guitare ou du violon ». Des conditions contraignantes seraient des situations (non-P & non-Q), qui contredisent l'énoncé, p. ex. « Jean est trop malade pour jouer d'un instrument ». Les alternatives seraient des situations (P & Q) ou les deux prémisses sont vraies, p.ex. « Jean est capable de jouer des deux instruments simultanément ». Des expérimentations sont toutefois nécessaires afin de valider cette structure.

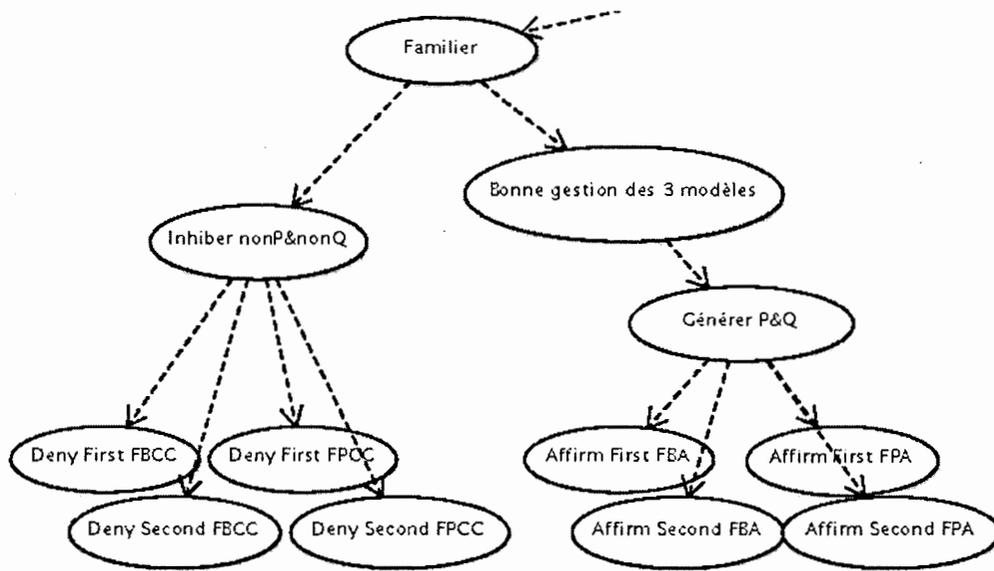


Figure 3.5 Portion projetée du RB pour le raisonnement disjonctif, contenu familier

Comme avec l'incompatibilité, les énoncés disjonctifs abstraits seraient composés de catégories imaginaires. Une fois de plus, des expérimentations devront être conduites afin de valider cette structure.

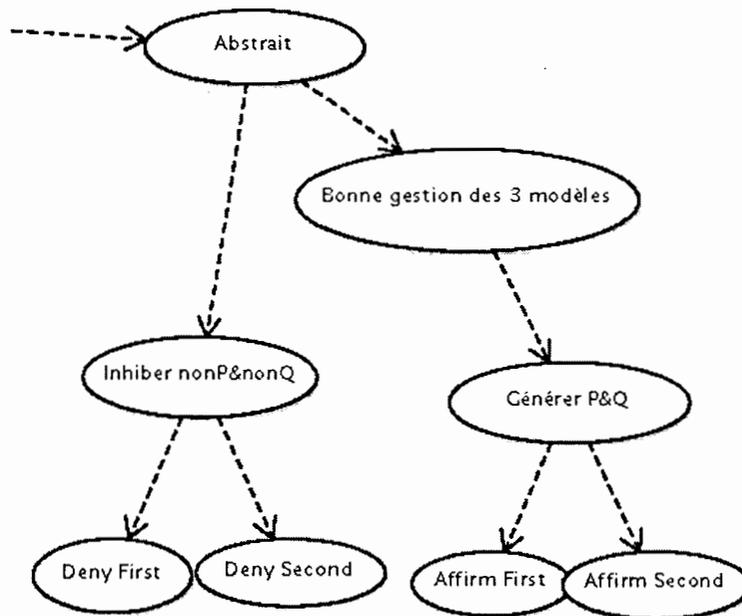


Figure 3.6 Portion projetée du RB pour le raisonnement disjonctif, contenu abstrait

L'évaluation du raisonnement disjonctif comporte cependant un défi particulier. Il est en fait possible de faire une interprétation inclusive (A ou B, ou les deux) ou exclusive (A ou B, mais pas les deux) d'une disjonction. Ces différentes interprétations influent grandement sur les conclusions qui peuvent être acceptées. Par exemple, à partir des prémisses "Jean joue de la guitare ou du violon" et "Jean joue de la guitare", la conclusion "Jean ne joue pas du violon" sera valide à partir d'une interprétation exclusive, mais invalide à partir d'une interprétation inclusive de la disjonction. Toute expérimentation future sur le raisonnement disjonctif devra donc prendre cette difficulté en considération.

3.11 Conclusion

Jusqu'à maintenant, nous avons développé le modèle de l'apprenant de Muse-Logique via un Réseau Bayésien basé sur les effets du contenu dans le raisonnement humain. Les trois grandes branches du réseau (conditionnelle, incompatibilité et disjonction inclusive) exposent les compétences fondamentales que nous avons déterminées à la lumière des phénomènes largement observés en psychologie du raisonnement (surtout au niveau du raisonnement conditionnel) ainsi que de nos propres expérimentations sur les incompatibilités. Cependant, un courant plus récent en psychologie du raisonnement expose un phénomène différent : celui de l'effet de la forme logique dans le raisonnement humain. Le prochain chapitre sera consacré à l'exploration de ce phénomène ainsi que sa pertinence dans le modèle de l'apprenant de Muse-Logique.

CHAPITRE IV

L'INTUITION LOGIQUE

Jusqu'à maintenant, nous avons abordé la modélisation du raisonnement logique par le biais des effets du contenu dans le raisonnement humain. Ceci s'explique par le fait que plusieurs décennies de recherche en psychologie du raisonnement ont été consacrées à ces effets. Ils s'agit donc de phénomènes largement observés, se rapprochant le plus d'un lieu commun dans une communauté connue pour ses nombreuses divergences.

Cependant, une nouvelle perspective a vu le jour dans la dernière décennie de recherche en psychologie du raisonnement : la sensibilité à la forme logique des inférences. Plus précisément, l'hypothèse de l'intuition logique suggère qu'à partir d'une exposition répétée à certains principes logiques de base, les raisonneurs pourraient développer une compréhension intuitive de ces derniers (De Neys, 2012, 2014). Cette perspective s'inscrit dans le cadre théorique des théories à processus duaux. Ces dernières postulent que les individus possèdent deux systèmes distincts (Evans, 2007; Evans, Over et Handley, 2005; Sloman, 1996; Stanovich & West, 2000). Le système 1 (souvent appelé système heuristique) génère des jugements ou des déductions à l'aide de processus intuitifs qui permettent la génération rapide de réponses. Le système 2 (souvent appelé système analytique) permet un processus plus lent et coûteux sur le plan cognitif. Il existe différentes versions de ces théories et on observe un faible consensus quant aux caractéristiques précises de ces deux systèmes. Cependant, les études menées dans le cadre de cette approche examinent généralement les interactions entre les processus de raisonnement analytique et les processus intuitifs.

L'intérêt pour la question de l'intuition logique est croissant dans la communauté des psychologues du raisonnement, ce qui a engendré des études de plus en plus nombreuses sur la question. Il semble donc pertinent de nous pencher sur ce volet et d'extraire les phénomènes pouvant enrichir le modèle de l'apprenant de Muse-Logique.

Dans l'article qui suit, nous ferons une revue de la littérature sur la question de l'intuition logique et nous présenterons les résultats d'expérimentations concernant la portée et les limites de l'intuition logique au sein du raisonnement conditionnel. Cet article a été publié dans la revue *Psychologica Belgica*.

4.1 Abstract

Empirical evidence for the capacity to detect conflict between biased reasoning and normative principles has led to the proposal that reasoners have an intuitive grasp of some basic logical principles. In two studies, we investigate the boundary conditions of these logical intuitions by manipulating the logical complexity of problems where logical validity and conclusion believability conflict or not. Results pointed to evidence for successful conflict detection on the basic Modus Ponens (MP) inference, but also showed evidence for such a phenomenon on the more complex Modus Tollens (MT) inference. This suggests that both the MP and the MT inferences are simple enough for reasoners to have an intuitive grasp of their logical structure. The boundaries of logical intuition might thus reside in problems of greater complexity than these inferences. We also observed that on the invalid Affirmation of the Consequent (AC) and Denial of the Antecedent (DA) inferences, participants showed higher accuracy on the inference that was expected to be more complex (DA), and no evidence for successful conflict detection was found on these forms. Implications for the logical intuition framework are discussed.

4.2 Introduction

The biased nature of human inferential processes has been extensively demonstrated in decades of research on reasoning and decision making. Educated adults often violate the elementary principles of logic, probability or mathematics and favor fast and intuitive rules-of-thumb, called heuristics, to more deliberative thinking (Gilovich, Griffin & Kahneman, 2002). While an intuitive response can be congruent with the normative one, it sometimes conflicts with basic normative principles. A striking example of such a situation is the bat-and-ball problem (Frederick, 2005):

“A bat and a ball together cost \$1.10. The bat costs \$1 more than the ball. How much does the ball cost?”

Obviously, the correct answer is that the ball costs 5 cents and the bat costs \$1.05 for a total of \$1.10, but an intuitive answer “The ball costs 10 cents” is given by a vast majority

of educated university students (Bourgeois-Gironde & Van der Henst, 2009). Of course, anyone who is familiar with the most basic principles of algebra should be able to come up with the right answer. So why do so many educated adults miss the goal? One answer would be that the parsing of \$1.10 in \$1 and 10 cents comes to mind so naturally that the intuitive “10 cents” answer becomes irresistible to many people (Kahneman, 2011).

However, it is unlikely that these biased people have no access to the normative response whatsoever. One question that arises from this is that when reasoners give an intuitive response that conflicts with a normative principle, are they perhaps aware of this conflict? Numerous studies have pursued this question by examining the detection of conflict in reasoning (e.g., De Neys & Glumicic, 2008; Pennycook, Fugelsang, & Koehler, 2015; Thompson & Johnson, 2014). The key manipulation of these studies is to present participants with problems for which the intuitive response conflicts with the normative one (like the above bat-and-ball problem) and a control version where both responses are the same. A no-conflict version of this problem could be:

“A bat and a ball together cost \$1.10. The bat costs \$1. How much does the ball cost?”

In this case, both mathematics and the intuitive parsing of \$1.10 in \$1 and 10 cents would lead to the same “10 cents” response.

Many studies have indicated that reasoners seem to process the conflict problems for which they gave the heuristic response differently than the no-conflict ones. For example, reasoners who answer intuitively to conflict problems need more time (Bonner & Newell, 2010; De Neys & Glumicic, 2008; Pennycook, Trippas, Handley, & Thompson, 2014; Villejoubert, 2009; Stupple, Ball, Evans, & Kamal-Smith, 2011), are less confident about their response (Bago & De Neys, 2017; De Neys, Cromheeke, & Osman, 2011; Gangemi, Bourgeois-Gironde, & Mancini, 2015; Johnson, Tubau, & De Neys, 2016; Thompson & Johnson, 2014) and show increased activation of brain areas assumed to mediate conflict and error monitoring (De Neys, Vartanian, & Goel, 2008; Simon, Lubin, Houdé, & De Neys, 2015) compared to when they give the normative answer to the no-conflict ones. These studies thus provide basic evidence for the presence of conflict detection in biased

reasoners. That is, even when reasoners fall for an erroneous, intuitively cued response, they seem to show sensitivity to the fact that it is logically inappropriate.

This literature has led to the proposal that reasoners have an intuitive grasp of some basic logical principles (De Neys, 2012, 2014). The basic idea of this proposal is that when reasoners give a biased response to a logical problem, they can intuitively detect that something is wrong with their answer, but subsequently fail to override the prepotent biased response. Some form of comprehension of the logical principle at stake would thus be implied in the intuitive detection stage.

4.3 Evidence for logical intuition: the “logical bias”

Many studies provide more general empirical support for this proposal. For example, some scholars have observed what we could call a “logical bias”. This bias is related to the well-known belief bias, that is the tendency to judge the validity of an argument based on the accordance between its conclusion and one’s beliefs rather than its logical structure. This heuristic makes people more prone to endorse an invalid argument when its conclusion is believable and reject a valid one when its conclusion contradicts one’s beliefs. However, recent studies have explored the possibility of a reversed phenomenon, that is, the possibility that the logical validity of an argument may bias judgments of believability. These studies presented participants with inferences where logic and belief were in accordance or in conflict and instructed them to evaluate their conclusions’ believability (Handley & Trippas, 2015; Trippas, Thompson, & Handley, 2017) or likability and brightness (Trippas, Handley, Verde, & Morsanyi, 2016) in a short period of time. They found that people needed more time and were more prone to errors when the logical validity of the problems conflicted with the conclusion’s believability. These studies thus provide empirical evidence for the presence of a “logic bias”. If logical processing would necessarily require slow and deliberate processing, then it should obviously not interfere with the evaluation of intuitive beliefs. These findings consequently lend credence to the proposal that people have an intuitive grasp of basic logical principles.

4.4 Additional supportive finding: The two-response paradigm

Additional supportive studies of the logical intuition hypothesis use a two-response paradigm where participants are presented with reasoning problems and are first asked to give the first, intuitive response that comes to mind. To make sure that the initial response is truly intuitive in nature participants are given a short deadline or are imposed a cognitive load task (Bago & De Neys, 2017; Newman, Gibb, & Thompson, 2017). They are then given as much time as they need to give their final answer. These studies found that even in the fast and challenging conditions, participants generated correct logical responses and accepted valid inferences more often than the invalid ones, thus suggesting that a logical response can come from fast and intuitive processes.

4.5 Critiques and boundary conditions of logical intuition

However, some critiques have been raised against studies on conflict detection and the logical intuition proposal. Some argue that the results supporting conflict detection could be due to a confound between the manipulation of conflict and the specific content or format of the problems (Aczel, Szollosi, & Bago, 2016; Klauer & Singmann, 2013; Singmann, Klauer, & Kellen, 2014). Others claim that incorrect responses to conflict problems, rather than being due to a failure to override the intuitive response, might arise earlier in the reasoning process and need to be attributed to an inaccurate comprehension of the problem at hand (Mata, Ferreira, Voss, & Kollei, 2017; Mata, Schubert, & Ferreira, 2014). Others address the extent of the proposal, arguing that the capacity to detect conflict might be limited to tasks where the contrast between the intuitive and the normative answers is amplified (Pennycook, Fugelsang, & Koehler, 2012) or when the underlying principle is simple (Travers, Rolison, & Feeney, 2016). These latter critiques lead to a key open question, that will be the major focus of the current studies, namely to define the boundary conditions of logical intuitions. Can we assume that reasoners have a logical intuition about each and every problem? Our stance on this question is rather that, except for highly trained logicians, logical intuitions arise only for simple problems to which people have been exposed frequently enough to develop an intuitive sense of their structure

(De Neys, 2012, 2014). We thus posit the main hypothesis that, when presented with logical problems that contain a more complex structure, reasoners will be less likely to show conflict detection since it will be less likely that a logical intuition will arise from the task.

For completeness, it should be noted that the logical intuition concept plays a key role in several (related) recent dual process models. For example, as the basis of a so-called hybrid dual-process account (De Neys, 2018), as a component of the three-stage model (Pennycook, 2018), or occurring alongside belief-based thinking in a parallel-processing model (Trippas & Handley, 2018). Rather than being a test for theory, the following studies are an empirical enquiry regarding the scope of logical intuitions, and can thus be a useful contribution to all of these models.

In continuity with previous studies on conflict detection in logical reasoning (Bago & De Neys, 2017; De Neys & Franssens, 2009; De Neys, Moyens, & Vansteenwegen, 2010) we used simple categorical syllogisms that were first introduced by Markovits and Nantel (1989) and popularized through the work of Stanovich and West (1998, 2000; West, Toplak, & Stanovich, 2008) and others. The first syllogism starts with the premises “All As are Bs; All Cs are As” and leads to the valid conclusion “All Cs are Bs”. For example:

“All dogs have legs

Labradors are dogs

Therefore, Labradors have legs”

The second starts with the premises “All As are Bs; All Cs are Bs” and lead to the invalid conclusion “All Cs are a As”. Such an inference could be:

“All dogs have legs

Labradors have legs

Therefore, Labradors are dogs”

This latter conclusion is invalid since many animals besides dogs can have legs. The conclusion thus follows possibly, but not necessarily, from the premises.

These syllogisms are categorical versions of two conditional inferences. The first is equivalent to the valid Modus Ponens or MP (If P then Q, P is true, therefore Q is true) and the second is equivalent to the invalid Affirmation of the Consequent or AC (If P then Q, Q is true, therefore P is true). Previous developmental research has suggested that categorical syllogisms are simpler to process than their propositional counterpart (Markovits, 2017). We thus started with these items as a lower bound in our enquiry for evidence of logical intuitions. Note that we will refer to these syllogisms as MP and AC, respectively.

One way to directly look for an upper boundary conditions of logical intuitions would be to increase these problems' complexity. These MP and AC inferences are indeed simple and one element of their simplicity is that their premises are always affirmed. One way to increase their complexity is to include negations in their premises and conclusion, as it is well established that the presence of a negation adds a cognitive burden to the reasoning process (Schaeken & Schroyens, 2000; Schroyens, Schaeken, Fias & d'Ydewalle, 2000; Schroyens, Schaeken & d'Ydewalle, 2001). Adding a negation to our previous MP syllogism could result in:

“All dogs have legs

Cats are not dogs

Therefore, cats don't have legs”

The syllogism would then become invalid and be equivalent to the Denial of the Antecedent conditional inference (If P then Q, P is false, therefore Q is false). Moreover, adding a negation to the AC syllogism could result in:

“All dogs have legs

Cats don't have legs

Therefore, cats are not dogs”

Which would equate to the valid Modus Tollens (If P then Q, Q is false, therefore P is false). Note that we will refer to these two syllogisms as DA and MT, respectively.

We will thus use these four inference forms to manipulate logical complexity in order to test for the boundary conditions of logical intuition, MP and AC being the simple inferences and MT and DA the more complex ones. Note that, while conflict detection on propositional versions of MP and MT has been previously observed (Trippas et al., 2017), to our knowledge, our studies are the first to manipulate complexity with these categorical items and with these four inferences forms.

We hypothesized that for more complex inferences, it will be less likely that people will have an intuitive grasp of their logical structure. Consequently, previously observed conflict detection effects (e.g., increased response latencies for conflict problems) should be less likely with the complex inferences than with the simpler ones.

4.6 Pretest

To test our hypothesis, we first constructed four sets of 16 categorical syllogisms for a pretest. Within each set, conclusion believability was in conflict with validity for half of the problems (i.e., conflict problems; two unbelievable MP, two unbelievable MT, two believable AC and two believable DA). For the other half, believability was consistent with validity (i.e., no-conflict problems; two believable MP, two believable MT, two unbelievable AC and two unbelievable DA). In each problem set, two additional conflict and two no-conflict syllogisms were added, for a total of 20 syllogisms in each set. This addition was done in order to be able to afterwards select (for the main study) a final set of 64 conclusions shown to be maximally believable or unbelievable in the pretest.

To minimize the possibility that content related variability would affect our results, we crossed the item content with the conflict status and logical form complexity across the four sets. That is, with the same major premise, we switched the order of the minor premise and the conclusion to manipulate problem validity, thus turning MP problems into an AC

form and MT problems into a DA form, and vice versa. To make sure that the minor premise was always believable, we also changed the minor term of the unbelievable conclusions. Consequently, across the four blocks, the same major premise was used to construct a different type of problem. A major premise that was used in one block to construct a simple conflict problem would be used to construct a complex no-conflict problem in another block, etc. This is a first step to minimize the possibility that simple item content differences bias the effect of problem complexity or conflict.

The conclusion believability classification in our items sets was based on previously published classifications and our own intuitions. To validate the classification and avoid confounds, we ran an extensive believability rating pretest for the 80 conclusions in our item sets.

Twelve participants took part in the pretest (3 females, 9 males, Mean age = 38 years, 4 months). We asked them to rate the believability of each conclusion on a scale of 0 to 10 (0 being totally unbelievable and 10 being totally believable). Conclusions from the item sets were presented in four different blocks, each block consisting of the 20 conclusions of one item set. Questions within each set were presented to participants in a randomized fixed order.

We first calculated the mean believability ratings for each of the 80 conclusions. We discarded 16 conclusions with moderate ratings, so that each conclusion in our item sets would be maximally believable (close to 10) or unbelievable (close to 0) with as little variance as possible.

To check whether the average believability of the selected material did not systematically differ, we performed a 2 (Conflict: conflict, no-conflict) X 2 (Complexity: easy, hard) within-subject ANOVA on believability ratings. Average ratings for conflict ($M = 5.12$, $SE = 1.02$) and no-conflict problems ($M = 5.11$, $SE = 1.05$) and simple ($M = 5.292$, $SE = 1.1$) and complex problems ($M = 4.93$, $SE = 0.97$) were very close. The ANOVA showed no significant effect of Conflict, $F(1, 15) < 1$, nor Complexity, $F(1, 15) = 2.532$, $p = 0.13$, and no significant interaction between Conflict and Complexity $F(1, 15) = 2.409$, $p = 0.14$. We can thus minimize the possibility that effects resulting from our manipulation of

conflict and complexity on conflict detection will be attributable to differences in conclusions believability. APPENDIX F gives a complete overview of the 16 selected problems in the four item sets and APPENDIX G gives a overview of the mean believability ratings for each of their conclusions.

4.7 Study 1

4.7.1 Method

4.7.1.1 Participants

A total of 95 participants (53 males, 42 females, Mean age = 32 years, 1 month) were recruited via the online Crowdfunder platform and received \$0.30 for their participation. Only native English speakers from the USA or Canada were allowed to participate in the study. A total of 31.6% of participants reported high school as highest completed educational level, 2.1% reported not having a high school degree and 66.3% reported having a post-secondary education degree. Note that previous studies have shown that both laboratory based (e.g., De Neys & Glumicic, 2008) and online settings (e.g., Frey, Johnson, & De Neys, 2017) give similar latency results. This online administration is thus in continuity with the relevant literature on conflict detection.

4.7.1.2 Materials and procedure

The four item sets selected in the pretest were used in this experiment. Participants were randomly divided into four groups to which one item set was assigned. Items were presented to them in a randomized order. Hence, each participant solved a total of 16 problems. Half of these were conflict problems (i.e., two unbelievable MP, two unbelievable MT, two believable AC and two believable DA) and half were no-conflict problems (i.e., two believable MP, two believable MT, two unbelievable AC and two

unbelievable DA). No time limit was imposed. All participants were first given the following instructions:

“In this experiment, you will need to solve a number of reasoning problems. In each problem you are going to get two premises, which you have to assume being true. Below the premises you will see a conclusion. We ask you to determine whether the conclusion follows logically from the premises or not. You have to assume that the premises are all true. This is very important.

Below you can see an example of the problems:

Premise 1: All dogs have four legs

Premise 2: Puppies are dogs

Conclusion: Puppies have four legs

Does the conclusion follow logically?

Yes

No

Once you have made up your mind we ask you to immediately click on the corresponding answer option. Then you have to click on the red “Next” button to advance to the next problem. Press "Next" if you are ready to start the practice session!”

Participants then received a practice session (not analyzed) of one additional item. They then proceeded to answer the 16 reasoning problems. Half of participants were randomly assigned to solve 8 easy problems first followed by 8 hard problems and half were randomly assigned to a reversed order of complexity. They were then all asked to provide basic demographic information.

4.7.2 Results and discussion

A preliminary data check showed that 4 participants took an unusually short time to read the instructions page (less than 2 seconds, whereas average reading time was 78 s, SD = 314s). We therefore decided to discard the data of these four participants from further analysis.

4.7.2.1 Manipulation check: Accuracy findings

We first wanted to verify whether our complexity manipulation was successful. Are the complex inferences really harder to solve than the simple ones? We therefore first looked at accuracy. Table 4.1 shows the results. Overall, participants performed better on the easy (M=65.1%, SD=21%) than on the hard problems (M=59.6% SD=18.1%), $t(90) = 2.81$, $p < 0.01$. However, examination of Table 4.1 indicates that this did not hold for all problem types. On the valid MP/MT problems, we did observe the expected pattern with higher accuracies throughout on the easy MP problems than on the complex MT problems, (M=78.6%, SD=23.4%; M=63.7%, SD=31.5%, respectively), $t(90) = 4.57$, $p < 0.001$. However, the difference between the invalid AC and DA problems was not significant (M=51.7%, SD=32.2%; M=55.5%, SD=34.1%, respectively), $t(90) = -1.41$, $p = 0.16$.

Our manipulation of complexity was thus in line with expectations for MP and MT but not for AC and DA. Surprisingly, response accuracy did not indicate that the DA inference was harder than the AC inference. Obviously, the accuracy findings complicate a test of our main conflict detection hypothesis for the AC/DA problems. Since problem complexity was a key manipulation for our hypothesis, further analysis on conflict detection will be done for valid MP/MT and invalid AC/DA inferences separately.

4.7.2.2 Conflict detection analysis.

Preliminary analysis of latencies showed positive skewness for failed conflict MP (skewness = 3.05, $SE = 0.51$) and AC (skewness = 3.01, $SE = 0.51$) as well as succeeded no conflict MT (skewness = 2.20, $SE = 0.51$). All latencies were thus log-transformed before further analysis. See Table 4.2 for an overview of the raw latency findings.

Valid inferences. Consistent with previous conflict detection studies (e.g., De Neys & Glumicic, 2008; De Neys et al., 2010; Frey, Johnson & De Neys, 2017), to test for conflict detection, we analyzed response latency on MP and MT for participants who failed the conflict items and succeeded the no-conflict items. We then performed a 2 (Conflict: incorrect conflict, correct no-conflict) X 2 (Complexity: MP, MT) within-subject ANOVA

on response latency. First, a main effect of Conflict showed that when participants failed the conflict problems, they took more time ($M= 10.93$, $SD=10.85$) than when they succeeded the no-conflict ones ($M= 8.26$, $SD= 6.65$), $F(1, 27)= 5.073$, $p < 0.05$, *partial eta*² = 0.16. We found no significant effect of Complexity, $F(1, 27) < 1$, and, critically, no significant interaction between Conflict and Complexity, $F(1, 27) < 1$.

First, these results suggest that reasoners are sensitive to the conflict between their biased response and the logical one, which is in line with our general hypothesis on conflict detection. However, the absence of interaction between Conflict and Complexity suggests that, contrary to our predictions, our manipulation of complexity had no significant impact on conflict detection. Indeed, as Table 4.2 indicates, if anything there was a trend towards a slightly stronger conflict detection effect on the harder MT than on the easier MP problem.

Invalid inferences. We then ran the same analysis on the invalid problems. A 2 (Conflict: incorrect conflict, correct no-conflict) X 2 (Complexity: AC, DA) within-subject ANOVA on response latency. A main effect of Complexity showed that participants took more time to solve the DA ($M= 10.75$, $SD= 7.09$) than the AC ($M= 8.32$, $SD= 5.65$) inference, $F(1, 33)= 11.599$, $p < 0.01$, *partial eta*² = 0.26. Moreover, a marginally significant main effect of Conflict suggested that, contrary to our predictions, participants took more time to solve the correct no-conflict problems ($M= 10.206$, $SD= 6.05$) than the incorrect conflict ones ($M=8.87$, $SD=6.69$), $F(1, 33)= 3.358$, $p = 0.076$, *partial eta*² = 0.09. No significant interaction between Conflict and Complexity was found, $F(1, 33)= 1.094$, $p = 0.3$. These results thus show that we did not replicate the evidence for conflict detection on the AC inference and that we did not find such evidence for the DA inference.

These unexpected results on accuracy and conflict detection with the invalid inferences are puzzling. Study 2 was run to test the robustness of these effects. Moreover, one possible limitation of Study 1 is that we did not control for differences in premises reading time. Stimulus presentation was thus slightly modified in Study 2 in order to provide a finer measurement of reasoning latencies.

4.8 Study 2

4.8.1 Method

4.8.1.1 Participants

A total of 96 participants (50 males, 45 females, Mean age = 34 years, 10 months, one participant failed to indicate demographic information) were recruited via the online Prolific Academic platform and received \$ 0.68 for their participation. Only native English speakers from the United States, Canada, United Kingdom, New Zealand and Australia were allowed to participate in the study. A total of 35.4% of participants reported high school as highest completed educational level and 63.6 % reported having a post-secondary education degree.

4.8.1.2 Materials and procedure

We used the exact same material and procedure as in Study 1, except that we presented the premises and conclusion serially. That is, participants first saw the major and minor premises and were instructed to click “next” once they finished reading them. The conclusion and question were then added to the premises, which completed the reasoning problem. Response latencies were calculated from then. We reasoned that removing premises’ reading times from the response latencies might result in more accurate measurement of the actual reasoning time.

4.8.2 Results and discussion

As in Study 1, preliminary data checking indicated that 2 participants showed a deviant short time to read the instructions page (less than 2.1 seconds; with an average reading time of 50 s, SD = 64 s). Data from these two participants was discarded from further analysis.

4.8.2.1 Manipulation check: accuracy findings

We first looked at our complexity manipulation. Table 4.1 shows the results. As in Study 1, participants performed better overall on the easy ($M=77.5\%$, $SD=22.6\%$) than on the hard problems ($M=72.2\%$, $SD=21.9\%$), $t(93) = 2.75$, $p < 0.01$. As expected, we observed higher accuracies on MP ($M=81.2\%$, $SD=23.6\%$) than on the MT problems, ($M=68.4\%$, $SD=32.8\%$), $t(93) = 3.88$, $p < 0.001$. Once again the difference between the AC ($M=73.9\%$, $SD=29.3\%$) and DA ($M=76.1\%$, $SD=30.4\%$) inferences was not significant, $t(93) = -0.79$, $p < 0.43$. These results thus replicate what was found in Study 1. Our manipulation of complexity being supported only for MP/MT, conflict detection will again be analyzed separately for valid and invalid forms.

4.8.2.2 Conflict detection analysis.

As in study 1, preliminary analysis of latencies showed positive skewness for some variables, namely failed conflict MP (skewness = 2.26, $SE = 0.55$), MT (skewness = 3.07, $SE = 0.55$), AC (skewness = 3.15, $SE = 0.55$) and DA (skewness = 2.46, $SE = 0.55$) as well as succeeded no conflict DA (skewness = 3.55, $SE = 0.55$). Further analyses were thus conducted on log-transformed data (see Table 2 for raw latencies).

Valid inferences. We then analyzed response latency on MP and MT for participants who failed the conflict items and succeeded the no-conflict one. As in Study 1, we performed a 2 (Conflict: incorrect conflict, correct no-conflict) X 2 (Complexity: MP, MT) within-subject ANOVA on log-transformed response latencies. A marginally significant main effect of Conflict suggested that when participants failed the conflict problems, they took more time ($M = 6.99$, $SE = 1.2$) than when they succeeded the no-conflict ones ($M = 4.3$, $SE = 0.34$), $F(1, 27) = 3.933$, $p = 0.058$, $partial\ eta^2 = 0.13$. No significant effect of Complexity, $F(1, 27) = 1.794$, $p = 0.19$, and no significant interaction between Conflict and Complexity were found, $F(1, 27) = 1.305$, $p = 0.26$. Although the results were marginally significant, this study suggests a replication of the conflict detection findings for both the MP and MT inferences, with our manipulation of complexity having no significant impact on the effect.

Invalid inferences. We then ran the same analysis on the invalid problems. A 2 (Conflict: incorrect conflict, correct no-conflict) X 2 (Complexity: AC, DA) within-subject ANOVA on response latency showed the same pattern as observed in Study 1. A main effect of Complexity showed that participants took more time to solve the DA ($M= 7.11$, $SE= 0.77$) than the AC ($M= 4$, $SE= 1.07$) inference, $F(1, 28)= 27.296$, $p < 0.001$, $partial\ eta^2 = 0.49$. There was a marginal significant main effect of Conflict, $F(1, 28) = 3.682$, $p = 0.07$ that pointed to *longer* latencies on the no-conflict problems. Finally, Conflict and Complexity did not interact, $F(1, 28) < 1$.

The purpose of this study was to validate the unexpected findings of Study 1, that is, evidence for conflict detection on the valid MP and MT inference with no difference across complexity and no evidence for conflict detection for the invalid AC and DA forms. The results found here did replicate those findings, although conflict detection on the valid forms was marginally significant. Provided that both studies showed the same tendencies, we decided to pool the data obtained in both studies and apply the same analysis with increased power.

4.8.2.3 Pooled conflict detection analysis

Regarding conflict detection on the valid inferences, a 2 (Conflict: incorrect conflict, correct no-conflict) X 2 (Complexity: MP, MT) within-subject ANOVA on response latency gave a significant main effect of Conflict, which showed again that when participants failed the conflict problems, they took more time ($M= 8.97$, $SE= 0.97$) than when they succeeded the no-conflict ones ($M= 6.28$, $SE= 0.6$), $F(1, 55) = 8.79$, $p < 0.01$, $partial\ eta^2 = 0.14$. We found no significant effect of Complexity, $F(1, 55) = 1.713$, $p = 0.2$, and no significant interaction between Conflict and Complexity, $F(1, 55) < 1$.

With respect to the invalid problems, a 2 (Conflict: incorrect conflict, correct no-conflict) X 2 (Complexity: AC, DA) within-subject ANOVA on response latency gave a main effect of Complexity, which showed that participants took more time to solve the DA ($M= 9.08$, $SE= 0.71$) than the AC ($M= 6.31$, $SE= 0.55$) inference, $F(1, 62)= 36.029$, $p < 0.001$, $partial\ eta^2 = 0.37$. There was also a main effect of Conflict, but in contrast with the conflict

detection hypothesis, the effect was again reversed so that when participants failed the conflict problems, they took less time ($M= 7.11$, $SE= 0.76$) than when they succeeded the no-conflict ones ($M= 8.28$ $SE= 0.65$), $F(1, 62)= 7.143$, $p < 0.05$, $partial\ eta^2 = 0.10$. No significant interaction between Conflict and Complexity was found, $F(1, 55) = 1.228$, $p = 0.27$.

Table 4.1 Percentage of logically correct responses on overall, conflict and no-conflict problems for Study 1, Study 2, and pooled data (standard deviations in parentheses)

	Logical form	Overall	Conflict	No-conflict
Study 1	MP	78.6 (23.4)	63.2 (40.7)	94 (16.4)
	MT	63.7 (31.5)	54.9 (40.2)	70.3 (38.7)
	AC	51.7 (32.2)	33. (41)	72.5 (37.5)
	DA	55.5 (34.1)	42.9 (43.8)	68.1 (40.5)
Study 2	MP	81.1 (23.7)	63.3 (45.3)	98.9 (7.3)
	MT	68.4 (32.8)	60.6 (44.5)	88.8 (24.5)
	AC	73.9 (29.3)	59. (46.4)	76.1 (34.2)
	DA	76.1 (30.4)	65.4 (42.8)	86.7 (27.6)
Pooled data	MP	79.9 (23.5)	63.2 (43)	96.5 (12.8)
	MT	66.1 (32.2)	57.8 (42.4)	79.7 (33.5)
	AC	63 (32.7)	46.2 (45.6)	74.3 (35.8)
	DA	66 (33.8)	54.3 (44.6)	77.6 (35.7)

Table 4.2 Response latency (in seconds) for incorrect conflict and correct no-conflict problems (standard deviations in parentheses)

	Logical form	Conflict	Accuracy	Response time	Conflict detection effect*
Study 1	MP	Conflict	Incorrect	10.09 (9.88)	1.83
		No-Conflict	Correct	8.26 (6.79)	
	MT	Conflict	Incorrect	11.78 (11.82)	3.57
		No-Conflict	Correct	8.27 (6.52)	
	AC	Conflict	Incorrect	7.5 (6.1)	-1.65
		No-Conflict	Correct	9.15 (5.21)	
DA	Conflict	Incorrect	10.24 (7.29)	-1.03	
	No-Conflict	Correct	11.26 (6.9)		
Study 2	MP	Conflict	Incorrect	6.29 (6.58)	2.58
		No-Conflict	Correct	3.71 (2.42)	
	MT	Conflict	Incorrect	7.7 (10.73)	2.78
		No-Conflict	Correct	4.89 (2.43)	

	AC	Conflict	Incorrect	3.46 (3.67)	-0.98
		No-Conflict	Correct	4.44 (3.45)	
	DA	Conflict	Incorrect	6.61 (8.14)	-1
		No-Conflict	Correct	6.61 (7.56)	
Pooled	MP	Conflict	Incorrect	8.19 (8.53)	2.21
		No-Conflict	Correct	5.98 (5.55)	
	MT	Conflict	Incorrect	9.74 (11.38)	3.16
		No-Conflict	Correct	6.58 (5.16)	
	AC	Conflict	Incorrect	5.64 (5.48)	-1.34
		No-Conflict	Correct	6.98 (5.04)	
	DA	Conflict	Incorrect	8.57 (7.84)	-1.01
		No-Conflict	Correct	9.58 (7.38)	

Note. * Incorrect conflict minus correct no-conflict trials latency difference. More positive values indicate stronger detection effect.

4.8.2.4 Conflict and believability confound

As explained in the presentation of the material, our conflict problems were created by crossing problem believability with logical validity, so that both believable–invalid problems and unbelievable–valid problems were constructed. Given the unexpected accuracy findings, we analysed the valid and invalid problems separately. One possible limitation that arises from this is that our manipulation of conflict was collapsed with conclusion believability. That is, in the valid problems, the conflict versions are always unbelievable whereas no-conflict problems were believable. Similarly, for invalid problems the conflict versions were always believable whereas no-conflict problems were unbelievable. Hence, one alternative explanation for our conflict findings is simply that people take more time to evaluate unbelievable conclusions. This would lead to a “conflict detection” effect on valid problems and a reversed effect (i.e., longer latencies for no-conflict problems) on invalid problems – exactly the pattern we observed in our pooled analysis. One way to control for this potential confound is to test for an effect of conflict for problems that are matched on believability (i.e., valid unbelievable vs. invalid unbelievable and invalid believable vs. valid believable). Hence, conflict and no conflict problems differ in validity but not in believability. In other words, in this matched control analysis the “conflict” effects always contrast valid and invalid problems rather than

validity and believability. If participants are sensitive to logical validity and not simply believability per se, they should still show longer latencies for conflict versions here. Therefore, in a control analysis on our pooled Study 1 and 2 data we also tested for the effect of this conflict factor on response latencies. Believability was entered as a separate factor in the design. This resulted in a 2 (Conflict: conflict, no-conflict) X 2 (Believability: believable, unbelievable) ANOVA on response latencies. A main effect of Believability showed that overall, participants took less time to solve the believable ($M = 8.12$, $SE = 0.52$) than the unbelievable problems ($M = 10.1$, $SE = 0.78$), $F(1, 184) = 22.836$, $p < 0.001$, $partial\ eta^2 = 0.11$. Critically, a main effect of Conflict, $F(1, 184) = 12.238$, $p = 0.001$, $partial\ eta^2 = 0.06$, showed that participants still took more time to solve the conflict, ($M = 9.39$, $SE = 0.63$) than the no conflict problems, ($M = 8.83$, $SE = 0.72$), even when the believability status was matched. No interaction between Conflict and Believability was found, $F(1, 184) < 1$. This establishes that our overall results are not driven by a believability confound but by the presence or absence of conflict between logical validity and believability per se.

4.8.2.5 Bayesian null-effect complexity test

With respect to the valid inferences, our pooled analysis corroborated the presence of conflict detection for both the MP and MT inferences without any effect of complexity. If anything, the effect tended to be more pronounced on the harder MT than on the easier MP inference. However, the critical conclusion with respect to a lack of complexity effect is based on a null finding. Even though we boosted power in a pooled analysis, the p-value significance testing approach presented here cannot quantify a degree of support for the null hypothesis. To address this issue, we relied on Bayesian hypothesis testing using Bayes factors (e.g., Masson, 2011; Wagenmakers, 2007). Using the JASP package (JASP Team, 2017), we ran a 2 (Conflict: incorrect conflict, correct no-conflict) x 2 (Complexity: MP, MT) Bayesian ANOVA on response latencies for the valid inferences with default priors (e.g., Cauchy prior width $r = .707$). This showed that the model with a main effect of Conflict received the most support against the Null model ($BF_{10} = 1.52$). Adding the interaction with Complexity to the model decreased the degree of support against the Null

model ($BF_{10} = 0.138$). The model with a main effect of Conflict was thus preferred to the interaction model by a Bayes factor of 11.01. These data thus provide good evidence against the hypothesis that conflict detection is modulated by complexity on the MP and MT inferences (see Wetzels et al., 2011 for a classification of Bayes factors).

4.9 General Discussion

In these studies, we aimed to test the boundary conditions of logical intuitions (De Neys, 2012). To this end, we manipulated logical complexity and expected that conflict detection would be smaller for inferences of greater complexity. However, some of our results were unexpected and contrary to predictions.

The most unclear and surprising results concerned the invalid AC and DA inferences. First, our accuracy results indicated that the DA was not harder than the AC inference. We don't have a clear explanation for this finding, but these results were robust amongst both our studies and should clearly call for further investigation. One suggestion would be that the effect of negation on AC and DA inferences is less clear than for their valid counterparts, as previous studies have shown variable frequencies of endorsement for these inferences (see Evans, 1993 for a review).

We also failed to observe the expected conflict detection effect on the invalid AC form. Note that previous studies that reported successful conflict detection with simple AC and MP problems typically ran analyses in which performance over both inference types was collapsed (e.g., Bago & De Neys, 2017; De Neys et al., 2010; De Neys & Franssens, 2009). Hence, the present results suggest that these effects were primarily driven by the MP inference.

Our results on the valid MP and MT inferences did provide additional support for the general idea that biased reasoners can show sensitivity to the conflict between their heuristic response and logical principles. However, contrary to our predictions, we observed conflict detection for both the easy MP form and for the more difficult MT form. This leads to a number of interesting considerations. First, our basic manipulation of

problem complexity was the presence/absence of negations. While there is empirical support for the additional cognitive challenge that negation provides (Schaeken & Schroyens, 2000; Schroyens et al., 2000, 2001), it might not be strong enough for us to observe differences in conflict detection. This possibility would be congruent with the results of Trippas et al. (2017), who indicated in their supplementary material that they found no significant difference in conflict detection between the propositional versions of MP and MT, but still reported smaller conflict detection effects for syllogisms of greater complexity than MT. Combined with these results, our study suggests that MT might be simple enough for people to develop an intuitive grasp of its structure. With hindsight one could argue here that even though MT might be harder than MP, the MT inference is still a basic form of argument refutation and similar patterns (like the *reductio ad absurdum* proof) are taught in school. Consequently, through such repeated exposure, MT might still be simple and frequent enough to allow for the development of logical MT intuitions. Second, although the effects were not significant, our conflict detection results indicated that conflict detection was even more pronounced on MT than MP. Provided that conflict can be detected both intuitively and deliberately, one speculative possibility is that MT's slightly greater complexity triggered more deliberative reasoning processes, thus making reasoners more likely to detect conflict with these inferences. If this was the case, complexity and conflict detection could be linked by a reversed U-shape relation, thus making conflict detection optimal when problem complexity is intermediate and then dropping as complexity increases. Of course, further studies would be needed to investigate this possibility.

As explained above, our manipulation of conflict was collapsed with conclusion believability. While we extensively pretested our material and an additional control analysis showed that conflict effects were observed even when the believability status was matched, a potential believability confound can still be stressed as a possible limitation for our studies. Note, however, that this confound is inherent to the manipulation of belief with the categorical syllogisms used here. When crossing validity and believability the content of the categorical syllogisms needs to be altered so that a potential believability confound can never be ruled out completely. Future studies on the complexity question might thus adopt different types of problems, where the conflict manipulation allows a full

counterbalancing of the content material (e.g., base-rate problems, e.g., Pennycook et al., 2015). Moreover, since the logical intuition proposal goes beyond mere reasoning with categorical syllogisms, testing a wider range of basic logical principles would not only be methodologically but also theoretically relevant to clearly delineate its extents and limits.

In sum, it is important to stress that the present results do clearly not entail that there are no boundary conditions for logical intuitions. Indeed, the logical intuition proposal is a post hoc theoretical inference posited to account for observed empirical data (De Neys, 2012). Further studies will thus be needed in order to explore these boundaries. This study, however, has informed this question by indicating that they presumably reside in problems of greater complexity than the MT inference.

4.10 Intuition logique et modèle de l'apprenant

Tel qu'exposé dans l'article précédent, trois mesures de l'intuition logique ont été utilisées précédemment dans la littérature. Il s'agit des temps de réponse, de la conductivité de la peau (mesure de l'excitation) et des techniques IRM pour observer les régions du cerveau responsables de la gestion du conflit.

La mesure du temps de réponse est sans aucun doute la plus accessible et abordable qui puisse être ajoutée au modèle de l'apprenant de Muse-Logique. Pour ce faire, il faudrait d'abord ajouter les items pertinents. La banque d'items créé dans l'article précédent serait parfaitement adaptée à ce besoin. En effet, elle contient les quatre formes de syllogismes conflictuels et consistants pour l'implication: MPP, MTT, AC et DA avec conclusions croyables et incroyables et dont le contenu est contrebalancé. À partir de ces items, il serait possible de mesurer les temps de réponse des apprenants et de calculer la différence entre les réponses incorrectes aux problèmes conflictuels et les réponses correctes aux problèmes consistants. Une différence entre les moyennes de temps de réponse à un certain nombre d'item indiquerait une détection du conflit, et donc une sensibilité à la forme logique de l'inférence. Un autre calcul pourrait être fait pour vérifier si la tendance à inhiber les biais de croyance change au fur et à mesure que l'apprenant progresse dans le tutoriel. Avec la même banque de problèmes, il serait possible de calculer la différence entre les problèmes conflictuels réussis et les problèmes consistants réussis. De même, une différence entre les moyennes de temps de réponse à un certain nombre d'item indiquerait une inhibition du biais de croyance.

La Figure 4.1 ci-dessous expose une extension possible au RB de Muse-Logique afin de prendre ces paramètres en considération.

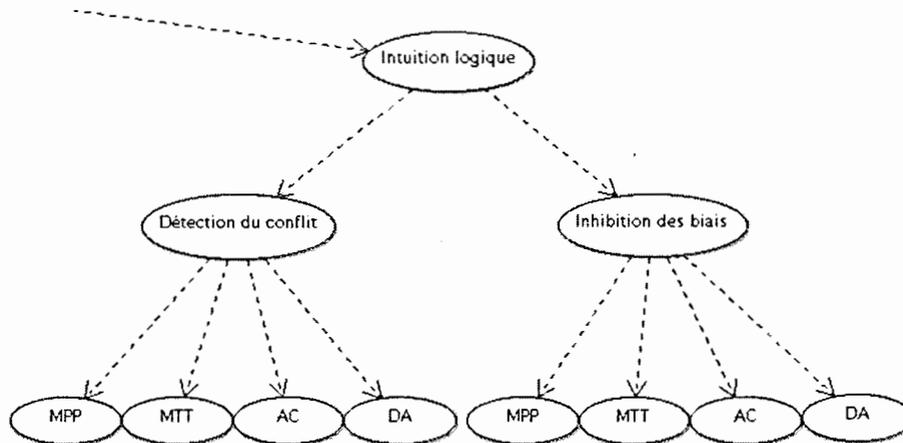


Figure 4.1 Portion du RB portant sur l'intuition logique

L'intuition logique étant un processus intuitif (rapide et inconscient), alors que les compétences identifiées jusqu'à maintenant concernent des processus analytiques (conscients, plus lents, contrôlés). Cette compétence devrait donc constituer une branche supplémentaire dans le RB, et ce, pour chacun des nœuds correspondants aux trois connecteurs logiques. La branche représentée dans la Figure 4.1 pourrait donc être ajoutée à chacune des trois branches principales du réseau (compétence à l'implication, à l'incompatibilité et à la disjonction inclusive).

CONCLUSION

Tout au long de cette thèse, nous avons rempli l'objectif suivant : construire un outil d'évaluation automatique du raisonnement déductif dans le cadre de la logique classique, et ce, à la lumière des résultats de la psychologie expérimentale du raisonnement.

Pour ce faire, au premier chapitre, nous avons identifié des compétences fondamentales au raisonnement logique à la lumière de deux perspectives théoriques mises en commun. Une approche formelle abordant le raisonnement déductif en logique classique (Robert & Brisson, 2016), combinée à une théorie descriptive du raisonnement humain (Markovits & Barouillet, 2002; Markovits, 2013), nous a permis d'identifier les compétences fondamentales quant au raisonnement logique : la reconnaissance des inférences certaines et incertaines à partir des connecteurs d'implication, de disjonction inclusive et d'incompatibilité.

Au second chapitre, nous avons présenté la conception et la validation d'un outil d'évaluation automatique du raisonnement déductif dans le cadre de l'apprentissage de la logique classique. Nous avons exposé la structure d'un réseau bayésien pour l'évaluation du raisonnement à partir d'une implication. Ce réseau est basé sur l'approche cognitive exposée au premier chapitre. Nous avons présenté une initialisation et ainsi qu'une validation empirique de ce RB à partir d'un *Cognitive Diagnosis Model* (CDM). Ce CDM a fourni un appui significatif au modèle psychologique sous-jacent au RB. Il a aussi permis un retour critique sur la gradation des niveaux de contenu initialement adoptée : le niveau contrefactuel semble plus difficile à maîtriser que le niveau abstrait au sein du raisonnement conditionnel. Nous avons aussi projeté les travaux futurs pour le RB. Dans un premier temps, il serait souhaitable de procéder à l'apprentissage automatique des distributions de probabilités pour la portion du RB qui concerne le raisonnement conditionnel. Dans un second temps, et suite aux expérimentations projetées, les extensions du RB pour le raisonnement à partir d'une incompatibilité et d'une disjonction inclusive devront être validées empiriquement.

Bien que les processus de raisonnement humain à partir d'une implication soient largement étudiés dans la littérature, le raisonnement à partir d'une incompatibilité n'a pas été étudié. Au troisième chapitre, nous avons donc présenté nos expérimentations à ce sujet. Les résultats ont corroboré les hypothèses avancées : la propension à reconnaître l'incertitude des inférences invalides à partir d'une incompatibilité croît avec le nombre de contre-exemples récupérables à partir de la mémoire à long terme. Nous avons ensuite présenté la portion du RB pouvant prendre en compte ce type de raisonnement et ses effets de contenu. Des expérimentations supplémentaires devront toutefois être menées afin d'explorer l'effet des conditions contraignantes sur les inférences valides. De plus, de possibles effets de contenu dans le cadre du raisonnement disjonctif devront être investigués.

Au quatrième et dernier chapitre, nous avons abordé l'effet de la forme logique sur le raisonnement et la possibilité de développer une compréhension intuitive de certains principes logiques de base. Nous avons présenté nos expérimentations concernant la portée et les limites de l'intuition logique dans le cadre du raisonnement conditionnel. Les résultats suggèrent la présence d'une intuition logique pour les deux inférences valides (MPP, MTT). Cependant, les résultats concernant les inférences invalides (AC, DA) sont peu probants. Des expérimentations supplémentaires devront être conduites afin de clarifier cette question. À la lumière de ces travaux ainsi que de la littérature actuelle, nous avons présenté un ajout possible à la structure du RB pour ce type de phénomène. Des expérimentations devront également être conduites concernant la possibilité d'une intuition logique au sein du raisonnement disjonctif et à partir d'une incompatibilité.

Lorsque complété, le Réseau bayésien de Muse-Logique sera en mesure de fournir une évaluation automatique du raisonnement déductif conditionnel, disjonctif et à partir d'une incompatibilité. Le STI Muse-Logique sera alors doté d'un modèle de l'apprenant complet et visant ce que nous considérons comme étant les compétences fondamentales au raisonnement déductif dans le cadre de la logique classique des propositions.

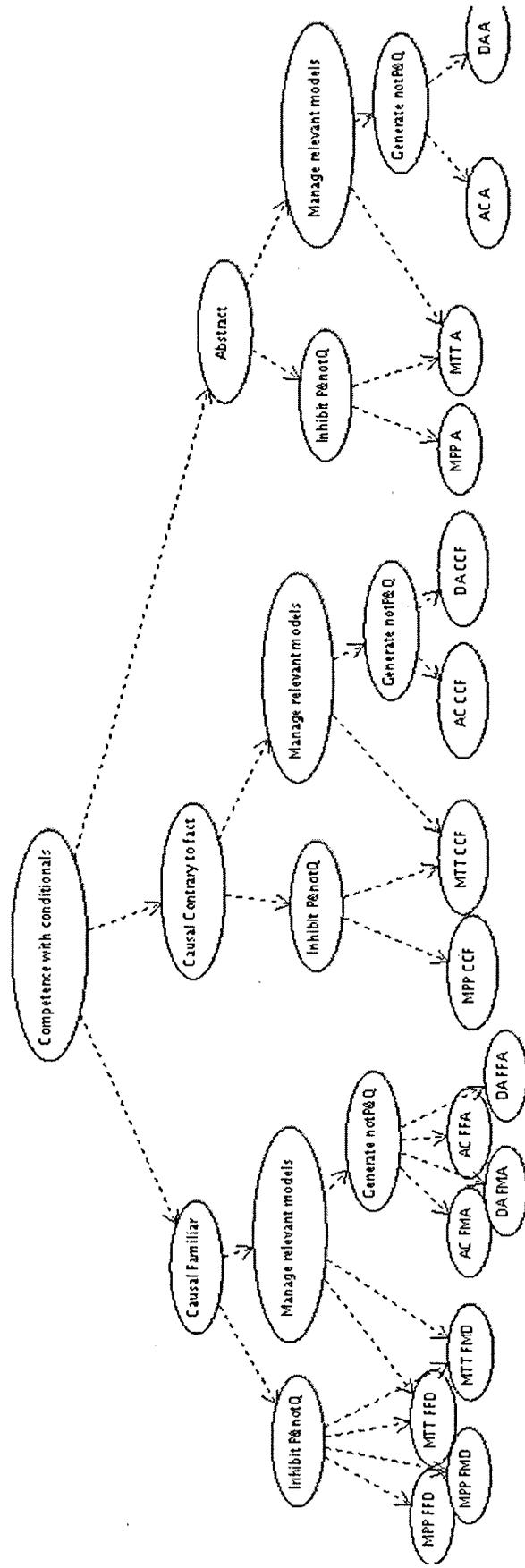
APPENDICE A

Tableau A.1 Q-matrix for conditional reasoning

	Causal Familiar			Causal Contrary to Fact			Abstract		
	Inhibit	Generate	Manage	Inhibit	Generate	Manage	Inhibit	Generate	Manage
MPP FFD	1	0	0	0	0	0	0	0	0
MPP FMD	1	0	0	0	0	0	0	0	0
MTT FFD	1	0	1	0	0	0	0	0	0
MTT FMD	1	0	1	0	0	0	0	0	0
AC FMA	0	1	1	0	0	0	0	0	0
AC FFA	0	1	1	0	0	0	0	0	0
DA FMA	0	1	1	0	0	0	0	0	0
DA FFA	0	1	1	0	0	0	0	0	0
MPP CCF	1	0	0	1	0	0	0	0	0
MTT CCF	1	0	1	1	0	1	0	0	0
AC CCF	0	1	1	0	1	1	0	0	0
DA CCF	0	1	1	0	1	1	0	0	0
MPP A	1	0	0	1	0	0	1	0	0
MTT A	1	0	1	1	0	1	1	0	1
ACA	0	1	1	0	1	1	0	1	1
DA A	0	1	1	0	1	1	0	1	1

APPENDICE B

Figure B.1 BN's structure for conditional reasoning



APPENDICE C

Tableau C.2 Complete item set for CDM

Familiar content		
<i>Few disablers, Many alternatives</i>		
Major premise 1		If one jumps in a pool, then one is wet.
1. MPP	Observation	Bob jumped in a pool.
	Valid conclusion	Bob is wet.
2. MTT	Observation	Paul it not wet.
	Conclusion	Paul did not jump in a pool.
3. AC	Observation	John is wet.
	Conclusion	One cannot conclude whether John jumped in a pool or not.
4. DA	Observation	Mark did not jump in a pool.
	Conclusion	One cannot conclude whether Mark is wet or not.
<i>Few disablers, few alternatives</i>		
Major premise 1		If a drum is hit, then it will sound.
5. MPP	Observation	The drum was hit.
	Conclusion	The drum will sound.
6. MTT	Observation	The drum did not sound.
	Conclusion	The drum was not hit.
7. AC	Observation	The drum sounded.
	Conclusion	One cannot conclude whether the drum was hit or not.
8. DA	Observation	The drum was not hit.
	Conclusion	One cannot conclude whether the drum will sound or not.
Major premise 2		If a dog has fleas, then it will scratch itself constantly.
9. MPP	Observation	Fido has fleas.
	Conclusion	Fido scratches itself constantly.
10.	Observation	Rex doesn't scratch itself constantly.
MTT	Conclusion	Rex doesn't have fleas
11. AC	Observation	Muzzle scratches itself constantly.
	Conclusion	One cannot conclude whether Muzzle has fleas or not.
12. DA	Observation	Buddy doesn't have fleas.
	Conclusion	One cannot conclude whether Buddy scratches itself constantly or not.
<i>Many disablers, many alternatives</i>		
Major premise 1		If the ignition key is turned, then the car will start.
13.	Observation	The ignition key was turned.
MPP	Conclusion	The car starts.
14.	Observation	The car is not started.
MTT	Conclusion	The ignition key was not turned.

15. AC	Observation Conclusion	The car starts. One cannot conclude whether the ignition key was turned or not.
16. DA	Observation Conclusion	The ignition key was not turned. One cannot conclude whether the car starts or not.
Major premise 2		If a rock is thrown at a window, then the window will break.
17. MPP	Observation Conclusion	A rock was thrown at a window. The window is broken.
18. MTT	Observation Conclusion	The window is not broken. No rock was thrown at the window.
19. AC	Observation Conclusion	The window is broken. One cannot conclude whether a rock was thrown at the window or not.
20. DA	Observation Conclusion	No rock was thrown at a window. One cannot conclude whether the window is broken or not.
<i>Many disablers, few alternatives</i>		
Major premise 1		If a match is scraped, then it will light up.
21. MPP	Observation Conclusion	A match was scraped. The match will light up.
22. MTT	Observation Conclusion	The match is not lit. The match was not scraped.
23. AC	Observation Conclusion	The match is lit. One cannot conclude whether the match was scraped or not.
24. DA	Observation Conclusion	The match was not scraped. One cannot conclude whether the match is lit or not.
Counterfactual content		
Major premise 1		If ketchup is thrown on a shirt, then the shirt will be clean.
25. MPP	Observation Conclusion	Luke threw ketchup on his shirt. Luke's shirt is clean.
26. MTT	Observation Conclusion	Mary's shirt is not clean. Mary did not throw ketchup on her shirt.
27. AC	Observation Conclusion	Peter's shirt is clean. One cannot conclude whether Peter threw ketchup on his shirt or not.
28. DA	Observation Conclusion	Lucy did not throw ketchup on her shirt. One cannot conclude whether Lucy's shirt is clean or not.
Major premise 2		If one brushes one's teeth regularly, then one will have cavities.
29. MPP	Observation Conclusion	Vince brushes his teeth regularly. Vince has cavities.
30. MTT	Observation Conclusion	Sarah does not have cavities. Sarah does not brush her teeth regularly.

31. AC	Observation	Roxane has cavities.
	Conclusion	One cannot conclude whether Roxane brushes her teeth regularly or not.
32. DA	Observation	Max does not brush his teeth regularly.
	Conclusion	One cannot conclude whether Max has cavities or not.
Major premise 3		If one is hot, then one's teeth will chatter.
33.	Observation	Simon is hot.
MPP	Conclusion	Simon's teeth are chattering.
34.	Observation	Nathalie's teeth are not chattering.
MTT	Conclusion	Nathalie is not hot.
35. AC	Observation	Judy's teeth are chattering.
	Conclusion	One cannot conclude whether Judy is hot or not.
36. DA	Observation	William is not hot.
	Conclusion	One cannot conclude whether William's teeth are chattering or not.

Abstract content

Major premise 1		If one blops, then one will become plede.
37.	Observation	Mary-Ann bloped.
MPP	Conclusion	Mary-Ann will become plede.
38.	Observation	Frank did not become plede.
MTT	Conclusion	Frank did not blop.
39. AC	Observation	Michelle became plede.
	Conclusion	One cannot conclude whether Michelle bloped or not.
40. DA	Observation	Peter did not blop.
	Conclusion	One cannot conclude whether Peter will become plede or not.
Major premise 2		If it burdons, then the ground will be brandup.
41.	Observation	It burdons.
MPP	Conclusion	The ground is brandup
42.	Observation	The ground is not brandup.
MTT	Conclusion	It does not burdon.
43. AC	Observation	The ground is brandup.
	Conclusion	One cannot conclude whether it burdons or not.
44. DA	Observation	It doesn't burdon.
	Conclusion	One cannot conclude whether the ground will be brandup or not.
Major premise 3		If a fymolyte is touched, then it will become mandola.
45.	Observation	A fymolyte was touched.
MPP	Conclusion	The fymotyle will become mandola.
46.	Observation	A fymolyte did not become mandola.
MTT	Conclusion	The fymotyle was not touched.
47. AC	Observation	A fymolyte became mandola.
	Conclusion	One cannot conclude whether the the fymolyte was touched or not.
48. DA	Observation	A fymolyte was not touched.

Conclusion One cannot conclude whether the the fymolyte will become
mandola or not.

APPENDICE D

Tableau D.3 Mean number of Third options and Exceptions generated (maximum of 5) for each premise (translated from the original French) by Type (Many, Few)

Many Third options	Third options	Exceptions
Having a car accident is incompatible with being on time at a meeting. Conditional version: If one has a car accident, then one will be late at a meeting.	4.00	1.78
Being on a plane is incompatible with sleeping well. Conditional version: If one is on a plane, then one will have poor sleep.	3.75	1.89
Drinking coffee in the evening is incompatible with falling asleep easily. Conditional version: If one is drinks coffee in the evening, then it will be harder to fall asleep.	3.58	1.89
Few Third options	Third options	Exceptions
Drinking a lot of alcohol is incompatible with being sober. Conditional version: If one drinks a lot of alcohol, then one will be drunk.	2.42	1.11
Being in a warm environment is incompatible with feeling cold. Conditional version: If one is in a warm environment, then one will feel warm.	2.5	1.78
A misplaced rail is incompatible with the train staying on track. Conditional version: If a rail is misplaced, then the train will derail.	2.5	1.44

APPENDICE E

Tableau E.4 Mean number of Third options generated for each parent category and its base-level categories (translated from the original French) by Type (Broad, Reduced, Close to binary)

Broad		
Parent category	Base-level categories	Third options
Vegetables	Broccoli; Pepper	8.41
Animals	Dog; Horse	10.71
World cities	Montreal; Paris	16.47
Fruits	Grape; Strawberry	9.18
Reduced		
Parent category	Base-level categories	Third options
Root vegetables	Potato; Carrot	2.29
Polar animals	Polar bear; Penguin	2.25
German cities	Berlin; Hamburg	1.07
Fruits with pits	Peach; Cherry	2.59
Dessert of the day in a restaurant	Chocolate cake; Lemon pie	5.65
Close to binary		
Parent category	Base-level categories	Third options
Bet on "Heads and tails"	Heads; Tails	0.75
Paths at a fork (road)	Right; Left	1.18
Bet on "Even, Odd" dice game	Even; Odd	0.33
Vote on a bill	For; Against	1.11

APPENDICE F

Tableau F.5 Item sets for logical intuition

Item set A	Item set B	Item set C	Item set D
All mammals can walk Whales are mammals Whales can walk (MP Conflict)	All mammals can walk Birds can walk Birds are mammals (AC No-conflict)	All mammals can walk Whales cannot walk Whales are not mammals (MT Conflict)	All mammals can walk Birds are not mammals Birds cannot walk (DA No-conflict)
All flowers need water Roses are flowers Roses need water (MP No-conflict)	All flowers need water Roses need water Roses are flowers (AC Conflict)	All flowers need water Rocks don't need water Rocks are not flowers (MT No-conflict)	All flowers need water Rocks are not flowers Rocks don't need water (DA Conflict)
All trees have roots Oaks have roots Oaks are trees (AC Conflict)	All trees have roots Oaks are trees Oaks have roots (MP No-conflict)	All trees have roots Apples are not trees Apples don't have roots (DA Conflict)	All trees have roots Apples don't have roots Apples are not trees (MT No-conflict)
All vehicles have wheels Trolley suitcases have wheels Trolley suitcases are vehicles (AC No-conflict)	All vehicles have wheels Trolley suitcases are not vehicles Trolley suitcases don't have wheels (DA No-conflict)	All vehicles have wheels Boats are vehicles Boats have wheels (MP Conflict)	All vehicles have wheels Boats don't have wheels Boats are not vehicles (MT Conflict)

Item sets for logical intuition (continued)

Item set A	Item set B	Item set C	Item set D
All birds can fly Penguins cannot fly Penguins are not birds (MT Conflict)	All birds can fly Planes are not birds Planes cannot fly (DA No-conflict)	All birds can fly Planes can fly Planes are birds (AC No-conflict)	All birds can fly Penguins are birds Penguins can fly (MP conflict)
All African countries are warm The North Pole is not warm The North Pole is not an African country (MT No-conflict)	All African countries are warm The North Pole is not an African country The North Pole is not warm (DA conflict)	All African countries are warm Congo is an African country Congo is warm (MP No-conflict)	All African countries are warm Congo is warm Congo is an African country (AC conflict)
All fruits can be eaten Knives are not fruits Knives cannot be eaten (DA conflict)	All fruits can be eaten Knives cannot be eaten Knives are not fruits (MT No-conflict)	All fruits can be eaten Strawberries can be eaten Strawberries are fruits (AC conflict)	All fruits can be eaten Strawberries are fruits Strawberries can be eaten (MP No-conflict)
All metals are solid Ceramic not a metal Ceramic is not solid (DA No-conflict)	All metals are solid Mercury is not solid Mercury is not a metal (MT conflict)	All metals are solid Mercury is a metal Mercury is solid (MP conflict)	All metals are solid Ceramic is solid Ceramic is a metal (AC No-conflict)

Item sets for logical intuition (continued)

Item set A	Item set B	Item set C	Item set D
All trees will become tall Bonsai are trees Bonsai will become tall (MP Conflict)	All trees will become tall Skyscrapers under construction will become tall Skyscrapers under construction are trees (AC No-conflict)	All trees will become tall Bonsai won't become tall Bonsai are not trees (MT Conflict)	All trees will become tall Skyscrapers under construction are not trees Skyscrapers under construction won't become tall (DA No-conflict)
All dogs have snouts Labradors are dogs Labradors have snouts (MP No-conflict)	All dogs have snouts Labradors have snouts Labradors are dogs (AC Conflict)	All dogs have snouts Ducks don't have snouts Ducks are not dogs (MT No-conflict)	All dogs have snouts Ducks are not dogs Ducks don't have snouts (DA Conflict)
All things made of wood can be used as fuel Trees can be used as fuel Trees are made of wood (AC Conflict)	All things made of wood can be used as fuel Trees are made of wood Trees can be used as fuel (MP No-conflict)	All things made of wood can be used as fuel Glass is not made of wood Glass cannot be used as fuel (DA Conflict)	All things made of wood can be used as fuel Glass cannot be used as fuel Glass is not made of wood (MT No-conflict)
All things made of metal shine Diamonds shine Diamonds are made of metal (AC No-conflict)	All things made of metal shine Old pennies don't shine Old pennies are not made of metal (MT Conflict)	All things made of metal shine Diamonds are not made of metal Diamonds don't shine (DA No-conflict)	All things made of metal shine Old pennies are made of metal Old pennies shine (MP Conflict)

Item sets for logical intuition (continued)

Item set A	Item set B	Item set C	Item set D
All humans have two legs	All humans have two legs	All humans have two legs	All humans have two legs
Leg amputees don't have two legs	Leg amputees are humans	Monkeys are not humans	Monkeys have two legs
Leg amputees are not humans (MT Conflict)	Leg amputees have two legs (MP Conflict)	Monkeys don't have two legs (DA No-conflict)	Monkeys are humans (AC No-conflict)
All things that are smoked are bad for your health	All things that are smoked are bad for your health	All things that are smoked are bad for your health	All things that are smoked are bad for your health
Salad is not smoked	Salad is not bad for your health	Cigarettes are smoked	Cigarettes are bad for your health
Salad is not bad for your health (DA Conflict)	Salad is not smoked (MT No-conflict)	Cigarettes are bad for your health (MP No-conflict)	Cigarettes are smoked (AC Conflict)
All sports require equipment	All sports require equipment	All sports require equipment	All sports require equipment
Talking does not require equipment	Talking is not a sport	Hockey requires equipment	Hockey is a sport
Talking is not a sport (MT No-conflict)	Talking does not require equipment (DA Conflict)	Hockey is a sport (AC Conflict)	Hockey requires equipment (MP No-conflict)
All things that need oxygen have lungs	All things that need oxygen have lungs	All things that need oxygen have lungs	All things that need oxygen have lungs
Dead people don't need oxygen	Fire needs oxygen	Dead people have lungs	Fire doesn't have lungs
Dead people don't have lungs (DA No-conflict)	Fire has lungs (MP Conflict)	Dead people need oxygen (AC No-conflict)	Fire doesn't need oxygen (MT Conflict)

APPENDICE G

Tableau G.6 Mean conclusion believability (on a scale of 0 to 10) for each conclusion

Item set A	CB	Item set B	CB	Item set C	CB	Item set D	CB
Whales can walk	0	Birds are mammals	0	Whales are mammals	0	Birds cannot walk	0.67
Roses need water	10	Roses are flowers	10	Rocks are not flowers	10	Rocks don't need water	9.83
Oaks are trees	9.17	Oaks have roots	10	Apples don't have roots	8.67	Apples are not trees	8.92
Trolley suitcases are vehicles	1.83	Trolley suitcases don't have wheels	1.58	Boats have wheels	1.67	Boats are not vehicles	0.83
Penguins are not birds	2.5	Planes cannot fly	1.17	Planes are birds	0.83	Penguins can fly	1.75
The North Pole is not an African country	8.42	The North Pole is not warm	9.67	Congo is warm	9.75	Congo is an African country	10
Knives cannot be eaten	7.17	Knives are not fruits	9.08	Strawberries are fruits	9.5	Strawberries can be eaten	9.67
Ceramic is not solid	0.42	Mercury is not a metal	3	Mercury is solid	1.33	Ceramic is a metal	0.25

Mean conclusion believability (on a scale of 0 to 10) for each conclusion (continued)

Item set A	CB	Item set B	CB	Item set C	CB	Item set D	CB
Bonsai will become tall	1.25	Skyscrapers construction are trees	0.17	Bonsai are not trees	0.08	Skyscrapers construction won't become tall	under 2
Labradors have snouts	8.67	Labradors are dogs	9.83	Ducks are not dogs	7.5	Ducks don't have snouts	7.17
Trees are made of wood	9.42	Trees can be used as fuel	8.83	Glass cannot be used as fuel	8.25	Glass is not made of wood	9.67
Diamonds are made of metal	1	Old pennies are not mat of metal	0.83	Diamonds don't shine	2.75	Old pennies shine	3.08
Leg amputees are not humans	0.42	Leg amputees have two legs	2.5	Monkeys don't have two legs	1.5	Monkeys are humans	0.92
Salad is not bad for your health	9.75	Salad is not smoked	7.42	Cigarettes are bad for your health	9.83	Cigarettes are smoked	8.08
Talking is not a sport	8.86	Talking does require equipment	not 7.86	Hockey is a sport	9.92	Hockey requires equipment	9.58
Dead people have lungs	1.83	Fire has lungs	0.08	Dead people need oxygen	0.42	Fire doesn't need oxygen	0.08

BIBLIOGRAPHIE

- Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. *Cognition*, 158, 90-109. <http://dx.doi.org/10.1016/j.cognition.2016.10.014>
- Baratgin, J., Over, D. E., & Politzer, G. (2013). Uncertainty and the de Finetti tables.
- Barnes, T., et Stamper, J. (2010). Automatic Hint Generation for Logic Proof Tutoring Using Historical Data. *Educational Technology & Society*, 13 (1), 3–12.
- Brisson, J., Markovits, H., Robert, S., & Schaeken, W. (2018). Reasoning from an incompatibility: False dilemma fallacies and content effects. *Memory & cognition*, 46(5), 657-670.
- Brisson, J., Schaeken, W., Markovits, H., & De Neys, W. (2018). Conflict detection and logical complexity. *Psychologica Belgica*, 58(1), 318.
- Conati, C. 2010. Bayesian student modeling. In R Nkambou, R. Mizoguchi & J.Bourdeau (dir), *Advances in intelligent tutoring systems*, Berlin : Springer, 281-299.
- Conati, C., Gertner, A. & VanLehn, K. 2002. Using Bayesian Networks to Manage Uncertainty in Student Modeling. *User Modeling and User-Adapted Interaction*, 12(4), 371-417.
- Cummins, D. D. (1995). Naive theories and causal deduction. *Memory & Cognition*, 23 (5), 646–658.
- Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & cognition*, 19 (3), 274–282.
- De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science*, 7(1), 28-38. <http://dx.doi.org/10.1177/1745691611429354>

- De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning*, 20(2), 169-187. <http://dx.doi.org/10.1080/13546783.2013.854725>
- De Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in doubt: Conflict and decision confidence. *PloS one*, 6(1), e15954. <https://doi.org/10.1371/annotation/1ebd8050-5513-426f-8399-201773755683>
- De Neys, W., Schaeken, W., & D'ydwalle, G. (2002). Causal conditional reasoning and semantic memory retrieval: A test of the semantic memory framework. *Memory & cognition*, 30 (6), 908–920.
- De Neys, W., Schaeken, W., & d'Ydwalle, G. (2003). Causal conditional reasoning and strength of association: The disabling condition case. *European Journal of Cognitive Psychology*, 15 (2), 161–176.
- Do, C., & Batzoglou, S. 2008. What is the expectation maximization algorithm? *Nature Biotechnology*, 26(8), 897–9.
- Evans, J. St. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, 11(3), 295-306. doi : 10.3758/BF03196976
- Evans, J. S. B. T., & Over, D. E. (2004). Oxford cognitive science series. If. New York, NY, US: Oxford University Press.
- Evans, J. St. B., Over, D. E., & Handley, S. J. (2005). Suppositionals, extensionality, and conditionals: A critique of the mental model theory of Johnson-Laird and Byrne (2002). *Psychological Review*, 112(4), 1040–1052. doi:10.1037/0033-295X.112.4.1040
- Handley, S. J., & Trippas, D. (2015). Chapter Two-Dual processes and the interplay between knowledge and structure: a new parallel processing model. *Psychology of learning and motivation*, 62, 33-58.
- Hirt, E. R., Kardes, F. R., & Markman, K. D. (2004). Activating a mental simulation mind-set through generation of alternatives: Implications for debiasing in related and unrelated domains. *Journal of Experimental Social Psychology*, 40(3), 374-383. doi : 10.1016/j.jesp.2003.07.009
- Janveau-Brennan, G., & Markovits, H. (1999). The development of reasoning with causal conditionals. *Developmental Psychology*, 35(4), 904-911. doi : 10.1037/0012-1649.35.4.904

- Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in cognitive sciences*, 5(10), 434-442.
- Johnson-Laird, P., & Byrne, R.M.J. (1991). *Deduction*. Hillsdale, NJ : Lawrence Erlbaum Associates.
- Johnson-Laird, P., & Byrne, R.M.J. (2002). Conditionals : A theory of meaning, pragmatics and inference. *Psychological Review*, 109, 646-678. doi : 10.1037/0033-295X.109.4.646
- Kenfack, C., Nkambou, R., Robert, S., Tato, A. A. N., Brisson, J., & Kissok, P. (2016). A Brief Overview of Logic-Muse, an Intelligent Tutoring System for Logical Reasoning Skills. In *Intelligent Tutoring Systems* (p. 511).
- Lesta, L. et K. Yacef (2002). "An Intelligent Teaching-Assistant System for Logic". *Proceedings of International Conference on Intelligent Tutoring Systems (ITS'02)* S. Cerri et F. Paraguo (eds) Biarritz, France, Springer-Verlag.
- Mayo, M., & Mitrovic, A. (2001). Optimising ITS behaviour with Bayesian networks and decision theory.
- Markovits, H. (1985). Incorrect conditional reasoning among adults: Competence or performance? *British Journal of Psychology*, 76(2), 241-247. doi : 10.1111/j.2044-8295.1985.tb01948.x
- Markovits, H. (2000). A mental model analysis of young children's conditional reasoning with meaningful premises. *Thinking & Reasoning*, 6(4), 335-347. doi : 10.1080/135467800750038166
- Markovits, H. (Éd.). (2013). *The developmental psychology of reasoning and decisionmaking*. New York, NY: Psychology Press.
- Markovits, H. (2013). How to develop a logical reasoner: A hierarchical model of the role of divergent thinking in the development of conditional reasoning. In *The developmental psychology of reasoning and decision-making* (pp. 156-172). Psychology Press.
- Markovits, H., & Barrouillet, P. (2002). The development of conditional reasoning: A mental model account. *Developmental Review*, 22(1), 5-36. doi : 10.1006/drev.2000.0533
- Markovits, H., Fleury, M. L., Quinn, S., & Venet, M. (1998). The development of conditional reasoning and the structure of semantic memory. *Child development*, 69(3), 742-755. doi : 10.1111/j.1467-8624.1998.tb06240.x

- Markovits, H., & Quinn, S. (2002). Efficiency of retrieval correlates with “logical” reasoning from causal conditional premises. *Memory & Cognition*, 30(5), 696-706. doi : 10.3758/BF03196426
- Markovits, H., & Vachon, R. (1990). Conditional reasoning, representation, and level of abstraction. *Developmental Psychology*, 26(6), 942-951. doi : 10.1037/0012-1649.26.6.942
- Markovits, Venet, Janveau-Brennan, Malfait, Pion, & Vadeboncoeur, 1996
- Mitchell, T. 1997. *Machine Learning*. MacGraw-Hill Companies. Inc, 1, 1997.
- Naïm, P., Wuillemin, P.-H., Leray, P., Pourret, O. & Becker, A. 2007. Réseaux Bayésiens. Paris : Eyrolles.
- Newman, I. R., Gibb, M., & Thompson, V. A. (2017). Rule-based reasoning is fast and belief-based reasoning can be slow: Challenging current explanations of belief-bias and base-rate neglect. *Journal of experimental psychology: learning, memory, and cognition*, 43(7), 1154. <http://dx.doi.org/10.1037/xlm0000372>
- Nkambou, R., Brisson, J., Kenfack, C., Robert, S., Kissok, P., & Tato, A. (2015). Towards an intelligent tutoring system for logical reasoning in multiple contexts. In *Design for Teaching and Learning in a Networked World* (pp. 460-466). Springer, Cham.
- Nkambou, R., Kenfack, C., Robert, S., & Brisson, J. (2015, June). The Design Rationale of Logic-Muse, an ITS for Logical Reasoning in Multiple Contexts. In *International Conference on Artificial Intelligence in Education* (pp. 738-742). Springer, Cham.
- Nkambou, R., Tato, A. A. N., Brisson, J., Kenfack, C., Robert, S., & Kissok, P. (2016). On the Evaluation of the Expert and the Learner Models of Logic-Muse Tutoring System. In *Intelligent Tutoring Systems* (p. 506). Springer.
- Oakhill, J., Johnson-Laird, P. N., & Garnham, A. (1989). Believability and syllogistic reasoning. *Cognition*, 31(2), 117-140. doi : 10.1016/0010-0277(89)90020-6
- Oaksford, M., & N. Chater. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences*, 5, 349-357. doi : 10.1016/S1364-6613(00)01699-5
- Oaksford, M., & N. Chater. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford: Oxford University Press.

- Over, D. E. (2009). New paradigm psychology of reasoning. *Thinking and Reasoning*, 15(4), 431- 438. doi : 10.1080/13546780903266188
- Pennycook, G., Trippas, D., Handley, S. J., & Thompson, V. A. (2014). Base rates: both neglected and intuitive. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(2), 544. <http://dx.doi.org/10.1037/a0034887>
- Pfeiffer, N. (2013). The new psychology of reasoning: A mental probability logical perspective. *Thinking and Reasoning*, 19(3-4), 329-345. doi : 10.1080/13546783.2013.838189.
- Quinn, S. & Markovits, H. (1998). Conditional reasoning, causality, and the structure of semantic memory: Strength of association as a predictive factor for content effects. *Cognition*, 68 (3), B93-B101.
- Robert, S., & Brisson, J. (2016). The Klein group, squares of opposition and the explanation of fallacies in reasoning. *Logica Universalis*, 10(2-3), 377-392.
- Tato, A., Nkambou, R., Brisson, J., Kenfack, C., Robert, S., & Kissok, P. (2016, September). A Bayesian Network for the Cognitive Diagnosis of Deductive Reasoning. In *European Conference on Technology Enhanced Learning* (pp. 627-631). Springer, Cham.
- Tato, A., Nkambou, R., Brisson, J., & Robert, S. (2017, June). Predicting Learner's Deductive Reasoning Skills Using a Bayesian Network. In *International Conference on Artificial Intelligence in Education* (pp. 381-392). Springer, Cham.
- Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory & Cognition*, 22(6), 742-758. doi : 10.3758/BF03209259
- Thompson, V. A. (1995). Conditional reasoning: The necessary and sufficient conditions. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 49(1), 1-60. doi : 10.1037/1196-1961.49.1.1
- Trippas, D., Handley, S. J., Verde, M. F., & Morsanyi, K. (2016). Logic brightens my day: Evidence for implicit sensitivity to logical validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(9), 1448. <http://dx.doi.org/10.1037/xlm0000248>
- Trippas, D., Thompson, V. A., & Handley, S. J. (2017). When fast logic meets slow belief: Evidence for a parallel-processing model of belief bias. *Memory & cognition*, 45(4), 539. <http://dx.doi.org/10.3758/s13421-016-0680-1>