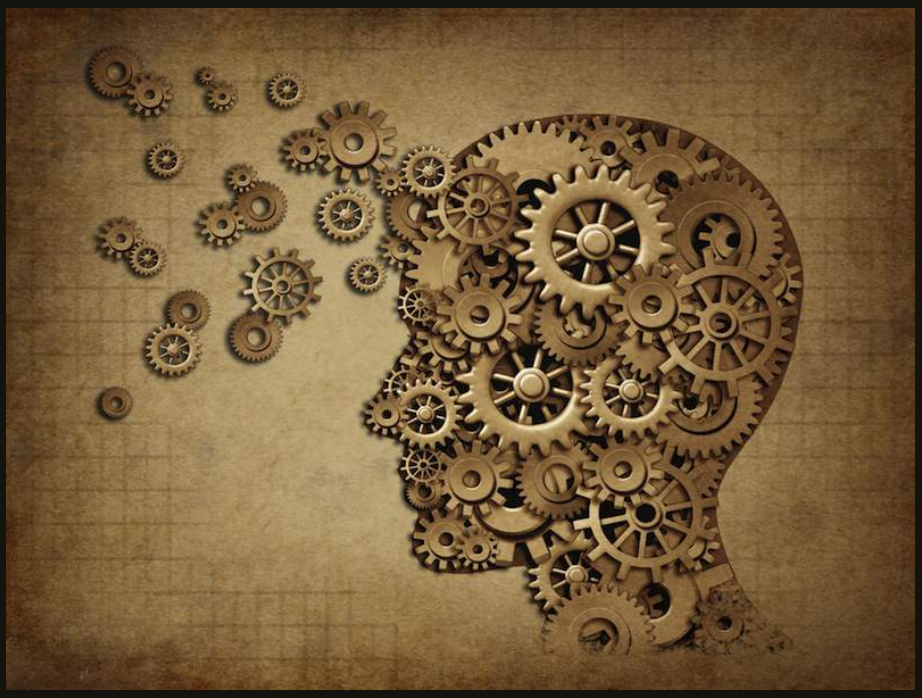
Logic and Commonsense Reasoning

Guillaume Aucher

Guillaume Aucher



Logic and Commonsense Reasoning

Lecture Notes

Master, Philosophy

Rennes, 2016 – 2017

Picture taken from http://epilepsyu.com/wp-content/uploads/2014/04/brain-mechanism.jpg

Contents

General Introduction **1**

I A Logic Compendium **5**

Introduction **7**

1 Logic: Basic Concepts **9**
1.1 Introduction 9
1.2 A Short History of Logic 10
1.2.1 The Origins of Logic 10
1.2.2 Second Age: Mathematical Logic (late 19^{th} to mid 20^{th} Century) . 11
1.2.3 Third Age: Logic in Computer Science (mid 20^{th} Century to Now) 12
1.3 Propositional, Modal and First-Order Logic 15
1.3.1 Propositional Logic (PL) 15
1.3.2 Modal Logic (ML) 16
1.3.3 First-order Logic (FO) 19
1.3.4 Truth, Logical Consequence, Validity, Satisfiability 22
1.4 Axioms, Inference Rules and Completeness 22
1.4.1 Deductive Calculus 22
1.4.2 Axiomatizing the Validities of PL, ML and FO 23
1.4.3 Increasing the Deductive Power of Modal Logics 25
1.5 List of Logics 26
1.6 Further Reading 28

2 Decidability, Complexity and Expressiveness **29**
2.1 Introduction 29
2.2 Decidability 29
2.2.1 Tableau Method: a Decision Procedure 30
2.2.2 Finite Model Property 31
2.3 Expressive Power and Invariance 35
2.3.1 First-order Logic: Ehrenfeucht-Fraïssé games 35

2.3.2 Modal Logic: Bisimulation Games . . . 38
2.3.3 Using Model Comparison Games: Definability and Expressiveness 39
2.4 Computation and Complexity . . . 41
2.5 Further Reading . . . 42

II Representing and Reasoning about Uncertainty 43

Introduction 45

3 Reasoning Alone about Uncertainty 47
3.1 Introduction . . . 47
3.2 Representing Uncertainty . . . 48
3.2.1 Probability Measures . . . 48
3.2.2 Dempster-Shafer Belief Functions . . . 50
3.2.3 Possibility Measures . . . 52
3.2.4 Ranking Function . . . 53
3.2.5 Preferential Structures . . . 54
3.2.6 Justifying Probability Numbers . . . 54
3.2.7 Choosing a Representation . . . 56
3.3 Reasoning about Uncertainty . . . 56
3.3.1 Plausibility Measures: an Abstract Framework . . . 56
3.3.2 Logics for Quantitative Reasoning . . . 58
3.3.3 Logics for Qualitative Reasoning . . . 59
3.4 Updating Uncertainty . . . 60
3.4.1 Conditioning Probabilities . . . 60
3.4.2 Conditioning Sets of Probabilities, Inner and Outer Measures . . . 62
3.4.3 Conditioning Belief Functions . . . 65
3.4.4 Conditioning Possibility Measures and Ranking Functions . . . 66
3.4.5 Jeffrey's Rule . . . 66
3.5 Further Reading . . . 68

4 Reasoning with Others about Uncertainty 69
4.1 Introduction . . . 69
4.2 Representing and Reasoning about Uncertainty: Epistemic Logic . . . 72
4.2.1 Syntax and Semantics . . . 73
4.2.2 Axiomatization . . . 77
4.2.3 Decidability . . . 79
4.3 Updating Uncertainty: Dynamic Epistemic Logic . . . 79
4.3.1 Public Events: Public Annoucement Logic . . . 80
4.3.2 Arbitrary Events: Event Model and Product Update . . . 83
4.3.3 A General Language . . . 86
4.4 Further Reading . . . 87

III Commonsense Reasoning **89**

Introduction **91**

5 Conditionals **93**
5.1 Introduction . . . 93
5.2 The Problem . . . 94
5.3 Truth-functional Semantics: Material Implication . . . 95
5.4 Modal Semantics: Strict Conditional . . . 97
5.5 Selection Functions . . . 98
5.6 Systems of Spheres . . . 99
5.7 From Systems of Spheres to Selection Functions . . . 100
5.8 Other Semantics . . . 101
5.9 Some Familiar Conditional Logics . . . 102
5.9.1 Selection Functions Semantics . . . 102
5.9.2 Systems of Spheres Semantics . . . 102
5.10 Proof Systems for Conditionals . . . 103
5.10.1 Tableaux Methods . . . 103
5.10.2 Hilbert Systems . . . 105
5.11 Further Reading . . . 106

6 Default Reasoning **107**
6.1 Introduction . . . 107
6.2 Logics for Defaults . . . 108
6.3 Proof System P . . . 109
6.4 From Defaults to Conditionals . . . 111
6.5 From Conditionals to Counterfactuals . . . 112
6.6 Further reading . . . 114

7 Belief Revision **115**
7.1 Introduction . . . 115
7.2 Expansion . . . 116
7.3 Revision . . . 118
7.4 "Two Sides of the Same Coin" . . . 120
7.5 Further Reading . . . 121

IV Appendix **123**

A Set Theory: Basic Notions and Notations **125**
A.1 Sets and Elements . . . 125
A.2 Defining Sets . . . 125
A.2.1 Lists of Elements . . . 125
A.2.2 Elements Satisfying a Condition . . . 126

A.2.3 Elements of a Given Form 126
A.3 Equality and Subsets 127
A.4 Operations: Union, Intersection, Difference 127
A.5 Functions 128
A.6 Relations and Cartesian Product 129
A.7 Further Reading 129

Bibliography **130**

Index **139**

General Introduction

> *"L'idée de la logique pratique, logique en soi, sans réflexion consciente ni contrôle logique, est une* contradiction dans les termes*, qui défie la logique logique."*
> – Pierre Bourdieu, *Le sens pratique*, 1980

For a very long time, logic was not formalized and the various patterns of reasonings studied were expressed in natural language. The formalization of logic began in the nineteenth century as mathematicians attempted to clarify the foundations of mathematics. Around that time, the formal languages of propositional logic and first-order logic were developed by Boole and then Frege. Frege's primary concern was to construct a logical system, formulated in an idealized language called *Begriffsschrift*, which was adequate for mathematical reasoning (Frege, 1879). The connective $\varphi \to \psi$, usually called *material implication*, played an essential role in his Begriffsschrift and Frege defined this connective formally. It was taken up enthusiastically by Russell, Wittgenstein and the logical positivists, and it is now found in every logic textbook. Later on, and in order to capture even more faithfully the actual reasoning performed by mathematicians, specific proof systems were developed by Gentzen (1935): the so-called sequent calculi for natural deduction.

Frege's material implication $\varphi \to \psi$ is true if, and only if, φ is false or ψ is true. The truth-functional feature of propositional and predicate logics was essential to Frege's approach. If we want the conditional "if φ, then ψ" to be truth-functional, this is the right truth function to assign to it: of the sixteen possible truth-functions of φ and ψ, it is the only serious candidate. It is sometimes told in a first course in logic that conditionals may be represented as the material implication $\to$. However, there are some obvious objections to this claim. According to the definition of material implication, if ψ is true then $\varphi \to \psi$ is also true, and if φ is false then $\varphi \to \psi$ is true. So, in that case, the following statements would be true, although they definitely appear to be dubious:

> If Rennes is in the Netherlands then 2+2=4.
>
> If Rennes is in France then World War II ended in 1945.
>
> If World War II ended in 1941 then gold is an acid.

For the purpose of doing mathematics, Frege's proposal to interpret conditionals as material implications was probably correct. The main defects of material implication do not show up in mathematics: mathematical inference is such that it never depends

on unlikely or doubtful premises. There are obviously some peculiarities, but as long as we are aware of them, they can be lived with. And arguably, the gain in simplicity and clarity more than offsets the oddities. They are harder to tolerate when we consider conditional statements about matters dealing with everyday life. The difference is that in reasoning about the world, we often accept and reject propositions with degrees of confidence less than certainty. The kind of statement "I think, but am not sure, that φ" plays no central role in mathematical thinking. In everyday life, we often use conditionals whose antecedent we think is likely to be false. Still, we are nevertheless able to make sound and plausible inferences based on these uncertain premises. In fact, in everyday life the way we update and infer information is quite different from the actual reasoning of mathematicians.

This observation has lead philosophers and researchers in artificial intelligence and computer science from the 1960s on to develop logical theories that study and formalize the so-called "commonsense reasoning". The rationale underlying the development of such theories was that it would ultimately help us understand our everyday life reasoning and the way we update our beliefs. For computer scientists, the resulting work could subsequently lead to the development of tools that could be used for example by artificial agents in order to act autonomously in an uncertain and changing world like internet or the real world. A number of theories have been proposed to capture different kinds of updates and the reasoning styles that they induce, using different formalisms and under various assumptions: dynamic epistemic logic (van Benthem, 2011; van Ditmarsch et al., 2007), default and non-monotonic logics (Makinson, 2005; Gabbay et al., 1998), belief revision theory (Gärdenfors, 1988), conditional logic (Nute and Cross, 2001), *etc.*

These lecture notes are an introduction to logic and commonsense reasoning. The notes are divided into three parts: *A Logic Compendium* (Part I), *Reasoning about Uncertainty* (Part II) and *Commonsense Reasoning* (Part III). We outline below the content and objectives of each part.

Part I: Logic. This part will introduce the basic concepts and methods of logic. The objective is to provide the logical background that will be necessary to deal with the rest of the content: most of the formalisms introduced will indeed be *logical* or *logic-based*. This means, in particular, that they will all have a similar structure based on a syntax and a semantics and that the problems that we will address and formalize will often be expressed as standard decision problems in these logics.

This part contains two chapters: Chapter 1, titled "*Logic: Basic Concepts*" and Chapter 2, titled "*Decidability, Complexity and Expressiveness*".

Part II: Representing and reasoning about uncertainty. This part will present the main logical formalisms that have been developed for specifying and reasoning about MAS and for representing and reasoning about uncertainty. Reasoning about uncertainty can sometimes be subtle and it requires a careful and rigorous analysis, especially if this reasoning is used to take decisions. Reasoning about uncertainty is complex in a multi-agent setting, since we have to take into account not only

the uncertainty and beliefs that the agents have about the surrounding world, but also their uncertainty and beliefs about the other agents' uncertainty and beliefs. This is even more complex when we introduce events and communication between agents, since we have to deal with the way the agents update and sometimes revise their beliefs when they get new pieces of information, which possibly contradicts their previous beliefs. Thus, this part will propose formal accounts and logic-based formalizations of communication, belief revision and update, incomplete information, information dynamics, *etc.*

This part contains two chapters: Chapter 3, titled *Reasoning Alone about Uncertainty* and Chapter 4, titled *Reasoning with Others about Uncertainty.*

Part III: Commonsense reasoning. This part will present some of the most familiar conditional logics for conditionals and counterfactuals introduced by the philosophers Stalnaker and Lewis, as well as a generic logical framework based on plausibility measures for dealing with non-monotonic and default reasoning introduced by the computer scientists Friedman and Halpern. Conditionals and non-monotonic reasoning are tricky topics and have been the subject of numerous debates in the history of logic. Even if these two topics cannot be taken in isolation, conditionals have traditionally been studied by philosophers whereas non-monotonic and default reasoning is rather a topic on the research agenda of computer scientists and more specifically researchers in artificial intelligence. Finally, we will present the basics of belief revision theory and its connection with non-monotonic reasoning, formalized via the Ramsey test.

This part contains three chapters: Chapter 5, titled *Conditionals*, Chapter 6, titled *Default Reasoning* and Chapter 7, titled *Belief Revision.*

The material of these lecture notes sometimes stems from neighboring fields, in particular computer science and artificial intelligence. This can be explained by the fact that commonsense reasoning and the representation of uncertainty are also the subject of investigations in these other fields, even if the overall approach and the objectives are sometimes different. For computer science and artificial intelligence, the rationale for studying these issues is that it can lead to the development of rational or software agents that can act autonomously in an uncertain and changing world like internet or even the real world. In that respect, computer scientists are more concerned with computational and decidability (implementability) issues than philosophers. This said, the boundary between philosophy and artificial intelligence is sometimes very fuzzy.

The material of these lecture notes is quite introductory and we will often only scratch the surface of a field which is in fact much more advanced and developed than what we present. For this reason, a number of pointers for further reading is provided at the end of each chapter. The main references for the lectures notes are the books of Halpern (2003), Goble (2001), Priest (2011) and van Benthem (2010).

Notes:

- These lecture notes are self-contained, no other material or book is needed to understand and study them. In particular, the usual notions and notations of set theory are recalled in the Appendix (Chapter A).
- Some parts of these lectures notes are largely based on or copied verbatim from publications of other authors. When this is the case, these parts are mentioned at the end of each chapter in the section "*Further reading*". For this reason, please do not circulate these lecture notes. Some parts of the begining of this general introduction stem from (Edgington, 2014).
- These lecture notes are supported by a website where all the homework (DM), exercise labs (TD) and the slides of presentations of articles made in class can be found. This website is accessible via moodle (`https://foad.univ-rennes1.fr/login/index.php`).

Part I

A Logic Compendium

Introduction to Part I

"Tous les chats sont mortels, Socrates est mortel, donc Socrates est un chat."
– Eugène Ionesco, *Rhinocéros*, 1959

The Sophists were the first lawyers in the world and they sought to devise an objective system of inference rules that could be applied in a dispute in order to confound their adversary. In reaction to the development of fallacies by the Sophists, Aristotle developed the syllogisms (Aubenque, 2012). From that moment on and for a very long time, logic dealt with the issue of determining the valid *forms* of reasoning, that is, determining whether a reasoning is 'valid' or 'good' independently of the specific content of this reasoning. For example, the reasoning "all X are Y, all Y are Z, therefore all X are Z" is a valid reasoning, independently of what X, Y and Z stand for. This means that if we assume that the *premises* "all X are Y" and "all Y are Z" are true, then we must necessarily infer that the *conclusion* "all X are Z" is also true. The symbols X, Y and Z are abstract symbols which stand for any kind of objects. In particular, this reasoning holds if we replace X with "women", Y with "humans" and Z with "mortal". Generally speaking, a reasoning is represented by a set of *rules of inference*, like the following one:

Premise:	All X are Y
Premise:	All Y are Z
Conclusion:	All X are Z

A rule of inference represents a 'pattern' of reasoning and a specific set of rules of inference will define a specific kind of reasoning.

For a very long time, logic was considered as a *normative* discipline whose goal was to set the standards of 'correct' reasoning and to identify the valid rules of inference: "Cette science des lois nécessaires de l'entendement et de la raison en général ou, ce qui est la même chose, de la simple forme de la pensée en général, nous la nommons: *Logique*" (Kant, 1800, p. 12–13). Nowadays, no logician would agree with this definition of logic. Hintikka and Sandu tell us that "[i]t is far from clear what is meant by logic or what should be meant by it. It is nevertheless reasonable to identify logic as the study of inferences and inferential relations" (Hintikka and Sandu, 2007, p. 13). In fact, the object of study of logic remains the reasoning, but the normative definition of logic is replaced by a *descriptive* one: the objective is now to characterize different kinds of reasoning occurring in different contexts and situations. Numerous logics have been developed

over the years, each of them modeling a specific kind of reasoning: the reasoning about time, about knowledge, about programs, *etc.* (see the list of logics in Section 1.5). Nevertheless, the concepts of *rule of inference*, *truth*, *validity*, *logical consequence*, *etc.* remain common to all the logics, because these concepts are characteristic of any form of reasoning. Logic is a unified and unifying discipline. On the one hand, the unity of logic is guaranteed by its common interest in reasoning. On the other hand, its unifying power is based on the fact that reasoning permeates a wide range of activities.

In computer science, the formal descriptive language of modern logic serves as a working tool. Logic is used as a means to represent and reason about problems and specific applications. For a specific application domain, a logic can be devised to address the problems that must be solved (see again Section 1.5). This is sometimes called "logic engineering". Once the logic is defined, the problems at stake can be formulated in logical terms and this reformulation often permits to reuse results or algorithms that have been obtained for other logics. This transfer of results is facilitated by the unified aspect of logic and its common methodology: as we said, all logics have the same format and deal with the same notions of Truth, validity, logical consequence, *etc.*

Once a logic is defined, we can study its properties. In particular, we can study whether it is more expressive than another logic, that is, whether it can express more things about a given model than another logic (this is called *expressiveness*). We can also study whether it is possible to find an algorithm that will solve a specific problem formulated in this logic (this is called *decidability*) and how complex and hard it will be to solve this problem with such an algorithm (this is called *computational complexity*).

The elementary concepts of logic will be dealt with in Chapter 1 and the notions of decidability, expressiveness and computational complexity that are used to study the properties of a given logic will be addressed in Chapter 2.

Chapter 1

Logic: Basic Concepts

"'Contrariwise,' continued Tweedledee, 'if it was so, it might be; and if it were so, it would be; but as it isn't, it ain't. That's logic.' "

– Lewis Caroll, *Through the Looking-Glass*, 1871

1.1 Introduction

There are three possible equivalent approaches for formally introducing and defining a logic equipped with a proof system. We present them below. The different concepts that are highlighted will be given a precise and rigorous meaning in the rest of the chapter. In the first two approaches, we start by defining a *logical language* which consists of a set of well-formed formulas (often defined by a grammar). Then, there are two different alternatives: a *semantically-driven* alternative and a *syntactically-driven* alternative.

1. In the *semantically-driven* alternative, we start by providing a semantics to the well-formed formulas by means of a class of models and a *satisfaction relation*. This semantics gives meaning to well-formed formulas and defines at the same time a set of *validities*: the well-formed formulas which are satisfied in every model. To capture the set of validities, we define a *proof system*, which is a (finite) set of *axiom schemata* and *inference rules* from which we can derive specific well-formed formulas called *theorems*.

2. In the *syntactically-driven* alternative, we start by providing a proof system that defines a set of theorems. This set of theorems is another means to characterize and define the logic. Then, we define a semantics for this logic which defines in turn a set of validities.

The coincidence between the set of validities and the set of theorems is captured by the notions of *soundness* and *completeness* (one notion for each inclusion). In that case, we say that the proof system *axiomatizes* the logical language for the semantics that we have defined.

3. In the third approach, we proceed the other way around. We first define a semantics consisting of a specific class of models and then a logical language with a satisfaction relation to 'talk about' these models. This defines in turn a set of validities that we can axiomatize with a proof system, as in the first approach.

In all cases, the three approaches lead to the same outcome: a logical language equipped with a syntax and semantics, together with a proof system axiomatizing the set of validities of the logic (*i.e.*, a proof system such that its set of theorems is the set of validities of the logic).

The chapter is organized as follows. Our presentation will follow a semantically-driven approach. After a short history of logic (Section 1.2), we introduce the syntax and semantics of propositional logic (PL), modal logic (ML) and first-order logic (FO) (Section 1.3). Then, we recall some basic notions of proof theory by introducing the key concepts of axiom, inference rule, proof, soundness and completeness (Section 1.4). We end the chapter with a panorama of well known logics which have been introduced in the literature (Section 1.5). They can all be seen as variants or combinations of propositional, modal or first-order logic. Finally, we give pointers for further readings (Section 1.6).

1.2 A Short History of Logic

We split up the history of logic into three distinct ages as follows. Section 1.2.1: the origins of logic (antiquity); Section 1.2.2: mathematical logic (late 19^{th} to 20^{th} century); and Section 1.2.3: logic in computer science (mid 20^{th} century to now).

1.2.1 The Origins of Logic

The study of logic was begun by the ancient Greeks whose educational system stressed competence in reasoning and in the use of language. Along with rhetoric and grammar, logic formed part of the *trivium*, the first subjects taught to young people. Logic was inaugurated by Aristotle with his *Organon* where he introduced the so-called syllogisms. A syllogism is a kind of argument in which one statement (the conclusion) is inferred from two or more others (the premises) of a specific form. Here is an example of one form of syllogism:

Major Premise:	All men are mortal.
Minor Premise:	Socrates is a man.
Conclusion:	Socrates is mortal.

Originally, syllogisms were developed by the Sophists for practical reasons. The Sophists were the first lawyers in the world and they sought to devise an objective system of inference rules (the syllogisms) that could be applied in a dispute in order to confound their adversary (Aubenque, 2012). Logic was devised for the purpose to determine beyond any doubt who had won an argument.

For a long time, that is from around 300 BC right until the end of 19^{th} century, the systems of reasonings formulated by Aristotle and the Stoic philosophers were the only forms of reasoning studied. They were expressed in natural language. Independent traditions arose around 300 BC also in China and India, which produced famous figures like the Buddhist logician Dignaga, or Gangesa, and this long tradition lives on in some philosophical schools today. Through translations of Aristotle, logic also reached the Islamic world. The work of the Persian logician Avicenna around 1000 AD was still taught in madrassa's by 1900.

1.2.2 Second Age: Mathematical Logic (late 19^{th} to mid 20^{th} Century)

The formalization of logic began in the nineteenth century as mathematicians attempted to clarify the foundations of mathematics. Around that time, the formal languages of propositional logic and first-order logic were developed by Boole and then Frege. One trigger was the discovery of non-Euclidean geometries: replacing Euclid's parallel axiom with another axiom resulted in a different theory of geometry that was just as consistent as that of Euclid. Proof systems – axioms and rules of inference – were developed with the understanding that different sets of axioms would lead to different theorems. The questions investigated included:

- *Consistency*: a proof system is consistent if it is impossible to prove both a formula and its negation.
- *Independence*: The axioms of a proof system are independent if no axiom can be proved from the others.
- *Soundness*: All theorems that can be proved in the proof system are true.
- *Completeness*: All true statements can be proved in the proof system.

Clearly, these questions will only make sense once we have formally defined the central concepts of *truth* and *proof*. During the first half of the twentieth century, logic became a full-fledge topic of modern mathematics. The framework for research into the foundations of mathematics was called *Hilbert's program* (named after the great mathematician David Hilbert). His central goal was to prove that mathematics, starting with arithmetic, could be axiomatized in a system that was both consistent and complete. This program was shattered by a number of significative results:

- Gödel's First Incompleteness Theorem showed in 1931 that this goal cannot be achieved: any consistent axiomatic system for arithmetic is incomplete since it contains true statements that cannot be proved within the system.
- Gödel's Second Incompleteness Theorem proved that a proof system powerful enough to form statements of about arithmetics cannot prove its own consistency.

- Church and Turing showed that there are some problems that no algorithm could ever solve. If such problems exist, then there could be no hope of finding a single algorithm to produce all mathematical truths.

In parallel, proof systems were also developed, notably by Gentzen, in order to capture the actual mathematical reasoning performed by mathematicians: the so-called sequent calculi for natural deduction (Gentzen, 1935). However, in everyday life, the way we update and revise information is quite different from the actual reasoning of mathematicians. This has lead researchers in artificial intelligence and computer science from the 1980s on to develop logical theories that study and formalize belief change and the so-called "common sense reasoning". The rationale underlying the development of such theories was that it would ultimately help us understand our everyday life reasoning and the way we update our beliefs, and that the resulting work could subsequently lead to the development of tools that could be used for example by artificial agents in order to act autonomously in an uncertain and changing world. A number of theories have been proposed to capture different kinds of updates and the reasoning styles that they induce, using different formalisms and under various assumptions: dynamic epistemic logic (van Benthem, 2011; van Ditmarsch et al., 2007), default and non-monotonic logics (Makinson, 2005; Gabbay et al., 1998), belief revision theory (Gärdenfors, 1988), conditional logic (Nute and Cross, 2001), *etc.*

1.2.3 Third Age: Logic in Computer Science (mid 20^{th} Century to Now)

Modal notions of necessity, possibility, and contingency that were standard fare in traditional logic up to the 19^{th} century went out the door in the work of the founding fathers of modern logic, like Boole and Frege. They reappeared on the logical agenda of philosophical logicians like Prior, Kripke, Hintikka, Lewis or Stalnaker in the 1950s. This is the period where labels like "modal logic", "epistemic logic", "deontic logic", "temporal logic",... were coined. In the 1970s, this philosophical phase was consolidated into a beautiful mathematical theory by authors like Blok, Fine, Gabbay, Segerberg and Thomason. But simultaneously, modal logic crossed over to linguistics, when "Montague semantics" gave the study of intensional expressions in natural language pride of place, using mixes of modal logic with type theory and other tools from mathematical logic (ter Meulen and van Benthem, 2010). In the same decade, and especially through the 1980s, modal notions found their way into computer science in the study of programs (Pratt, 1976), and into economics in the study of knowledge of players in games (Aumann, 1976). The formal descriptive language of modern logic served as a working tool for computer science.

In fact, many computer scientists claim that logic is "the calculus of computer science" (Manna and Waldinger, 1985; Halpern et al., 2001). The significance and importance of logic for computer science, the science of computing, is indeed overwhelming (Abramsky et al., 1992; Gabbay and Robinson, 1998; Ben-Ari, 2012). As a matter of fact, in the 1980's, as recalled by Dov Gabbay, "[computer science and artificial intelligence]

Aristotle

Evert Willem Beth

George Boole

Lewis Carroll

Alonzo Church

Gottlob Frege

Gerhard Gentzen

Kurt Gödel

Jacques Herbrand

David Hilbert

Jaakko Hintikka

Saul Kripke

Joachim Lambek

Gottfried Leibniz

David Lewis

Frank Ramsey

Bertrand Russell

Alfred Tarski

Alan Turing

Ludwig Wittgenstein

were under increasing commercial pressure to provide devices which help and/or replace the human in his daily activity. This pressure required the use of logic in the modeling of human activity and organization on the one hand and to provide the theoretical basis for the computer program constructs on the other" (Gabbay and Guenthner, 2001, p. vii). The result was that the research in (philosophical) logic has been applied to the needs of these active communities (and has been at the same time pushed forward).We illustrate it below with some of its contributions (see also (Gabbay and Guenthner, 2001, p. x-xiii),(Halpern et al., 2001; Vardi, 2009)). The three first ones are rather theoretically-driven contributions, while the remaining three ones are application-driven contributions.

Program verification and semantics: In 1969, the use of logic was propounded by Hoare as a means to prove the correctness of computer programs (Hoare, 1969). The use of logic transformed programming from arts and crafts to a science and helped to develop a theory of programming. Nowadays, temporal and dynamic logics are widely used in the industry in combination with *model-checking* techniques in order to verify sequential and concurrent programs (Baier and Katoen, 2008).

Complexity theory: Logic triggered the invention of the Turing machine. The Turing machine is now the standard model of algorithms and computation. This led in turn to the development of *computational complexity theory* which aims at determining how hard it is to solve a problem in terms of computations (Papadimitriou, 2003). As a twist of history, logic is now coming back to this theory with a research field called *descriptive complexity theory* (Immerman, 1999), where the different classes of complexity and their relationships are expressed in logical terms.

Type systems for programming language: Although type theory and Gentzen's logical calculus of natural deduction were for a long time considered as highly theoretical fields of research in philosophical and mathematical logic, they were used in the 1980s as a unifying conceptual framework for the design, analysis, and implementation of *programming languages*. This convergence was made possible by the Curry-Howard isomorphisms (Leeuwen, 1990, Chap. 8).

Databases: First-order logic was used for *relational databases* and it improved tremendously the efficiency of querying data from a database (Abiteboul et al., 1995). Nowadays, first-order logic lies at the core of modern database systems, and the standard query languages such as Structured Query Language (SQL) and Query-By-Example (QBE) are syntactic variants of first-order logic.

Logic programming: Logic also gave rise to its own programming language, PROLOG. *Logic programming* stems from an adaptation of a logical calculus for first-order logic called SDL-resolution (Gabbay and Robinson, 1998, Vol. 5). While initially aimed at natural language processing, the language has since then stretched far into other areas like theorem proving, expert systems, games, automated answering systems, ontologies.

Computer circuits: Propositional (Boolean) logic is used to design computer circuits. A logic gate is an idealized or physical device implementing a Boolean function. Logic gates are primarily implemented using diodes or transistors. They can be cascaded in the same way that Boolean functions can be composed, allowing the construction of a physical model of the algorithms that can be described with Boolean logic.

1.3 Propositional, Modal and First-Order Logic

A logic can be represented as a triple (language, class of models, satisfaction relation). The three logics will be presented by following this tri-partite representation: (1) language, (2) class of models, (3) satisfaction relation. The *semantics* of a logic is usually defined by the combination of (2) and (3) and its *syntax* by (1). The syntax defines the *logical language* which is a set of *well-formed formulas* of the logic.

1.3.1 Propositional Logic (PL)

In the sequel, $PROP$ is a countable set of *atoms* (propositional letters) denoted $p, q, r, \ldots$ and T and F are two symbols called *truth values* standing for True and False.

Definition 1.3.1 (Propositional language $\mathcal{L}_{\mathsf{PL}}$). The language $\mathcal{L}_{\mathsf{PL}}$ is the smallest set that contains $PROP$ and that is closed under negation and conjunction. That is,

- if $\varphi \in \mathcal{L}_{\mathsf{PL}}$, then $\neg\varphi \in \mathcal{L}_{\mathsf{PL}}$;
- if $\varphi, \psi \in \mathcal{L}_{\mathsf{PL}}$, then $(\varphi \wedge \psi) \in \mathcal{L}_{\mathsf{PL}}$.

In other words, the language $\mathcal{L}_{\mathsf{PL}}$ is defined by the following grammar in Backus-Naur Form (BNF):

$$\mathcal{L}_{\mathsf{PL}} : \quad \varphi \quad ::= \quad p \mid \neg\varphi \mid (\varphi \wedge \varphi)$$

where $p \in PROP$. We introduce the following abbreviations: $\bot := p \wedge \neg p$ for a chosen $p \in PROP$, $\top := \neg\bot$, $\varphi \vee \psi := \neg(\neg\varphi \wedge \neg\psi), \varphi \to \psi := \neg\varphi \vee \psi, \varphi \leftrightarrow \psi := (\varphi \to \psi) \wedge (\psi \to \varphi)$. To save parenthesis, we use the following ranking of binding strength: $\neg, \wedge, \vee, \to, \leftrightarrow$ (*i.e.*, $\neg$ binds stronger than $\wedge$, *etc.*). For example, $\neg p \wedge q \to r \vee s$ means $((\neg p) \wedge q) \to (r \vee s)$. □

If we wanted to consider *non-classical* logics, such as intuitionistic logic, relevant logic, many-valued logics, conditional logics,... then the connectives introduced as abbreviations should be introduced as primitives in the language.

Definition 1.3.2 (Interpretation). An *interpretation* is a total function $I : PROP \mapsto \{T, F\}$ that assigns one of the *truth values* T or F to *every* atom in $PROP$. The set of interpretations is denoted $\mathcal{C}_{\mathsf{PL}}$. □

Definition 1.3.3 (Satisfaction relation $\models_{\mathsf{PL}}$). The *satisfaction relation* $\models_{\mathsf{PL}} \subseteq \mathcal{C}_{\mathsf{PL}} \times \mathcal{L}_{\mathsf{PL}}$ is defined inductively as follows (we omit the subscript PL subsequently). Let $I \in \mathcal{C}_{\mathsf{PL}}$ and $\varphi, \psi \in \mathcal{L}_{\mathsf{PL}}$.

$$\begin{array}{lll} I \models p & \text{iff} & I(p) = T \\ I \models \neg\varphi & \text{iff} & \text{it is not the case that } I \models \varphi \\ I \models \varphi \wedge \psi & \text{iff} & I \models \varphi \text{ and } I \models \psi \end{array}$$ □

The inductive clauses defining a satisfaction relation are often called the *truth conditions*.

Example 1.3.1. Let $PROP := \{p, q, r\}$. We define the total function $I : PROP \to \{T, F\}$ as follows: $I(p) = I(q) := T$ and $I(r) := F$. Then, we have for example that $I \models p \wedge q \wedge \neg r$, $I \models \neg p \to r$, $I \models p \vee r$, $I \models p \vee q$ and $I \models r \to (q \to r)$ hold. □

1.3.2 Modal Logic (ML)

In the sequel, $AGTS := \{1, \ldots, n\}$ is a set of indices.

Definition 1.3.4 (Modal language $\mathcal{L}_{\mathsf{ML}}$). The multi-modal language $\mathcal{L}_{\mathsf{ML}}$ is defined inductively by the following grammar in BNF:

$$\mathcal{L}_{\mathsf{ML}} : \quad \varphi \quad ::= \quad p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_i\varphi$$

where $p \in PROP$ and $i \in AGTS$. The formula $\Diamond_i\varphi$ is an abbreviation for $\neg\Box_i\neg\varphi$. We use the same abbreviations as in Definition 1.3.1. The formulas that go in the making of φ are called *subformulas of* φ and its set is denoted $\mathsf{Sub}(\varphi)$. □

Example 1.3.2. Here are some examples of modal formulas:

- $\Box_i\bot$, $\Diamond_i\top$, $\Box_i\Diamond_i\top$.
- $p \wedge \Diamond_i\neg p$, $\Box_i p \vee \Box_i\neg p$, $p \to \Box_i p$.

Then, $\mathsf{Sub}(p \wedge \Diamond_i\neg p) = \{p, \Diamond_i\neg p, \neg p\}$ and $\mathsf{Sub}(\Box_i p \vee \Box_i\neg p) = \{\Box_i p, \Box_i\neg p, p, \neg p\}$. □

Now, we present the so-called *possible world semantics*.

Definition 1.3.5 (Kripke model and frame). A *Kripke model* $\mathcal{M}$ is a tuple $\mathcal{M} := (W, R_1, \ldots, R_n, V)$ where

- W is a non-empty set whose elements are called *possible worlds*;
- $R_1, \ldots, R_n$ are binary relations over W called *accessibility relations*;
- $V : PROP \times W \to \{T, F\}$ is a function called the *valuation function*.

If $w \in W$ and $j \in AGTS$, we write wR_jv or R_jwv for $(w, v) \in R_j$, and $R_j(w)$ denotes $\{v \in W : wR_jv\}$. We abusively write $w \in \mathcal{M}$ for $w \in W$. The pair $(\mathcal{M}, w)$ is called a *pointed Kripke model*. A *Kripke frame*, generally denoted F, is a Kripke model without valuation function. The class of all pointed Kripke models is denoted $\mathcal{C}_{\mathsf{ML}}$. □

Definition 1.3.6 (Satisfaction relation $\models_{\mathsf{ML}}$). We define the *satisfaction relation* $\models_{\mathsf{ML}} \subseteq \mathcal{C}_{\mathsf{ML}} \times \mathcal{L}_{\mathsf{ML}}$ inductively as follows (we omit the subscript ML subsequently). Let $(\mathcal{M}, w) \in \mathcal{C}_{\mathsf{ML}}$ and let $\varphi \in \mathcal{L}_{\mathsf{ML}}$.

$$\begin{array}{lll} \mathcal{M}, w \models p & \text{iff} & V(p, w) = T \\ \mathcal{M}, w \models \neg\varphi & \text{iff} & \text{it is not the case that } \mathcal{M}, w \models \varphi \\ \mathcal{M}, w \models \varphi \wedge \psi & \text{iff} & \mathcal{M}, w \models \varphi \text{ and } \mathcal{M}, w \models \psi \\ \mathcal{M}, w \models \Box_i\varphi & \text{iff} & \text{for all } v \in W \text{ such that } R_iwv, \mathcal{M}, v \models \varphi \end{array}$$

We write $F \models \varphi$ when $(F, V), w \models \varphi$ for all valuations V and all worlds $w \in F$. □

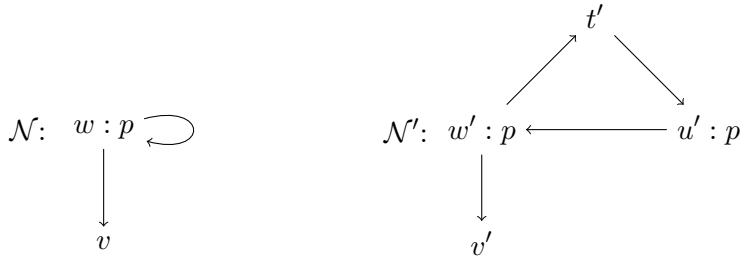
From these definitions, we can derive the following truth condition for the possibility modality $\Diamond_i$:

$$\mathcal{M}, w \models \Diamond_i\varphi \quad \text{iff} \quad \text{there is } v \in W \text{ such that } R_iwv \text{ and } \mathcal{M}, v \models \varphi$$

Example 1.3.3. Let us consider the Kripke models $\mathcal{N} := (W, R, V)$ and $\mathcal{N}' := (W', R', V')$ defined as follows (here, $PROP := \{p\}$):

$$\begin{array}{ll} W := \{w, v\} & W' := \{w', v', u', t'\} \\ R := \{(w, w), (w, v)\} & R' := \{(w', t'), (t', u'), (u', w'), (w', v')\} \\ V(p, w) := T & V'(p, w') = V'(p, u') := T \\ V(p, v) := F & V'(p, v') = V'(p, t') := F \end{array}$$

Here, we have a single accessibility relation R (unlike in the general Definition 1.3.5 where we have n accessibility relations). We represent graphically these Kripke models $\mathcal{N}$ and $\mathcal{N}'$ by the following figures:



Possible worlds are represented by Latin letters, the accessibility relation R is represented by arrows between pairs of possible worlds and the propositional letter p holds in a possible world when this possible world is labelled with p, and does not hold otherwise. Then, we have for example that the following hold:

- $\mathcal{N}, w \models \Diamond p \wedge \Diamond\neg p$:

  Indeed, $\mathcal{N}, w \models \Diamond p \wedge \Diamond\neg p$
  because $\mathcal{N}, w \models \Diamond p$ and $\mathcal{N}, w \models \Diamond\neg p$
  because Rww and $\mathcal{N}, w \models p$, and Rwv and $\mathcal{N}, v \models \neg p$
  because $V(p, w) = T$ and $V(p, v) = F$.

- $\mathcal{N}, v \models \Box\bot$ because there is no $u \in W$ such that Rvu (in other words, $R(w) = \{u : Rwu\} = \emptyset$): we recall that a universal quantification on an empty set is always true. □

1.3.2.1 'Game' Semantics

For the game semantics, we consider two players: **V** and **F**. The goal of player **V** is to show that the formula is true in the model and the goal of player **F** is to show that the formula is false in the model.

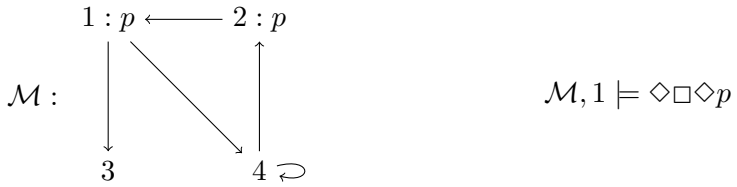
Definition 1.3.7 (Evaluation game). Let $\mathcal{M}$ be a Kripke model, let $w \in W^{\mathcal{M}}$ and φ a ML formula. The *evaluation game* denoted game$(\mathcal{M}, w, \varphi)$ starts at the world w. Each move is determined by the main operator of φ and we move to its subformulas:

atom p	:	*test* p at w: if true, then **V** wins, if false, then **F** wins,
$\varphi \vee \psi$	:	**V** *chooses* which disjunct to play
$\varphi \wedge \psi$	:	**F** *chooses* which conjunct to play
$\neg\varphi$	:	*role switch* between the two players, play continues w.r.t. φ
$\Diamond_i\varphi$	:	**V** picks an R_i–successor v of the current world, play continues w.r.t. φ at v
$\Box_i\varphi$	:	**F** picks an R–successor v of the current world, play continues w.r.t. φ at v.

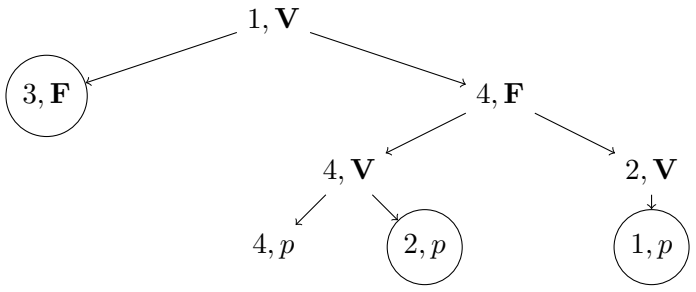
A player also loses when (s)he must pick a successor, but cannot do. □

Theorem 1.3.1. *For all Kripke model $\mathcal{M}$, all $w \in W^{\mathcal{M}}$, $\mathcal{M}, w \models \varphi$ if, and only if,* **V** *has a winning strategy for* game$(\mathcal{M}, w, \varphi)$.

Example 1.3.4.



So, **V** has a winning strategy for *game*$(\mathcal{M}, 1, \Diamond\Box\Diamond p)$. The circle nodes indicate the winning positions for *Verifier*:



□

There exist other alternative semantics for ML than the Kripke semantics: the algebraic semantics, the neighborhood semantics and the topological semantics (van Benthem and Blackburn, 2007).

Applications. The modality $\Box_i$ can have different intuitive interpretations depending on what we want to represent and reason about. For instance, in epistemic logic, $\Box_i\varphi$ reads as "agent i *knows* that φ holds". In dynamic logic, $\Box_i\varphi$ reads as "after every successful completion of action i, φ holds". Likewise for the accessibility relation.

1.3.3 First-order Logic (FO)

In the sequel, VAR is a set of *variables*, $CONS := \{c_1, \ldots, c_m\}$ is a finite set of *constants* and $PRED := \{R_1, \ldots, R_n\}$ is a set of *predicate symbols* of arity $k_1, \ldots, k_n$ respectively, whose one of them is the *identity predicate* $=$ of arity 2. We do not consider function symbols.

Definition 1.3.8 (First-order language $\mathcal{L}_{\mathsf{FO}}$). The language $\mathcal{L}_{\mathsf{FO}}$ of First-Order Logic is defined inductively by the following grammars in BNF.

$$\begin{array}{ll} \mathcal{T}: & t ::= c \mid x \\ \mathcal{L}_{\mathsf{FO}}: & \varphi ::= R(t_1, \ldots, t_k) \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \forall x\varphi \end{array}$$

where $c \in CONS$, $x \in VAR$, $R \in PRED$ and $t, t', t_1, \ldots, t_k \in \mathcal{T}$. Elements of $\mathcal{L}_{\mathsf{FO}}$ are called *formulas* and elements of $\mathcal{T}$ are called *terms*. Formulas of the form $R(t_1, \ldots, t_k)$ are called *atomic formulas*. The formula $t = t'$ is an abbreviation for $=(t, t')$ and $\exists\varphi$ is an abbreviation for $\neg\forall\neg\varphi$. Let $\varphi \in \mathcal{L}_{\mathsf{FO}}$. An occurrence of a variable x in φ is a *free variable* of φ if, and only if, x is not within the scope of a quantified variable x. A variable which is not free is *bound*. We say that a formula of $\mathcal{L}_{\mathsf{FO}}$ is a *sentence*, or is *closed*, when it contains no free variable. □

Example 1.3.5. No variable of φ below is *free*, they are all *bound*:

$$\varphi := \forall x \forall y (x < y \to \exists z (x < z \wedge z < y))$$

However, in formula ψ, y is a *free* variable, but x is *bound*:

$$\psi := \forall x (\neg(x = 0) \to x > y)$$ □

We present two kinds of semantics: the classical 'Tarskian' one and the game-theoretical one proposed first by Hintikka.

1.3.3.1 Tarskian Semantics

Definition 1.3.9 (Structure and assignment). A *structure* is a tuple $\mathcal{M} := (W^{\mathcal{M}}, R_1^{\mathcal{M}}, \ldots, R_n^{\mathcal{M}}, c_1^{\mathcal{M}}, \ldots, c_m^{\mathcal{M}})$ where:

- $W^{\mathcal{M}}$ is a non-empty set called the *domain*;
- $R_1^{\mathcal{M}}, \ldots, R_n^{\mathcal{M}}$ are relations over $W^{\mathcal{M}}$ with the same arity as $R_1, \ldots, R_n$ respectively;
- $c_1^{\mathcal{M}}, \ldots, c_m^{\mathcal{M}} \in W^{\mathcal{M}}$ are distinguished elements.

An *assignment* is a function $\sigma : VAR \to W^{\mathcal{M}}$. The function $\sigma[x := w]$ is the same assignment as σ except that x is mapped to w. The set of pairs of structures and assignments is denoted $\mathcal{C}_{\mathsf{FO}}$. □

Example 1.3.6. $(\mathbb{Q}, <^{\mathbb{Q}}, 0)$, $(\mathbb{R}, <^{\mathbb{R}}, 1)$, (labeled) graphs,... are structures. Note that a Kripke frame can also be seen as a structure. □

Definition 1.3.10 (Satisfaction relation $\models_{\mathsf{FO}}$). The *satisfaction relation* $\models_{\mathsf{FO}} \subseteq \mathcal{C}_{\mathsf{FO}} \times \mathcal{L}_{\mathsf{FO}}$ is defined inductively as follows (we omit the subscript FO subsequently). Let $\varphi \in \mathcal{L}_{\mathsf{FO}}$ and $(\mathcal{M}, \sigma) \in \mathcal{C}_{\mathsf{FO}}$.

$$\mathcal{M}, \sigma \models R_i(x_1, \ldots, x_{n_i}) \quad \text{iff} \quad (w_1, \ldots, w_{n_i}) \in R_i^{\mathcal{M}}$$

where for all k, $w_k := \begin{cases} c_l^{\mathcal{M}} & \text{if for some } l,\ x_k = c_l; \\ \sigma(x_k) & \text{otherwise.} \end{cases}$

$$\begin{array}{lll} \mathcal{M}, \sigma \models \neg\varphi & \text{iff} & \text{it is not the case that } \mathcal{M}, \sigma \models \varphi \\ \mathcal{M}, \sigma \models \varphi \wedge \psi & \text{iff} & \mathcal{M}, \sigma \models \varphi \text{ and } \mathcal{M}, \sigma \models \psi \\ \mathcal{M}, \sigma \models \forall x\varphi & \text{iff} & \mathcal{M}, \sigma[x := w] \models \varphi \text{ for all } w \in W^{\mathcal{M}} \end{array}$$

We say that the formula φ is *true in* $\mathcal{M}$ *under* σ or that $\mathcal{M}$ is *a model of* φ *under* σ when $\mathcal{M}, \sigma \models \varphi$. If $\varphi \in \mathcal{L}_{\mathsf{FO}}$ and x is free in φ, then $\mathcal{M} \models \varphi[x/w]$ means that we are evaluating w.r.t. an assignment that assigns w to x. □

From this definition, we derive that $\mathcal{M}, \sigma \models \exists x\varphi$ if, and only if, $\mathcal{M}, \sigma[x := w] \models \varphi$ for some $w \in W^{\mathcal{M}}$.

1.3.3.2 'Game' Semantics

Suppose two parties disagree about a sentence φ in some situation $\mathcal{M}$ under discussion: *Verifier* **V** claims that φ is true in $\mathcal{M}$, and *Falsifier* **F** claims that it is false.

Definition 1.3.11 (Evaluation game). Let $\mathcal{M}$ be a structure and let $\varphi \in \mathcal{L}_{\mathsf{FO}}$. The *evaluation game* denoted game$(\mathcal{M}, \varphi)$ is defined by induction on the structure of the formula φ by means of moves of defense and attack – with the following schedule depending on the formula under consideration:

$$\begin{array}{rcl} R(t_1, \ldots, t_k) & : & \textit{test} \text{ to determine who wins} \\ \neg\varphi & : & \textit{role switch} \text{ between the two players, play continues w.r.t. } \varphi \\ \varphi \wedge \psi & : & \mathbf{F} \textit{ chooses} \text{ which conjunct to play} \\ \forall x\varphi(x) & : & \mathbf{F} \text{ picks an element } w, \text{ play continues w.r.t. } \varphi(w) \end{array}$$ □

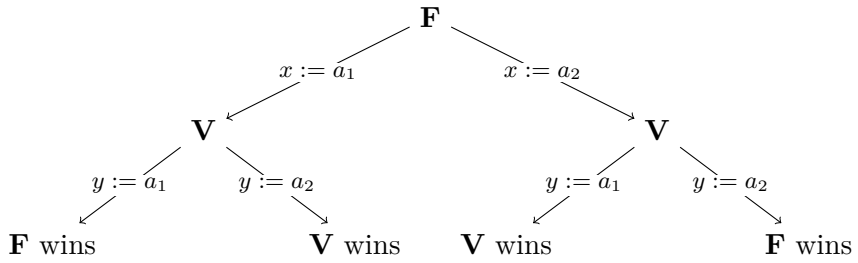
The problem of the role switch for negation can be circumvented by pushing negations inside to the atoms.

Example 1.3.7.

1. Consider the following structure $\mathcal{M}$ and sentence φ:

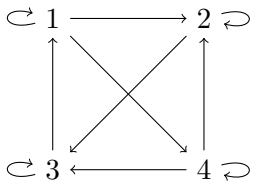
$$\mathcal{M}: \quad a_1 \qquad a_2 \qquad \varphi := \forall x \exists y \neg (x = y)$$

The evaluation game in extensive form on the structure $\mathcal{M}$ for the formula φ:



Falsifier starts, Verifier must respond. There are 4 possible plays, with 2 wins for each player. Verifier is the only one who has a *winning strategy.*

2. Communication network:



$$\forall x \forall y (Rxy \vee \exists z (Rxz \wedge Rzy))$$

Here is a possible run of the corresponding evaluation game:

Player	*move*	*next formula*
F	picks 2	$\forall y (R2y \vee \exists z (R2z \wedge Rzy))$
F	picks 4	$R24 \vee \exists z (R2z \wedge Rz4)$
V	chooses	$\exists z (R2z \wedge Rz4)$
V	picks 1	$R21 \wedge R14$
F	chooses	$R14$
test	**V** wins	

Question: does **V** have a winning strategy? □

Proposition 1.3.1. *For all structures $\mathcal{M}$, all sentence φ, $\mathcal{M} \models \varphi$ iff verifier **V** has a* winning strategy *for the evaluation game for φ played in $\mathcal{M}$.*

Proof. The proof is by induction on formulas. One shows simultaneously that if a formula φ is true in $\mathcal{M}$, **V** has a winning strategy; if a formula φ is false in $\mathcal{M}$, **F** has a winning strategy. □

1.3.4 Truth, Logical Consequence, Validity, Satisfiability

Definition 1.3.12. We define *propositional logic* PL as the triple $\mathsf{PL} := (\mathcal{L}_{\mathsf{PL}}, \mathcal{C}_{\mathsf{PL}}, \models_{\mathsf{PL}})$, *modal logic* as the triple $\mathsf{ML} := (\mathcal{L}_{\mathsf{ML}}, \mathcal{C}_{\mathsf{ML}}, \models_{\mathsf{ML}})$, and *first-order logic* as the triple $\mathsf{FO} := (\mathcal{L}_{\mathsf{FO}}, \mathcal{C}_{\mathsf{FO}}, \models_{\mathsf{FO}})$. □

The realm of logics is vast and PL, ML and FO are only but a few of them. We will introduce many others in these lecture notes (in Sections 3.3, 4.2.1, 5.8 for instance). However, all the logics have in common to deal with the same notions of *truth*, *logical consequence*, *validity* and *satisfiability*.

Definition 1.3.13. Let $\mathsf{L} = (\mathcal{L}, \mathcal{C}, \models) \in \{\mathsf{PL}, \mathsf{ML}, \mathsf{FO}\}$ and let $\Gamma \subseteq \mathcal{L}$, $\varphi \in \mathcal{L}$ and $\mathcal{M} \in \mathcal{C}$. We write $\mathcal{M} \models \Gamma$ when for all $\psi \in \Gamma$, we have $\mathcal{M} \models \psi$. Then, we say that

- φ is *true* (*satisfied*) at $\mathcal{M}$ or $\mathcal{M}$ is a *model* of φ when $\mathcal{M} \models \varphi$;
- φ is a *logical consequence* of Γ, written $\Gamma \models_{\mathsf{L}} \varphi$, when for all $\mathcal{M} \in \mathcal{C}$, if $\mathcal{M} \models \Gamma$ then $\mathcal{M} \models \varphi$;
- φ is *valid*, written $\models_{\mathsf{L}} \varphi$, when for all models $\mathcal{M} \in \mathcal{C}$, we have $\mathcal{M} \models \varphi$;
- φ is *satisfiable* when $\neg\varphi$ is not valid in $\mathcal{C}$, *i.e.* when there is a model $\mathcal{M} \in \mathcal{C}$ such that $\mathcal{M} \models \varphi$. □

1.4 Axioms, Inference Rules and Completeness

Validity of a formula φ is defined abstractly and non-constructively as truth of φ in each model (be it interpretation, Kripke model or structure). How can we describe the form of these validities more concretely and constructively? One concrete method is to provide a *proof system* (or *deductive calculus*) that will *axiomatize* these validities.

1.4.1 Deductive Calculus

In the sequel, a *logic* L is an element of $\{\mathsf{PL}, \mathsf{ML}, \mathsf{FO}\}$ and we consider the *language* $\mathcal{L} := \mathcal{L}_{\mathsf{L}}$.

Definition 1.4.1 (Proof system). A *proof system* $\mathcal{H}$ for $\mathcal{L}$ is a set of formulas of $\mathcal{L}$ called *axioms* and a set of *inference rules*. Let $\Gamma \subseteq \mathcal{L}$ and let $\varphi \in \mathcal{L}$. We say that φ is *provable* (from Γ) in $\mathcal{H}$ or a *theorem* of $\mathcal{H}$, written $\vdash_{\mathcal{H}} \varphi$ (resp. $\Gamma \vdash_{\mathcal{H}} \varphi$), when there is a *proof* of φ (from Γ) in $\mathcal{H}$, that is, a finite sequence of formulas ending in φ such that each of these formulas is:

1. either an instance of an axiom of $\mathcal{H}$ (or a formula of Γ);
2. or the result of applying a rule of inference to preceding formulas. □

A proof system should produce only valid principles (this property is called *soundness*) and hopefully all of them (this property is called *completeness*).

Definition 1.4.2 (Soundness and completeness). Let $\mathcal{H}$ be a proof system for $\mathcal{L}$. Then,

- $\mathcal{H}$ is *sound* for $\mathcal{L}$ w.r.t. $\mathcal{C}$ when for all $\varphi \in \mathcal{L}$, if $\vdash_{\mathcal{H}} \varphi$, then $\models_{\mathsf{L}} \varphi$.
- $\mathcal{H}$ is *(strongly) complete* for $\mathcal{L}$ w.r.t. $\mathcal{C}$ when for all $\varphi \in \mathcal{L}$ (and all $\Gamma \subseteq \mathcal{L}$), if $\models_{\mathsf{L}} \varphi$, then $\vdash_{\mathcal{H}} \varphi$ (resp. if $\Gamma \models_{\mathsf{L}} \varphi$, then $\Gamma \vdash_{\mathcal{H}} \varphi$)

A (modal) logic can also be sound and complete w.r.t. a class of *frames*. □

Note that there might be several proof systems which are sound and complete w.r.t. the same class of models. In any case, by the very definition of a proof, if a logic is sound and *strongly* complete, then it should also be *compact*:

Definition 1.4.3 (Compactness). A logic L is *compact* when for all $\Gamma \subseteq \mathcal{L}$ and all $\varphi \in \mathcal{L}$, the following equivalent statements hold:

- If $\Gamma \models \varphi$, then for some finite $\Gamma_0 \subseteq \Gamma$ we have $\Gamma_0 \models \varphi$;
- If every finite subset Γ_0 of Γ is satisfiable, then Γ is satisfiable. □

There are two main kinds of proof systems, namely *sequent calculi* (Gentzen calculi and natural deduction) and *Hilbert system*. In these lecture notes, we only consider Hilbert systems.

1.4.2 Axiomatizing the Validities of PL, ML and FO

Definition 1.4.4 (Proof systems $\mathcal{H}_{\mathsf{PL}}, \mathcal{H}_{\mathsf{ML}}$ and $\mathcal{H}_{\mathsf{FO}}$). The proof systems $\mathcal{H}_{\mathsf{PL}}, \mathcal{H}_{\mathsf{ML}}$ and $\mathcal{H}_{\mathsf{FO}}$ for the languages $\mathcal{L}_{\mathsf{PL}}, \mathcal{L}_{\mathsf{ML}}$ and $\mathcal{L}_{\mathsf{FO}}$ are defined in Figures 1.1, 1.2 and 1.3 respectively. The system $\mathcal{H}_{\mathsf{ML}}$ is often denoted K in the literature. □

Example 1.4.1 (Distribution axiom and rule).

(i) The following formulas are all provable in $\mathcal{H}_{\mathsf{PL}}$, *i.e.* they are all theorems of $\mathcal{H}_{\mathsf{PL}}$:

$$((p \to q) \wedge (p \to r)) \to (p \to (q \wedge r)) \quad \text{(PL1)}$$
$$p \to (q \to (p \wedge q)) \quad \text{(PL2)}$$
$$(p \wedge q) \to p \quad \text{(PL3)}$$
$$(p \wedge q) \to q \quad \text{(PL4)}$$
$$(p \to (q \to r)) \to ((p \wedge q) \to r) \quad \text{(PL5)}$$

We use these theorems to prove that $\Box(p \wedge q) \to (\Box p \wedge \Box q)$ is provable in $\mathcal{H}_{\mathsf{ML}}$ (formally, $\vdash_{\mathcal{H}_{\mathsf{ML}}} \Box(p \wedge q) \to (\Box p \wedge \Box q)$):

$\varphi \to (\psi \to \varphi)$ (Axiom 1)

$(\varphi \to (\psi \to \chi)) \to ((\varphi \to \psi) \to (\varphi \to \chi))$ (Axiom 2)

$(\neg\psi \to \neg\varphi) \to (\varphi \to \psi)$ (Axiom 3)

$$\frac{\varphi \to \psi \quad \varphi}{\psi}$$ (Modus Ponens)

Figure 1.1: Proof system $\mathcal{H}_{\mathsf{PL}}$ for $\mathcal{L}_{\mathsf{PL}}$

The axioms and the rule of inference of $\mathcal{H}_{\mathsf{PL}}$ ($\mathcal{H}_{\mathsf{PL}}$)

$\Box_i(\varphi \to \psi) \to (\Box_i\varphi \to \Box_i\psi)$ (Modal Distributivity)

$$\frac{\varphi}{\Box_i\varphi}$$ (Necessitation)

Figure 1.2: Proof system $\mathcal{H}_{\mathsf{ML}}$ for $\mathcal{L}_{\mathsf{ML}}$

The axioms and the rule of inference of $\mathcal{H}_{\mathsf{PL}}$ ($\mathcal{H}_{\mathsf{PL}}$)

$\forall x(\varphi(x) \to \psi(x)) \to (\forall x\varphi(x) \to \forall x\psi(x))$ (Distributivity)

$\forall x\varphi(x) \to \varphi(x)[x/t]$, where t is substitutable for x in φ (Universal Instantiation)

$\varphi \to \forall x\varphi$, where x does not occur free in φ (Vacuous Universal Generalization)

$$\frac{\varphi}{\forall x\varphi}$$ (Generalization)

Figure 1.3: Proof system $\mathcal{H}_{\mathsf{FO}}$ for $\mathcal{L}_{\mathsf{FO}}$ (without equality)

1	$(p \wedge q) \to p$	by (PL3)
2	$\Box\left((p \wedge q) \to p\right)$	Necessitation rule on 1
3	$\Box\left((p \wedge q) \to p\right) \to (\Box(p \wedge q) \to \Box p)$	Modal Distributivity
4	$\Box(p \wedge q) \to \Box p$	Modus Ponens on 2, 3
5	$(p \wedge q) \to q$	by (PL4)
6	$\Box\left((p \wedge q) \to q\right)$	Necessitation rule on 5
7	$\Box\left((p \wedge q) \to q\right) \to (\Box(p \wedge q) \to \Box q)$	Modal Distributivity
8	$\Box(p \wedge q) \to \Box q$	Modus Ponens on 6, 7
9	$(\Box(p \wedge q) \to \Box p) \to \Big((\Box(p \wedge q) \to \Box q) \to \big((\Box(p \wedge q) \to \Box p) \wedge (\Box(p \wedge q) \to \Box q)\big)\Big)$	by (PL2)
10	$(\Box(p \wedge q) \to \Box q) \to \big((\Box(p \wedge q) \to \Box p) \wedge (\Box(p \wedge q) \to \Box p)\big)$	Modus Ponens on 4, 9
11	$(\Box(p \wedge q) \to \Box p) \wedge (\Box(p \wedge q) \to \Box q)$	Modus Ponens on 8, 10
12	$(\Box(p \wedge q) \to \Box p) \wedge (\Box(p \wedge q) \to \Box q) \to (\Box(p \wedge q) \to (\Box p \wedge \Box q))$	by (PL1)
13	$\Box(p \wedge q) \to (\Box p \wedge \Box q)$	by Modus Ponens 11, 12

(ii) If $\varphi \to \psi$ is provable in $\mathcal{H}_{\mathsf{ML}}$, then so is $\Box\varphi \to \Box\psi$:

1	$\varphi \to \psi$	provable by assumption
2	$\Box(\varphi \to \psi)$	Necessitation rule on 1
3	$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$	Modal Distributivity
4	$\Box\varphi \to \Box\psi$	Modus Ponens on 2, 3

□

Theorem 1.4.1 (Soundness and completeness). *The proof system $\mathcal{H}_{PL}$ ($\mathcal{H}_{ML}$ and $\mathcal{H}_{FO}$) is sound and strongly complete for $\mathcal{L}_{PL}$ (resp. $\mathcal{L}_{ML}$ and $\mathcal{L}_{FO}$) w.r.t. $\mathcal{C}_{PL}$ (resp. $\mathcal{C}_{ML}$ and $\mathcal{C}_{FO}$).*

Corollary 1.4.1 (Compactness of PL, ML and FO). *The logics PL, ML and FO are compact.*

1.4.3 Increasing the Deductive Power of Modal Logics

We can increase the deductive strength of modal logic ML, by means of further axioms on top of our minimal proof system $\mathsf{K} := \mathcal{H}_{\mathsf{ML}}$.

Definition 1.4.5 (Proof systems T, S4, and S5). The proof system T adds the axiom schema $\Box_i\varphi \to \varphi$ to K, or equivalently, $\varphi \to \Diamond_i\varphi$. Next, $\mathsf{S4}$ adds the $\mathsf{4}$ axiom $\Box_i\varphi \to \Box_i\Box_i\varphi$ to T, or equivalently $\Diamond_i\Diamond_i\varphi \to \Diamond_i\varphi$. Finally, $\mathsf{S5}$ adds the following axiom to $\mathsf{S4}$: $\Diamond_i\varphi \to \Box_i\Diamond_i\varphi$, or equivalently, $\neg\Box_i\varphi \to \Box_i\neg\Box_i\varphi$. □

These axioms *correspond* to first-order frame properties:

Proposition 1.4.1. *Let F be a frame. We have*

- $F \models \Box_i p \to \Box_i \Box_i p$ *if, and only if, the relation* R_i *is* transitive*:* i.e., $\forall xyz(R_i xy \wedge R_i yz \to R_i xz)$*;*
- $F \models \Box_i p \to p$ *if, and only if, the relation* R_i *is* reflexive*:* i.e., $\forall x R_i xx$*;*
- $F \models \neg\Box_i p \to \Box_i \neg\Box_i p$ *if, and only if, the relation* R_i *is* euclidean*:* i.e., $\forall xyz(R_i xy \wedge R_i xz \to R_i yz)$*.*

The notion of correspondence is dealt with in *modal correspondence theory* (van Benthem, 2001). Among other questions, it addresses the following ones (See (Blackburn et al., 2001) for more details):

1. When does a given modal axiom have a first-order frame correspondent ?
2. When does a first-order frame property have a modal definition ?

Note that modal formulas may also sometimes have *second-order* correspondent (*e.g.* Löb's axiom $\Box_i(\Box_i\varphi \to \varphi) \to \Box_i\varphi$ corresponds to the property of transitivity and reverse well-foundedness of accessibility relations).

Our formulas T, $\mathsf{4}$ and $\mathsf{5}$ are *Sahlqvist formulas*. These formulas are such that their first-order frame correspondent can be computed algorithmically and such that we have:

Theorem 1.4.2. *Let* Γ *be a set of Sahlqvist axioms. The proof system* $\mathsf{K}\Gamma$ *is sound and strongly complete w.r.t. the first-order class of frames defined by* Γ*.*

Corollary 1.4.2. *The proof system* T *(resp.* $\mathsf{S4}$*,* $\mathsf{S5}$*) is sound and complete for* $\mathcal{L}_{\mathsf{ML}}$ *w.r.t. the class of frames with reflexive (resp. reflexive and transitive, equivalent) accessibility relations.*

1.5 List of Logics

We list below some well-known logics. This list is obviously non exhaustive. For more information about these logics, the interested reader can consult the following references as introductory texts and for further pointers: (Goble, 2001; Priest, 2011; van Benthem, 2010; Gabbay and Guenthner, 2001; Restall, 2000).

- Variants and weakenings of PL:
  1. *Intuitionistic logics*: the excluded middle $\varphi \vee \neg\varphi$ is not valid anymore.
  2. *Linear logics*: the contraction and weakening rules of Gentzen sequent calculus are not valid anymore. (In computer science, it is used for systems where ressources cannot be used infinitely often.)
  3. *Relevant logics*: the weakening rules of Gentzen sequent calculus are not valid anymore. Hence, $\varphi \to (\psi \to \varphi)$ is not valid (informally, ψ is not *relevant* to the derivation of φ here).

4. *Non-monotonic logics*: monotonicity, *i.e.* from $\Gamma \models \varphi$ infer $\Gamma \cup \{\psi\} \models \varphi$, is no longer valid;
5. *Conditional logics*: the definition of material implication as $\varphi \to \psi := \neg\varphi \vee \psi$ does not hold anymore. (It attempts to model more faithfully our intuitive understanding of conditionals and counterfactuals 'if φ then ψ'.)
6. *Many-valued logics*: logics with more truth values than just T and F. *Fuzzy logic* is a many-valued logic.
7. *Paraconsistent logics*: the principle of explosion (or *ex contradictione sequitur quodlibet*) $\psi \wedge \neg\psi \to \varphi$ is not valid anymore. (It attempts to deal with contradictions while avoiding trivial theories.)
8. *Quantum logic*: the distributive law $p \wedge (q \vee r) \leftrightarrow (p \wedge q) \vee (p \wedge r)$ is not valid anymore. (It attempts to respect the postulates of quantum mechanics in physics.)

Many of the above logics are *substructural logics* (Restall, 2000).

- Variants, weakenings and extensions of FO:
  1. *Free logics*: allow for terms that do not denote any object and for models that have an empty domain.
  2. *Independence-friendly logic*: it allows one to express independence relations between quantified variables as in the formula $\forall a \forall b \exists c/b \exists d/a \varphi(a, b, c, d)$ (x/y should be read as "x is independent of y"). It has a 'game' semantics based on imperfect information games.
  3. *Higher-order logics*: second-order logic (quantification over predicates), third-order logic (quantification over predicates of predicates),...

- Variants of ML:
  1. *Epistemic logics, preference and deontic logics, temporal logics*: to reason about knowledge, obligations and permissions, time respectively. (Temporal logics LTL, CTL,... are used in computer science for formal verification.)
  2. *Dynamic logics (e.g. PDL)*: to reason about programs, and also about actions and events in artificial intelligence.
  3. *Provability logics*: to reason about proofs.
  4. *Description logics*: to reason about the concepts of an application domain and their relations. It is also a fragment of FO.

- Other logics and combinations of logics:
  1. *First-order modal logic*: a combination of ML with FO (possible worlds of models are identified with first-order structures).

2. *Probabilistic logics, possibilistic logics*: to reason about probability and uncertainty.
3. *Hoare logics*: to reason about programs and in particular programs that manipulate pointer data structures (*Separation logic*).
4. *Spatial logics*: to reason about spatial notions.

1.6 Further Reading

Section 1.2 is based on (van Benthem et al., 2013; van Benthem, 2010; Ben-Ari, 2012) and some lecture notes of Moshe Vardi (in this section, some parts are directly copy-pasted from them). See these references for a general introduction to formal logic. For more details about modal logic, see the books of Blackburn et al. (2001) or van Benthem (2010) and for more advanced readers the handbook of modal logic (van Benthem et al., 2007). For a mathematical presentation of propositional logic and first-order logic, see (Smullyan, 1968; Kleene et al., 1971; Enderton, 1972), and for a more 'computer science' oriented perspective, see (Huth and Ryan, 2004; Ben-Ari, 2012). Note that the book of Huth and Ryan (2004) gives a presentation of logic where all the proof systems are not Hilbert systems like here but natural deduction systems. Finally, questions such as 'what is logic ?' that we raised in the introduction of this part are addressed in the philosophy of logic (Jacquette, 2007; Read, 1995).

Chapter 2

Decidability, Complexity and Expressiveness

"What can be said at all can be said clearly, and what we cannot talk about we must pass over in silence."

– Ludwig Wittgenstein, *Tractatus Logico-Philosophicus* , 1922

2.1 Introduction

Once a logic is defined, we can study its properties. In particular, we can study whether it is more expressive than another logic, that is, whether it can express more things about a given model than another logic (this is called *expressiveness*). We can also study whether it is possible to find an algorithm that will solve a specific problem formulated in this logic (this is called *decidability*) and how complex and hard it will be to solve this problem with such an algorithm (this is called *computational complexity*). These three problems will be the topic of this chapter and the chapter is organized accordingly.

Section 2.2 deals with decidability and introduces two standard techniques to prove decidability of a logic: the tableau method and the finite model property. Section 2.3 deals with expressiveness and introduces a number of games that can be used to prove that a logic is more expressive than another. Section 2.4 deals with computational complexity and introduces the main decision problems defined for logics, as well as the main complexity classes.

2.2 Decidability

Hilbert proof systems provide a means to (finitely) characterize and *recursively enumerate* the set of validities of a given logic. But they do not provide a way to *decide* or *test* whether a given formula is a validity of the logic.

Generally speaking, a *decision problem* in logic is a problem that takes some logical objects as input (formula, models) and yields as output a decision, "yes" or "no", depending on the nature of the problem. A *decision procedure* for a decision problem is an algorithm, a 'mechanical' method that terminates and gives an answer to any given

Algorithm 2.2.1.

Input: A formula $\varphi \in \mathcal{L}_{\mathsf{PL}} \cup \mathcal{L}_{\mathsf{ML}} \cup \mathcal{L}_{\mathsf{FO}}$.

Output: A tableau $\mathcal{T}$ for φ.

1. Initially, $\mathcal{T}$ is a tree consisting of a single root node labeled with φ, if $\varphi \in \mathcal{L}_{\mathsf{PL}} \cup \mathcal{L}_{\mathsf{FO}}$ (or with $(\ell\ \varphi)$, if $\varphi \in \mathcal{L}_{\mathsf{ML}}$).
2. Repeat the following steps as long as possible:
   (a) *Choose* a branch which is neither closed nor open and choose a formula ψ to decompose (resp. a labeled formula $(\ell\ \psi)$ or a pair of labeled formula $(\ell\ \psi)$ and relation term $(R\ \ell\ \ell')$) not selected before on this branch.
   (b) *Apply* the appropriate tableau rule of Figures 2.2, 2.3 and 2.4 to ψ (or the pair $(\ell\ \psi)$, $(R\ \ell\ \ell')$):
   - if the rule is a β-rule, add two successor nodes to the branch labeled with the instantiation of the denominator(s) of that rule,
   - otherwise, add a unique successor node labeled with the instantiation of the denominator(s) of that rule.

   (c) i. *Label* by $\times$ (*closed*) the (new) branches which contain a formula and its negation.
   ii. *Label* by $\odot$ (*open*) the (new) branches where there are no more formulas to decompose.

Figure 2.1: Construction of a tableau

instance of this problem. When such a decision procedure exists, the decision problem is *decidable*, and *undecidable* otherwise.

2.2.1 Tableau Method: a Decision Procedure

Tableaux are decisions procedure that provide an answer to the validity decision problem for PL and ML, but not for FO. Another well-known decision procedure is *resolution* which is the theoretical basis of *logic programming* (Gabbay and Robinson, 1998).

Definition 2.2.1 (Label, labeled formula and relation term). Let S be an infinite set whose elements are called *labels*. A *labeled formula* is an expression of the form $(\ell\ \varphi)$ where ℓ is a label and φ is a formula (typically of $\mathcal{L}_{\mathsf{ML}}$). A *relation term* is an expression of the form $(R\ \ell\ \ell')$ where $\ell, \ell' \in S$. □

Definition 2.2.2 (Tableau rules). A *(prefixed) tableau* is a finite tree whose nodes are labeled with formulas (resp. prefixed formulas). The tableau tree for a formula is

constructed as shown in Algorithm 2.2.1 of Figure 2.1. The *tableau rules of PL* are represented in Figure 2.2. The *tableau rules of FO* are obtained by adding the rules of Figure 2.3 to those of PL. The *tableau rules for ML* are obtained by prefixing the formulas of the rules of PL and by adding the rules of Figure 2.4.

In the tableau rules, the formulas above the horizontal lines are called *numerators* and those below are called *denominators*. If these are separated by vertical line(s), the tableau rule is called a *β-rule*. □

Theorem 2.2.1 (Soundness and completeness). *Let $\varphi \in \mathcal{L}_{PL} \cup \mathcal{L}_{ML} \cup \mathcal{L}_{FO}$. Then, φ is satisfiable if, and only if, the tableau for φ is open.*

Example 2.2.1. The tableau of Figure 2.5 shows that the formula $p \to (q \to (p \wedge q))$ is valid in PL. The tableau of Figure 2.6 shows that the formula $\Box(p \to q) \to (\Box p \to \Box q)$ is valid in ML. □

Theorem 2.2.2 (Termination). *The construction of a tableau for any formula of $\mathcal{L}_{PL}$ or $\mathcal{L}_{ML}$ terminates, but not necessarily for formulas of $\mathcal{L}_{FO}$.*

Corollary 2.2.1 (Decidability of PL and ML). *The validity problems of PL and ML are decidable.*

Proof. It follows easily from Theorems 2.2.1 and 2.2.2. □

Theorem 2.2.3 (Undecidability of FO). *The validity problem of FO (sometimes called the Hilbert's Entscheidungs problem) is undecidable.*

Many fragments of FO *are* decidable, such as *monadic* first-order logic (only unary predicates), the *two-variable* fragment of FO, the *guarded fragment*, and many more (Börger et al., 2001).

2.2.2 Finite Model Property

A problem is decidable if both the problem and its complement are recursively enumerable. We already know that the validity problem is recursively enumerable, because the set of validities is axiomatizable. To prove that the complement of the validity problem, *i.e.* the satisfiability problem, is recursively enumerable, it suffices to show that the logic has the *finite model property*: any satisfiable formula is satisfiable in a *finite* model. Indeed, if a logic has the finite model property, we can enumerate the set of finite models: if the formula is satisfiable it will eventually be satisfied by a finite model.

Proving that a given logic has the *effective* finite model property is another means to prove that its validity/satisfiability problem is decidable. In that case, we must find a computable bound on the size of the model that satisfies the formula. A general approach that works for a number of modal logics is based on the *filtration* method.

Definition 2.2.3 (Filtration of a model). Let $\mathcal{M} = (W, R, V)$ be a Kripke model and let $\varphi \in \mathcal{L}_{\mathsf{ML}}$. The (smallest) *filtration of model $\mathcal{M}$ by φ* is the Kripke model denoted

$$\frac{\varphi_1 \wedge \varphi_2}{\begin{array}{c}\varphi_1\\ \varphi_2\end{array}}\ \wedge \qquad \frac{\neg\neg\varphi}{\varphi}\ \neg\neg \qquad \frac{\neg(\varphi_1 \wedge \varphi_2)}{\neg\varphi_1 \mid \neg\varphi_2}\ \neg\wedge$$

$$\frac{\neg(\varphi_1 \vee \varphi_2)}{\begin{array}{c}\neg\varphi_1\\ \neg\varphi_2\end{array}}\ \neg\vee \qquad \frac{\neg(\varphi_1 \to \varphi_2)}{\begin{array}{c}\varphi_1\\ \neg\varphi_2\end{array}}\ \neg\to \qquad \frac{\varphi_1 \vee \varphi_2}{\varphi_1 \mid \varphi_2}\ \vee \qquad \frac{\varphi_1 \to \varphi_2}{\neg\varphi_1 \mid \varphi_2}\ \to$$

Figure 2.2: Tableau rules for PL (first raw) and derived rules (second raw)

$$\frac{\forall x\varphi}{\varphi[x/t]}\ \forall \text{ where } t \text{ is a variable-free term} \qquad \frac{\neg\forall x\varphi}{\neg\varphi[x/c]}\ \neg\forall \text{ where } c \text{ is a new constant}$$

$$\frac{\neg\exists x\varphi}{\neg\varphi[x/t]}\ \neg\exists \text{ where } t \text{ is a variable-free term} \qquad \frac{\exists x\varphi}{\varphi[x/c]}\ \exists \text{ where } c \text{ is a new constant}$$

Figure 2.3: Specific tableau rules for FO (first raw) and derived rules (second raw)

$$\frac{(\ell\ \Box\varphi) \quad (R\ \ell\ \ell')}{(\ell'\ \varphi)}\ \Box \qquad \frac{(\ell\ \neg\Box\varphi)}{\begin{array}{c}(R\ \ell\ \ell')\\ (\ell'\ \neg\varphi)\end{array}}\ \neg\Box$$

$$\frac{(\ell\ \neg\Diamond\varphi) \quad (R\ \ell\ \ell')}{(\ell'\ \neg\varphi)}\ \neg\Diamond \qquad \frac{(\ell\ \Diamond\varphi)}{\begin{array}{c}(R\ \ell\ \ell')\\ (\ell'\ \varphi)\end{array}}\ \Diamond$$

Figure 2.4: Specific tableau rules for ML (first raw) and derived rules (second raw)

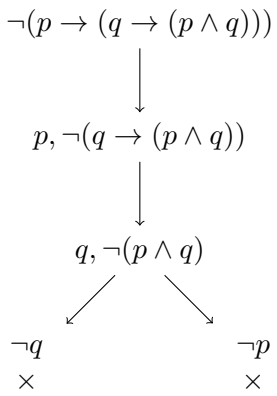


Figure 2.5: Tableau for the formula $\neg(p \to (q \to (p \wedge q)))$

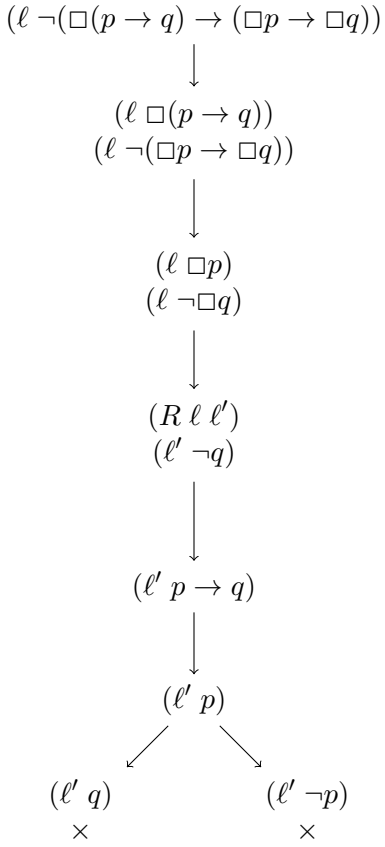


Figure 2.6: Tableau for the formula $\neg(\Box(p \to q) \to (\Box p \to \Box q))$

$\mathcal{M}^{\sim_\varphi} = (W^{\sim_\varphi}, R^{\sim_\varphi}, V^{\sim_\varphi})$ and defined as follows. First, we define the equivalence relation $\sim_\varphi$ between possible worlds by $\sim_\varphi := \{(w, v) : \mathcal{M}, w \models \psi$ iff $\mathcal{M}, v \models \psi$ for all $\psi \in \mathsf{Sub}(\varphi)\}$. Equivalence classes of $\sim_\varphi$ are denoted $w^{\sim_\varphi}, v^{\sim_\varphi}, \ldots$ Then,

- $W^{\sim_\varphi} := \{w^{\sim_\varphi} : w \in W\}$;
- $R^{\sim_\varphi} := \{(w^{\sim_\varphi}, v^{\sim_\varphi}) \in W^{\sim_\varphi} \times W^{\sim_\varphi} :$ there are $s \in w^{\sim_\varphi}$ and $t \in v^{\sim_\varphi}$ such that $(s, t) \in R\}$;
- $V^{\sim_\varphi}(p, w^{\sim_\varphi}) = T$ iff $V(p, w) = T$ if $p \in \mathsf{Sub}(\varphi)$. □

The definition of $R^{\sim_\varphi}$ can vary, giving rise to other definitions of filtrations. This one is the *smallest* filtration.

Fact 2.2.1. *For all Kripke models $\mathcal{M}$, all $w \in \mathcal{M}$ and all $\psi \in \mathsf{Sub}(\varphi)$, we have $\mathcal{M}, w \models \psi$ if, and only if, $\mathcal{M}^{\sim_\varphi}, w^{\sim_\varphi} \models \psi$. Moreover, $\mathcal{M}^{\sim_\varphi}$ contains at most $2^{|\mathsf{Sub}(\varphi)|}$ worlds.*

Theorem 2.2.4 (Effective finite model property of ML). *Every satisfiable modal formula $\varphi \in \mathcal{L}_{ML}$ is satisfiable on a finite Kripke model containing at most $2^{|\mathsf{Sub}(\varphi)|}$ worlds.*

Corollary 2.2.2. *The validity problem of ML is decidable.*

Proof. Consider $\varphi \in \mathcal{L}_{\mathsf{ML}}$. Enumerate all the Kripke models of size $2^{|\mathsf{Sub}(\varphi)|}$ (their number is finite) and check for each of them whether they make φ true. If one does, φ is satisfiable, otherwise φ is unsatisfiable. □

Theorem 2.2.5. *FO does not have the finite model property.*

Proof. Let φ be the formula that says that $<$ is an irreflexive transitive order where every point has a successor. The natural numbers with the relation $<$ is a model. However, φ has only infinite models: any finite transitive model in which each point has a successor must have loops, which are forbidden by the irreflexivity. □

However, FO satisfies the following property:

Theorem 2.2.6 (Löwenheim-Skolem). *If a countable set of formulas is satisfiable then it is satisfiable in a* countable *domain.*

Uncountable sets such as the real numbers can be described by countably many axioms (formulas). Thus formulas that describe real numbers also have a countable model in addition to the standard uncountable model ! Such models are called *non-standard* models. The *Löwenheim-Skolem* and the *Compactness* properties are often used to show undefinability of mathematical properties in FO: a standard example is *finiteness* of the domain. They are also characteristic of FO:

Theorem 2.2.7 (Lindström). *FO is the 'strongest logic' having both the Compactness and Löwenheim-Skolem properties.*

A recent characerization result also holds for ML: an 'abstract modal logic' L extending the modal logic ML equals ML if, and only if, L satisfies (a) Invariance for Bisimulation and (b) Compactness (see (van Benthem, 2010, Th. 73) for more details).

2.3 Expressive Power and Invariance

Independently from any language, structures have mathematical relations (isomorphism, bisimulations,...). A fundamental measure of the expressive power of a language is its 'power of distinction' between different structures, that is, to what extent the language can distinguish two different structures. A semantic relation of *invariance* between structures can be matched to a particular logic: e.g. bisimulation with ML, isomorphism with FO.

2.3.1 First-order Logic: Ehrenfeucht-Fraïssé games

Definition 2.3.1 (Isomorphism and partial isomorphism). Two structures $\mathcal{M}$ and $\mathcal{N}$ are *isomorphic* if there is a bijection F from $W^{\mathcal{M}}$ to $W^{\mathcal{N}}$ such that for all predicate R, all $w_1, \ldots, w_k \in W$, all $c \in CONS$

$$F(c^{\mathcal{M}}) = c^{\mathcal{N}}$$
$$(w_1, \ldots, w_k) \in R \quad \text{iff} \quad (F(w_1), \ldots, F(w_k)) \in R^{\mathcal{N}}$$

A *partial isomorphism* is an injective partial function (often finite) between subsets of the domains of two structures, which is an isomorphism if it is restricted to its domain and range. □

Proposition 2.3.1. *For all structures $\mathcal{M}$ and $\mathcal{N}$, condition* (1) implies *condition* (2)*:*

1. *$\mathcal{M}$ and $\mathcal{N}$ are isomorphic*
2. *$\mathcal{M}$ and $\mathcal{N}$ are* elementary equivalent*,* i.e. *they both make true the same sentences.*

If the structures $\mathcal{M}$ and $\mathcal{N}$ are finite*, then conditions* (1) *and* (2) *are in fact* equivalent*.*

The proposition does not hold with *infinite* structures: *finiteness* of domains is not expressible in FO (due to the *compactness* of FO).

Example 2.3.1. The following pairs of structures are *not* isomorphic, but

- $(\mathbb{Q}, <^{\mathbb{Q}})$ and $(\mathbb{R}, <^{\mathbb{R}})$ satisfy the same formulas of FO with $<$;
- $(\mathbb{N}, <^{\mathbb{N}})$ and $(\mathbb{N} + \mathbb{Z}, <^{\mathbb{N}+\mathbb{Z}})$ satisfy the same formulas of FO with $<$. □

FO can distinguish $\mathbb{Q}$ from $\mathbb{R}$ with a richer *vocabulary* if one adds the multiplication function: $\sqrt{2} \in \mathbb{R} - \mathbb{Q}$ is expressible by $\exists x : x \cdot x = 2$.

Definition 2.3.2 (Ehrenfeucht-Fraïssé games). Consider two structures $\mathcal{M}$ and $\mathcal{N}$ and two players called Duplicator **D** and Spoiler **S**. Fix $k \in \mathbb{N}$ representing the number of rounds. At each round:

1. **S** chooses one of the structures (say $\mathcal{M}$) and picks an element d in its domain;

2. **D** chooses an element e in the other structure, and the pair (d, e) is added to the current list of matched elements.

At the end of the k rounds, if the list of matched elements is a partial isomorphism, **D** wins; otherwise, **S** has won the game. □

Example 2.3.2. Wins for **S** are correlated with specific first-order formulas φ that bring out a difference between the two structures. The number of rounds of the game is also correlated with the *quantifier depth* of φ, written $\mathsf{qd}(\varphi)$, defined as follows: $\mathsf{qd}(\varphi) := 0$ for atomic formulas φ, $\mathsf{qd}(\neg\varphi) := \mathsf{qd}(\varphi)$, $\mathsf{qd}(\varphi \wedge \psi) := max\{\mathsf{qd}(\varphi), \mathsf{qd}(\psi)\}$, and $\mathsf{qd}(\forall x\varphi) := \mathsf{qd}(\varphi) + 1$.

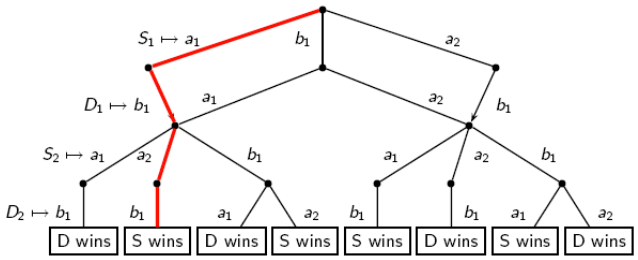
1. A simple example:

$$\mathcal{M}: \quad a_1 \qquad a_2 \qquad\qquad \mathcal{N}: \quad b_1$$

Round 1: **S** chooses a_1 in $\mathcal{M}$ **D** chooses b_1 in $\mathcal{N}$
Round 2: **S** chooses a_2 in $\mathcal{M}$ **D** chooses again b_1 in $\mathcal{N}$

If we stopped after round 1, **D** would trivially win. After round 2, **S** has won, as the map is not a partial isomorphism: the cardinalities do not match. In extensive form game:

$$\mathcal{M}: \quad a_1 \qquad a_2 \qquad\qquad \mathcal{N}: \quad b_1$$



Red: a (the) play induced by a winning strategy for Spoiler **S**.

Difference formula:

$$\varphi := \exists x \exists y \neg(x = y) \qquad\qquad \mathsf{qd}(\varphi) = 2$$

2. Cycles:

Round 1: **S** chooses 1 in $\mathcal{M}$ **D** chooses i in $\mathcal{N}$
Round 2: **S** chooses 2 in $\mathcal{M}$ **D** chooses again j in $\mathcal{N}$
Round 3: **S** chooses 3 in $\mathcal{M}$ **D** chooses again k in $\mathcal{N}$

Difference formula:

$$\varphi := \exists x \exists y \exists z (Rxy \wedge Ryz \wedge Rxz) \qquad \mathsf{qd}(\varphi) = 3$$

S has won, because this match is not a partial isomorphism. But **S** can do better:

Round 1: **S** chooses i in $\mathcal{N}$ **D** chooses 1 in $\mathcal{N}$
Round 2: **S** chooses k in $\mathcal{M}$ **D** chooses any element, and loses

Difference formula:

$$\psi := \exists x \exists y (\neg Rxy \wedge \neg Ryx \wedge \neg(x = y)) \qquad \mathsf{qd}(\psi) = 2$$

3. Integers versus rational: $(\mathbb{Z}, <^{\mathbb{Z}})$ and $(\mathbb{Q}, <^{\mathbb{Q}})$ are two linear orders with different properties: the latter is *dense* and the former is *discrete*.

$$\begin{array}{lccccccc} \mathbb{Z} & \cdots & & -1 & & 0 & 1 & \cdots \\ \mathbb{Q} & & \cdots & & 0 & \frac{1}{5}\ \frac{1}{3} & & \cdots \end{array}$$

D has a winning strategy for the EF-game over 2 rounds. But **S** has a winning strategy for the EF-game in 3 rounds:

Round 1: **S** chooses 0 in $\mathbb{Z}$ **D** chooses 0 in $\mathbb{Q}$
Round 2: **S** chooses 1 in $\mathbb{Z}$ **D** chooses $\frac{1}{3}$ in $\mathbb{Q}$
Round 3: **S** chooses $\frac{1}{5}$ in $\mathbb{Q}$ **D** chooses any element, and loses

The following *difference formula* characterizes the property of density:

$$\varphi := \forall x \forall y (x < y \to \exists z (x < z \wedge z < y)) \qquad \mathsf{qd}(\varphi) = 3 \qquad \square$$

Definition 2.3.3 (Winning strategy). A *winning strategy* for the Duplicator **D** in a k-round EF-game on $\mathcal{M}$ and $\mathcal{N}$ is a sequence $I_0, I_1, \ldots, I_r$ of non-empty sets of partial isomorphisms from $\mathcal{M}$ to $\mathcal{N}$ such that for all $i < r$,

Forth for all $f \in I_i$, all $w \in W^{\mathcal{M}}$, there is $v \in W^{\mathcal{N}}$ and $g \in I_{i+1}$ such that $f \cup \{(w, v)\} \subseteq g$;

Back for all $f \in I_i$, all $v \in W^{\mathcal{N}}$, there is $w \in W^{\mathcal{M}}$ and $g \in I_{i+1}$ such that $f \cup \{(w, v)\} \subseteq g$.

Theorem 2.3.1 (Adequacy theorem). *For all structures $\mathcal{M}, \mathcal{N}$, for all $k \in \mathbb{N}$, the following are equivalent:*

1. *Duplicator **D** has a winning strategy against Spoiler **S** in the k-round EF-game on the structures $\mathcal{M}$ and $\mathcal{N}$;*
2. *$\mathcal{M}$ and $\mathcal{N}$ make true the same sentences up to quantifier depth k.*

2.3.2 Modal Logic: Bisimulation Games

Definition 2.3.4 (Bisimulation). A *bisimulation* between two structures $\mathcal{M}, \mathcal{N}$ is a binary relation $Z \subseteq W^{\mathcal{M}} \times W^{\mathcal{N}}$ such that, whenever wZv:

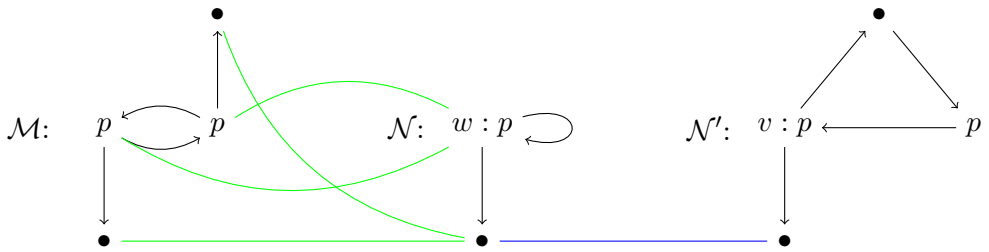
Atom: w and v make true the same propositional letters;

Forth: for all $w' \in W^{\mathcal{M}}$ such that $wR^{\mathcal{M}}w'$, there is $v' \in W^{\mathcal{N}}$ such that $vR^{\mathcal{N}}v'$ and $w'Zv'$;

Back: for all $v' \in W^{\mathcal{N}}$ such that $vR^{\mathcal{N}}v'$, there is $w' \in W^{\mathcal{M}}$ such that wRw' and $w'Zv'$.

Proposition 2.3.2 (Invariance). *Let $\mathcal{M}$ and $\mathcal{N}$ be two Kripke models. If Z is a bisimulation between $\mathcal{M}$ and $\mathcal{N}$ and wZv, then w and v make true the same modal formulas.*

Example 2.3.3. $\mathcal{M}$ and $\mathcal{N}$ are bisimilar: Z is in green. $\mathcal{N}$ and $\mathcal{N}'$ are also bisimilar: Z is in blue. $(\mathcal{N}, w)$ and $(\mathcal{N}', w')$ are not bisimilar: $\mathcal{N}, w \models \Diamond\Diamond\Box\bot$ but $\mathcal{N}', w' \models \neg\Diamond\Diamond\Box\bot$.



Proposition 2.3.3. *Let $\mathcal{M}$ and $\mathcal{N}$ be finite (or image finite) Kripke models. If $w \in \mathcal{M}$ and $v \in \mathcal{N}$ make the same modal formulas true, then there is a bisimulation Z between $\mathcal{M}$ and $\mathcal{N}$ with wZv.*

Definition 2.3.5 (Bisimulation game). Consider two Kripke models $\mathcal{M}$ and $\mathcal{N}$ and a pair $(w, v) \in W^{\mathcal{M}} \times W^{\mathcal{N}}$. Fix $k \in \mathbb{N}$ representing the number of rounds. At each round:

1. **S** chooses one of the models (say $\mathcal{M}$) and picks an element w' in its domain such that wR_iw';
2. **D** then responds with an element $v' \in W^{\mathcal{N}}$ of the other model (here $\mathcal{N}$) such that vR_iv'.

If at any round, w and v are different in their atomic properties or if **D** cannot find a successor, **S** wins. □

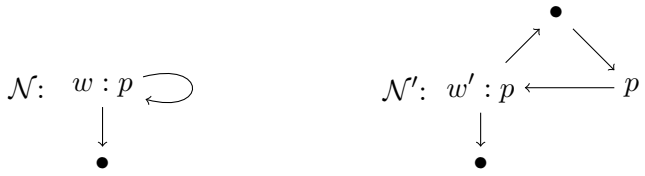
Unlike EF-games, bisimulation games restrict the selection of elements to successors of those previously matched. Bisimulation games are a modification of EF-games.

Theorem 2.3.2 (Adequacy theorem). *For all Kripke models $\mathcal{M}, \mathcal{N}$, all $w \in \mathcal{M}$, $v \in \mathcal{N}$, $k \in \mathbb{N}$, the following are equivalent:*

1. ***D*** *has a winning strategy in a k-round (infinite round) bisimulation game on* $(\mathcal{M}, w)$ *and* $(\mathcal{N}, v)$*;*

2. $(\mathcal{M}, w)$ *and* $(\mathcal{N}, v)$ *make true the same modal formulas up to modal depth k (resp. any modal depth).*

Example 2.3.4. $(\mathcal{N}, w)$ and $(\mathcal{N}', w')$ can be distinguished in a bisimulation game of *3* rounds. Why? Because $\mathcal{N}, w \models \Diamond\Diamond\Box\bot$ but $\mathcal{N}', w' \not\models \Diamond\Diamond\Box\bot$.



Other model comparison games: p-Pebble games, counting games, bijection games... By varying the rules of EF-games, one can investigate and characterize the expressive power of a wide variety of logical languages. For example, the p-Pebble games characterize the fact that two structures make true the same first-order sentences up to quantifier depth k which use at most the variables $x_1, \ldots, x_m$ (free or bound).

2.3.3 Using Model Comparison Games: Definability and Expressiveness

2.3.3.1 Definability

When we want to choose a logical language to specify and reason about a specific class of structures (representing some problems), it is important to know which logical language *defines* such a class.

Definition 2.3.6 (Definability). A class of structures $\mathcal{C}$ is *first-order definable* (*modally definable*) on a class of structures $\mathcal{C}'$ if, and only if, there is a sentence φ (resp. a formula $\varphi \in \mathcal{L}_{\mathsf{ML}}$) such that $\mathcal{C} = \{\mathcal{M} : \mathcal{M} \in \mathcal{C}' \text{ and } \mathcal{M} \models \varphi\}$. □

None of the following properties are definable on the class of finite graph: transitive closure, connectivity, acyclicity, planarity, Eulerian, k-colorability (for each $k \geq 2$), Hamiltonicity... The following corollary of Theorems 2.3.1 and 2.3.2 enables to show in particular that a class $\mathcal{C}$ of structures is not first-order or modally definable on another class $\mathcal{C}'$.

Corollary 2.3.1. *Let* $\mathcal{C}, \mathcal{C}'$ *be two classes of finite structures. The following are equivalent:*

- $\mathcal{C}$ *is* first-order definable *(modally definable) on* $\mathcal{C}'$*;*
- *there exists* $k \in \mathbb{N}$ *such that for all* $\mathcal{M}, \mathcal{N} \in \mathcal{C}'$*, if* $\mathcal{M} \in \mathcal{C}$ *and* $\mathcal{N} \notin \mathcal{C}$*, then* ***S*** *wins the EF-game (resp. bisimulation game) between* $\mathcal{M}, \mathcal{N}$ *in k rounds.*

Example 2.3.5. Proper succession, that is the class of structures satisfying the FO property $\exists y(Rxy \wedge \neg Ryx \wedge P(y))$ is modally undefinable: it is not invariant under bisimulation. □

Other invariances: ultraproducts. There are also other structural characterizations of FO and ML, mathematically deeper than the game analysis. *Keisler's theorem* (Hodges, 1997, Th. 8.5.10) says that a class of structures of FO is definable in FO if it is closed under the formation of *ultraproducts* and *potential isomorphisms*. Likewise for ML: a class of Kripke models $\mathcal{C}$ of ML is definable in ML if, and only if, both the class $\mathcal{C}$ and its complement $\overline{\mathcal{C}}$ are closed under bisimulations and ultraproducts (Blackburn et al., 2001, Th. 2.76). For ML, at the level of frames, we also have the *Goldblatt-Thomason theorem* (Blackburn et al., 2001).

2.3.3.2 Expressiveness

Definition 2.3.7 (Expressiveness). Let two logics $\mathsf{L}_1 = (\mathcal{L}_1, \mathcal{C}, \models_1)$ and $\mathsf{L}_2 = (\mathcal{L}_2, \mathcal{C}, \models_2)$ be given (interpreted in the same class of models $\mathcal{C}$). L_1 is *at least as expressive as* L_2, written $\mathsf{L}_1 \geq \mathsf{L}_2$, when for all $\varphi_2 \in \mathsf{L}_2$, there is $\varphi_1 \in \mathsf{L}_1$ such that $\{\mathcal{M} \in \mathcal{C} : \mathcal{M} \models \varphi_1\} = \{\mathcal{M} \in \mathcal{C} : \mathcal{M} \models \varphi_2\}$. L_1 is *more expressive than* L_2 when $\mathsf{L}_1 \geq \mathsf{L}_2$ but not $\mathsf{L}_2 \geq \mathsf{L}_1$. □

In this section, we consider a FO language over Kripke models $\mathcal{M} = (W, R, V)$ with one binary predicate letter R for the accessibility relation, and unary predicate letters $P, Q, \ldots$ matching propositional letters $p, q, \ldots$ Let variables $x, y, z, \ldots$ range over worlds.

Definition 2.3.8 (Standard translation). Let $\varphi \in \mathcal{L}_{\mathsf{ML}}$. The *standard translation* $ST(\varphi)$ of φ is a first-order formula with one free variable x defined inductively as follows:

$$\begin{array}{rcll} ST(p) & = & P(x) & \\ ST(\neg\varphi) & = & \neg ST(\varphi) & \\ ST(\varphi \wedge \psi) & = & ST(\varphi) \wedge ST(\psi) & \\ ST(\Box\varphi) & = & \forall y(xRy \rightarrow ST(\varphi)[y/x]) & \text{where } y \text{ is a new variable} \end{array}$$

□

Example 2.3.6.

- $ST(\Diamond\varphi) = \exists y(xRy \wedge ST(y))$,
- $ST(\Box\Diamond(p \vee q)) = \forall(xRy \rightarrow \exists z(yRz \wedge (P(z) \vee Q(z)))$. □

Proposition 2.3.4. *For all Kripke model $\mathcal{M}$ and all $w \in \mathcal{M}$, for all $\varphi \in \mathcal{L}_{\mathsf{ML}}$, it holds that*

$$\mathcal{M}, w \models \varphi \quad \textit{iff} \quad \mathcal{M} \models ST(\varphi)[x/w]$$

Corollary 2.3.2. *First-order logic is more expressive than modal logic.*

Proof. It follows from Proposition 2.3.4 and Example 2.3.5. □

From Proposition 2.3.4, we inherit a number of results for ML from FO: compactness, countable model and interpolation (although this later needs some fine-tuning for ML).

Theorem 2.3.3 (Craig's interpolation). *Let φ, ψ be two formulas of ML (or FO). If $\varphi \models \psi$, then there exists an "interpolant" α of ML (resp. FO) with $\varphi \models \alpha$ and $\alpha \models \psi$ such that every non-logical symbol in α occurs in both φ and ψ.*

The following theorem tells that the modal language may be viewed in a natural manner as the bisimulation-invariant fragment of first-order logic.

Theorem 2.3.4 (van Benthem). *Let $\varphi(x) \in \mathcal{L}_{FO}$ be a formula containing unary predicates $P(x)$ and binary predicates $R(x, y)$. The two following statements are equivalent:*

- *There exists a formula $\psi \in \mathcal{L}_{ML}$ such that $\models_{FO} \varphi \leftrightarrow ST(\psi)$;*
- *For all models $\mathcal{M}, \mathcal{N}$, all $w \in \mathcal{M}$, $v \in \mathcal{N}$ such that there exists a bisimulation Z between $\mathcal{M}$ and $\mathcal{N}$ such that wZv, we have $\mathcal{M} \models \varphi(x)[x/w]$ if, and only if, $\mathcal{N} \models \varphi(x)[x/v]$.*

2.4 Computation and Complexity

Even if the validity problem is decidable for propositional and modal logic, the resources needed to answer the problem might be quite different and it may be wildly infeasible in practice. To address such worries and answer many other questions about actual performance of algorithms, computer scientists developed Complexity Theory. One can measure the complexity of a task in terms of *time* (number of steps taken) and *space* (size of the memory employed). Measures for time and space involve some task-dependent variable: often the length of the input formula, or the size of some given finite model. What we are actually measuring are *rates of growth* rather than specific numbers. In Figure 2.7, we recall the main complexity classes. They are ordered according to the computational effort needed to solve problems of each class.

Definition 2.4.1 (Standard decision problems in logic). The three main decision problems in logic are:

Satisfiability: Determining validity, or equivalently, *testing for satisfiability* of given formulas, is answering the question: "Given a formula φ, determine whether φ has a model".

Model checking: Model checking, or *testing for truth* of formulas in given models, is answering the question: "Given a formula φ and a finite model $\mathcal{M}, w$, check whether $\mathcal{M}, w \models \varphi$".

Model equivalence: Model comparison, or *testing for equivalence* of given models, is answering the question: "Given two finite models $\mathcal{M}, w$ and $\mathcal{N}, v$, check if they satisfy the same formulas". □

Usually, the theorem provers for ML are not in the above axiomatic style, but they use a variety of other methods: (1) translation into FO plus "resolution" methods; (2)

P	NP	PSPACE	EXPTIME	...	Undecidable
2SAT Shortest path	3SAT Traveling Salesman	QBF Game of Geography	Sat of CTL Pebble Games		Sat of FO Halting problem
tractable (feasible)	untractable (unfeasible)				

Figure 2.7: Main complexity classes

Decision Problems:	Model Checking	Satisfiability	Model Comparison
Propositional Logic:	linear time (P)	NP	linear time (P)
Modal Logic:	P	PSPACE	P
First-order Logic:	PSPACE	undecidable	NP

Figure 2.8: Complexity profile of the logics PL, ML and FO

the "tableau methods" of the next section; (3) specially optimized modal calculi. (see `http://www.aiml.net` for more details)

Here is an important rule of logic. Modal logic is known for its good balance between expressiveness and computational complexity.

> **Golden Rule: the balance between expressive power and computational effort.** "The more expressive a logic is, the more complex its associated decision problems are, and vice versa."

For undecidable logics, there is also a kind of balance between expressive power and axiomatizability: second-order logic is more expressive than first-order logic, but second-order logic loses axiomatizability.

2.5 Further Reading

My presentation of the tableau method is based on (Bibel and Eder, 1993) and (Fitting, 1993). The rest of this chapter is based on a combination and adaptation of presentations from (van Benthem, 2010; Blackburn et al., 2001) (specially for modal logic), (Enderton, 1972; Ben-Ari, 2012) (for first-order logic). More information about the theory of computation and complexity theory can be found for instance in (Sipser, 2006; Papadimitriou, 2003) and about model theory in (Hodges, 1997).

Part II

Representing and Reasoning about Uncertainty

Introduction to Part II

"An investment in knowledge pays the best interest."
– Benjamin Franklin, *Poor Richard's Almanac*, c. 1750

This part is about logical models for knowledge representation and belief change. We propose logical systems which are intended to represent how agents perceive a situation and reason about it, and how they update their beliefs about this situation when events occur. These agents can be machines, robots, human beings... but they are assumed to be somehow autonomous.

The way a fixed situation is perceived by agents can be represented by statements about the agents' beliefs: for example 'agent A believes that the door of the room is open' or 'agent A believes that her colleague is busy this afternoon'. '*Logical* systems' means that agents can reason about the situation and their beliefs about it: if agent A believes that her colleague is busy this afternoon then agent A infers that he will not visit her this afternoon. We moreover often assume that our situations involve multiple agents which interact with each other. So these agents have beliefs about the situation (such as 'the door is open') but also about the other agents' beliefs: for example, agent A might believe that agent B believes that the door is open. These kinds of beliefs are called higher-order beliefs. *Epistemic logic* (Hintikka, 1962; Fagin et al., 1995; Meyer and van der Hoek, 1995), the logic of belief and knowledge, can capture all these phenomena and will be our main starting point to model such fixed ('static') situations.

Uncertainty can of course be expressed by beliefs and knowledge: for example agent A being uncertain whether her colleague is busy this afternoon can be expressed by 'agent A does not *know* whether her colleague is busy this afternoon'. But we sometimes need to enrich and refine the representation of uncertainty: for example, even if agent A does not know whether her colleague is busy this afternoon, she might consider it more probable that he is actually busy. So other logics have been developed to deal more adequately with the representation of uncertainty, such as probabilistic logic, belief functions, possibilistic logic, *etc.* and we will refer to some of them in this part. Things become more complex when we introduce events and changes in the picture. Issues arise even if we assume that there is a single agent. Indeed, there are situations where the agent might also have some uncertainty about the incoming information: for example, agent A might be uncertain due to some noise whether her colleague told her that he would visit her on Tuesday or on Thursday. In this part we also investigate such

phenomena.

Things are even more complex in a multi-agent setting because the way agents update their beliefs depends not only on their beliefs about the event itself but also on their beliefs about the way the other agents perceived the event (and so about the other agents' beliefs about the event). For example, during a private announcement of a piece of information to agent A, the beliefs of the other agents actually do not change because they believe nothing is actually happening; but during a public announcement all the agents' beliefs might change because they all believe that an announcement has been made. Such kind of subtleties have been dealt with in a field called *dynamic epistemic logic* (Baltag et al., 1998; van Ditmarsch et al., 2007; van Benthem, 2011). The idea is to represent by an event model how the event is perceived by the agents and then to define a formal update mechanism that specifies how the agents update their beliefs according to this event model and their previous representation of the situation.

So this part is more generally about information and information change. However, we will not deal with problems of how to store information in machines or how to actually communicate information. Such problems have been dealt with in information theory (Cover and Thomas, 1991) and Kolmogorov complexity theory (Li and Vitányi, 1993). We will just assume that such mechanisms are already available and start our investigations from there.

This part is divided into two chapters. In Chapter 3, we will introduce the main logical formalisms for reasoning about uncertainty in a context where there is a *single agent*. In Chapter 4, we will introduce the current most prominent logical formalisms for reasoning about uncertainty in a context where there are *multi-agent* systems, namely Dynamic Epistemic Logic (DEL).

Applicative perspective. Studying and proposing logical models for reasoning about uncertainty has applications in three areas. First, in artificial intelligence, where machines or robots need to have a formal representation of the surrounding world (which might involve other agents) and formal mechanisms to update this representation when they receive incoming information. Such formalisms are crucial if we want to design agents, able to act autonomously in the real world or in a virtual world (such as on the internet). Indeed, the representation of the surrounding world is essential for a robot to reason about the world, plan actions in order to achieve goals, *etc.* It must also be able to update and revise its representation of the world itself, in order to cope autonomously with unexpected events. Second, in game theory (and consequently in economics), where we need to model games involving several agents (players) having beliefs about the game and about the other agents' beliefs (such as agent A believes that agent B has the ace of spade, or agent A believes that agent B believes that agent A has the ace of hearts...), and how they update their representation of the game when events (such as showing a card privately or putting a card on the table) occur. Third, in cognitive psychology, where we need to model as accurately as possible the epistemic state of human agents and the dynamics of belief and knowledge, in order to explain and describe cognitive processes.

Chapter 3

Reasoning Alone about Uncertainty

"Do not expect to arrive at certainty in every subject which you pursue. There are a hundred things wherein we mortals... must be content with probability, where our best light and reasoning will reach no farther"

– Isaac Watts (1674 – 1748)

3.1 Introduction

As we shall see in this chapter, reasoning about uncertainty can be subtle and varied. This richness may be explained by the fact that uncertainty is present in almost every human activity. For example, a player may have no clue whether the dice of a backgammon are loaded or not, but this should not prevent him from making inferences and strategies about which moves to play. Likewise, an engineer may have difficulties assigning numbers corresponding to probabilities of occurrence of faulty events of a system that he is supposed to manage, and he may then prefer instead to compare the relative likelihood of these events. In general, depending on the access that the modeler has about the epistemic state of the agent or about the situation at stake, the representation of uncertainty will be different. In fact, the variety of representations of uncertainty reflects the variety of situations in everyday life in which uncertainty has to be represented by some means. This explains why so many formalisms have been introduced to account for the representation and reasoning about uncertainty. In this chapter, we will present some of these most well-known formalisms and we will see that they are in fact all instances of a very abstract framework based on the notion of plausibility measure introduced by Friedman and Halpern (2001).

The chapter is organized as follows. We will first focus in Section 3.2 on the *representation* of uncertainty by studying several formalisms introduced for different reasons. Then, we will introduce in Section 3.3 various logics for *reasoning* about the uncertainty represented in these formalisms. Finally, in Section 3.4, we will show how the representation of uncertainty can be updated as events occur and how these events change our representation of the world and our uncertainty about it. (However, we will not define logics for reasoning about these dynamic phenomena.)

3.2 Representing Uncertainty

In this section, we introduce various formalisms for representing uncertainty: probability spaces, lower/upper probabilities, inner/outer measures (Section 3.2.1), belief functions (Section 3.2.2), possibility measures (Section 3.2.3) and ranking function (Section 3.2.4). Each of them is suitable for representing uncertainty in a certain kind of situation and under specific modeling assumptions. Choosing the right formalism suitable for modeling a given situation will be the topic of Section 3.2.7.

In this chapter, W is a *finite* set of possible worlds. However, all the results of this chapter can be extended to a setting where W is an *infinite* set of possible worlds (modulo some slight changes, see (Halpern, 2003)).

3.2.1 Probability Measures

This section deals with the problem of representing uncertainty with probability. We will start by formally defining probability spaces. If we follow a perfect external approach, representing the uncertainty of the agent by means of a single probability measure (space) is often sufficient. Justifying where the probability numbers come from in that case is the topic of Section 3.2.6. On the other hand, if we follow an *imperfect* external approach (see Section 4.1 for more details on the different modeling approaches), then we need to represent the uncertainty of the agent by means of a *set* of probability measures. This second case will be dealt with by means of lower/upper probabilities or inner/outer measures in Section 3.2.1.1.

If a probability can be assigned to both sets U and V, then it is useful to be able to assume that a probability can also be assigned to $U \cup V$ and to $\overline{U}$. This leads us to define the notion of *algebra*.

Definition 3.2.1 (Algebra). An *algebra over* W is a set $\mathcal{F}$ of subsets of W that contains W and is closed under union and complementation, so that if U and V are in $\mathcal{F}$, then so are $U \cup V$ and $\overline{U}$. □

For example, $\mathcal{F} := 2^W$ is an algebra over W. Note that an algebra is also closed under intersection, since $U \cap V = \overline{\overline{U} \cup \overline{V}}$.

Definition 3.2.2 (Probability space). A *probability space* is a tuple $(W, \mathcal{F}, \mu)$, where $\mathcal{F}$ is an algebra over W and $\mu : \mathcal{F} \to [0; 1]$ satisfies the following two properties:

$$\mu(W) = 1 \tag{P1}$$

$$\mu(U \cup V) = \mu(U) + \mu(V) \qquad \text{if } U \text{ and } V \text{ are disjoint elements of } \mathcal{F}. \tag{P2}$$

Although (P2) applies only to pairs of sets, an easy induction argument shows that if $U_1, \ldots, U_k$ are pairwise disjoint elements of $\mathcal{F}$, then

$$\mu(U_1 \cup \ldots \cup U_k) = \mu(U_1) + \ldots + \mu(U_k) \tag{Finite Additivity}$$

3.2.1.1 Lower and Upper Probabilities, Inner and Outer Measures

Example 3.2.1. Suppose that a bag contains 100 marbles; 30 are known to be red, and the remainder are known to be either blue or violet, although the exact proportion of blue and violet is not known. What is the likelihood that a marble taken out of the bag is violet ?

This example can be modeled with three possible worlds: w_1 (for the *red* outcome), w_2 (for the *blue* outcome) and w_3 (for the *violet* outcome). It seems reasonable to assign probability 0.3 to w_1 and probability 0.7 to $\{w_2, w_3\}$:

$$\mu(\{w_1\}) := 0.3 \qquad\qquad \mu(\{w_2, w_3\}) := 0.7.$$

But what probability should be assigned to $\{w_2\}$ and $\{w_3\}$? If we apply the *principle of indifference*, then w_2 and w_3 should both be assigned the probability $0.35 = \frac{0.7}{2}$. This suggests that betting that a blue or a violet marble will be withdrawn is more likely than betting for a red marble. This is obviously counter-intuitive. □

To summarize the problem of the above example, we have an algebra $\mathcal{F} := \{\emptyset, \{w_1\}, \{w_2, w_3\}, \{w_1, w_2, w_3\}\}$ and a probability measure μ defined on this algebra. We want to extend this probability measure to the algebra $\mathcal{F}' := 2^{\{w_1, w_2, w_3\}}$.

We formalize the problem raised by the example. Let μ be a probability measure on $\mathcal{F}$. We want to extend this probability measure to an algebra $\mathcal{F}'$ such that $\mathcal{F} \subseteq \mathcal{F}'$. We define two *approximations* of such an extension (from above and from below) in two different ways.

Definition 3.2.3 (Lower/upper probability; Inner/outer measure). Let $\mathcal{F}, \mathcal{F}'$ be two algebras such that $\mathcal{F} \subseteq \mathcal{F}'$. Let μ be a probability measure on $\mathcal{F}$.

1. We define the *lower probability* and the *upper probability induced by μ on $\mathcal{F}'$* as follows. For that, we consider the *set of all extensions of μ to $\mathcal{F}'$* defined by $\mathcal{P}_\mu := \{\mu' : \mu' \text{ is a probability measure on } \mathcal{F}' \text{ and } \mu'(U) = \mu(U) \text{ for all } U \in \mathcal{F}\}$. Then, for all $U \in \mathcal{F}'$, we define

$$(\mathcal{P}_\mu)_*(U) := \inf\{\mu(U) : \mu \in \mathcal{P}\} \qquad \text{(Lower probability)}$$
$$(\mathcal{P}_\mu)^*(U) := \sup\{\mu(U) : \mu \in \mathcal{P}\}. \qquad \text{(Upper probability)}$$

2. We define the *inner measure* and the *outer measure induced by μ on $\mathcal{F}'$* as follows: for all $U \in \mathcal{F}'$,

$$\mu_*(U) := \sup\{\mu(V) : V \subseteq U, V \in \mathcal{F}\} \qquad \text{(Inner measure)}$$
$$\mu^*(U) := \inf\{\mu(V) : V \supseteq U, V \in \mathcal{F}\}. \qquad \text{(Outer measure)}$$

□

These two ways to define upper and lower bounds that approximate extensions are in fact equivalent:

Theorem 3.2.1. *Let μ be a probability measure on an algebra $\mathcal{F}$ and let $\mathcal{P}_\mu$ consist of all extensions of μ to an algebra $\mathcal{F}' \supset \mathcal{F}$. Then, $\mu_*(U) = (\mathcal{P}_\mu)_*(U)$ and $\mu^*(U) = (\mathcal{P}_\mu)^*(U)$ for all $U \in \mathcal{F}'$.*

Example 3.2.2. We have that $(\mathcal{P}_\mu)_*(\{w_2\}) = \mu_*(\{w_2\}) = 0$ and $(\mathcal{P}_\mu)^*(\{w_2\}) = \mu^*(\{w_2\}) = 0.7$. Likewise for $\{w_3\}$. This tells us that the probability to withdraw a blue marble (or a violet marble) is between 0 and 0.7. Also, we have that $(\mathcal{P}_\mu)_*(\{w_1\}) = \mu_*(\{w_1\}) = (\mathcal{P}_\mu)^*(\{w_1\}) = \mu^*(\{w_1\}) = 0.3$: the probability to withdraw a red marble is equal to 0.3. □

Sets of probabilities can be characterized by an argument similar to the Dutch book argument of the previous section. In that case, the postulate (RAT3) no longer holds. Note that μ_* and μ^* are not necessarily probability measures. They satisfy some weaker properties:

Proposition 3.2.1. *For all inner measure μ_* and outer measure μ^* induced by μ on $\mathcal{F}'$, for all $U, V \in \mathcal{F}'$, we have*

$$\mu_*(U \cup V) \geq \mu_*(U) + \mu_*(V) \qquad \text{(Super-additivity)}$$
$$\mu^*(U \cup V) \leq \mu^*(U) + \mu^*(V) \qquad \text{(Sub-additivity)}$$
$$\mu_*(U) = 1 - \mu^*(\overline{U}) \qquad \text{(Dual)}$$

Lower and upper probabilities satisfy also (Super-additivity), (Sub-additivity) *and* (Dual).

3.2.2 Dempster-Shafer Belief Functions

Just like an inner measure, $Bel(U)$ can be viewed as providing a lower bound on the likelihood of U.

Definition 3.2.4 (Belief function and plausibility function). A *belief function* $Bel : 2^W \to [0; 1]$ is a function satisfying the following three properties:

$$Bel(\emptyset) = 0 \qquad \text{(B1)}$$
$$Bel(W) = 1 \qquad \text{(B2)}$$
$$Bel\left(\bigcup_{i=1}^{n} U_i\right) \geq \sum_{i=1}^{n} \sum_{I \subseteq \{1,\ldots,n\}:|I|=i} (-1)^{i+1} Bel\left(\bigcap_{j \in I} U_j\right) \qquad \text{for all } n \in \mathbb{N}. \qquad \text{(B3)}$$

The *plausibility function* associated to Bel is defined by, for all $U \in 2^W$,

$$Plaus(U) := 1 - Bel\left(\overline{U}\right) \qquad \text{(Dual)}$$

We recall that $\bigcup_{i=1}^{n} U_i := U_1 \cup \ldots \cup U_n$, that $\bigcap_{i=1}^{n} U_i := U_1 \cap \ldots \cap U_n$ (see Chapter A) and that for any finite set of real numbers $\{x_1, \ldots, x_n\} \subseteq [0;1]$ we have $\sum_{i=1}^{n} x_i := x_1 + \ldots + x_n$.

Note that every probability measure and every inner measure is a belief function. But not every belief function is the inner measure of a probability measure.

Proposition 3.2.2. *Let Bel be a belief function and Plaus its associated plausibility function. Then, for all $U \in 2^W$, we have*

$$Bel(U) \leq Plaus(U)$$

$$Plaus\left(\bigcap_{i=1}^{n} U_i\right) \geq \sum_{i=1}^{n} \sum_{\{I \subseteq \{1,\ldots,n\} : |I| = i\}} (-1)^{i+1} Bel\left(\bigcup_{j \in I} U_j\right) \text{ for all } n \in \mathbb{N}. \tag{3.1}$$

In fact, plausibility measures are characterized by the properties $Plaus(\emptyset) = 0, Plaus(W) = 1$ and (3.1).

For any event U, the interval $[Bel(U); Plaus(U)]$ can be viewed as describing the range of possible values of the likelihood of U. The connection between belief functions, inner measures and lower probabilities is made more precise by the following theorem:

Theorem 3.2.2. *Given a belief function Bel defined on a space W, let $\mathcal{P}_{Bel} := \{\mu : \mu$ is a probability measure on 2^W and $\mu(U) \geq Bel(U)$ for all $U \subseteq W\}$. Then, $Bel = (\mathcal{P}_{Bel})_*$ and $Plaus = (\mathcal{P}_{Bel})^*$.*

Belief functions are part of a theory of *evidence*. Intuitively, evidence supports events to varying degrees. In general, evidence provides some degree of support (possibly 0) for each subset of W. The total amount of support is 1.

Definition 3.2.5 (Mass function). A *mass function* (sometimes called a *basic probability assignment*) on W is a function $m : 2^W \to [0;1]$ satisfying the following properties:

$$m(\emptyset) = 0 \tag{M1}$$

$$\sum_{U \subseteq W} m(U) = 1. \tag{M2}$$

Example 3.2.3. The information that there are exactly 30 red marbles provides support in degree 0.3 for w_1; the information that there are 70 violet and blue marbles does not provide any positive support for either $\{w_2\}$ or $\{w_3\}$, but does provide support 0.7 for $\{w_2, w_3\}$. So, we have $m(\{w_1\}) = 0.3$, $m(\{w_2\}) = m(\{w_3\}) = m(\{w_1, w_2, w_3\}) = 0$, and $m(\{w_2, w_3\}) = 0.7$. □

The belief that U holds, $Bel(U)$, is then the sum of all of the support on subsets of U. Intuitively, $Bel_m(U)$ is the sum of the probabilities of the evidence or observations that guarantee that the actual world is in U.

Definition 3.2.6 (Belief function based on a mass function). Given a mass function m, define the *belief function based on* m, Bel_m, by taking:

$$Bel_m(U) = \sum_{\{U' : U' \subseteq U\}} m(U')$$

The corresponding plausibility function $Plaus_m$ is defined as:

$$Plaus_m(U) = \sum_{\{U' : U' \cap U \neq \emptyset\}} m(U') \qquad \square$$

There is a one-to-one correspondence between belief functions and mass functions:

Theorem 3.2.3. *Given a mass function m on W, the function Bel_m is a belief function and $Plaus_m$ is the corresponding plausibility function. Moreover, given a belief function Bel on W, there is a unique mass function m on W such that $Bel = Bel_m$.*

Hence, there are three equivalent kinds of representation of uncertainty: mass functions m, associated belief/plausibility functions $(Bel_m, Plaus_m)$ and sets of probabilities $\mathcal{P}_{Bel_m}$. Note that if the set of probabilities $\mathcal{P}_{Bel} = \{\mu\}$ is a singleton, then $Bel_m = Plaus_m = \mu$.

3.2.3 Possibility Measures

Possibility measures are yet another approach to assigning numbers to sets. They are based on ideas of *fuzzy logic*.

Definition 3.2.7 (Possibility measure). A *possibility measure* $Poss : 2^W \to [0; 1]$ is a function satisfying the following three properties:

$$Poss(\emptyset) = 0 \qquad \text{(Poss1)}$$

$$Poss(W) = 1 \qquad \text{(Poss2)}$$

$$Poss(U \cup V) = \max\{Poss(U), Poss(V)\} \qquad \text{if } U \text{ and } V \text{ are disjoint.} \qquad \text{(Poss3)}$$

Unlike (P2), (Poss3) holds even if U and V are not disjoint. Like probability, if W is finite, then a possibility measure can be defined by its behavior on singleton sets: $Poss(U) = \max_{u \in U}(\{u\})$.

Proposition 3.2.3. *A possibility measure is a plausibility function (it satisfies* (3.1)*). The dual of possibility, called* necessity*, is defined as follows:*

$$Nec(U) := 1 - Poss(\overline{U}) \qquad \text{(Dual)}$$

Since Poss is a plausibility function, Nec is a belief function and $Nec(U) \leq Poss(U)$ for all $U \in 2^W$.

Since possibility measures are specific kind of plausibility measures, we could wonder to which kind of mass function it corresponds.

Definition 3.2.8 (Consonant mass function). A mass function m is *consonant* when for all $U, U' \in 2^W$, $m(U) > 0$ and $m(U') > 0$ implies that either $U \subseteq U'$ or $U' \subseteq U$. □

The following theorem shows that possibility measures are the plausibility functions that correspond to *consonant* mass functions:

Theorem 3.2.4. *If m is a consonant mass function on a finite space W, then $Plaus_m$, the plausibility function corresponding to m, is a possibility measure. Conversely, given a possibility measure $Poss$ on W, there is a consonant mass function m such that $Poss$ is the plausibility function corresponding to m.*

3.2.4 Ranking Function

Ranking functions are similar in spirit to possibility measures.

Definition 3.2.9 ((Ordinal) Ranking function). A *ranking function* $\kappa : 2^W \to \mathbb{N}^*$, where $\mathbb{N}^* := \mathbb{N} \cup \{\infty\}$, is a function satifying the following three properties:

$$\kappa(\emptyset) = \infty \tag{Rk1}$$
$$\kappa(W) = 0 \tag{Rk2}$$
$$\kappa(U \cup V) = \min\{\kappa(U), \kappa(V)\} \qquad \text{if } U \text{ and } V \text{ are disjoint} \tag{Rk3}$$

The numbers can be though as denoting degrees of surprise; that is, $\kappa(U)$ is the degree of surprise the agent would feel if the actual world were in U. 0 denotes "unsurprising", 1 denotes "somewhat surprising", 2 denotes "quite surprising", and so on; ∞ denotes "so surprising as to be impossible". For example, the uncertainty corresponding to tossing a coin with bias $\frac{1}{3}$ can be captured by a ranking function such as $\kappa(heads) = \kappa(tails) = 0$ and $\kappa(edge) = 3$, where *edge* is the event that the coin lands edge.

Like possibility measures, the third property (Rk3) holds even if U and V are not disjoint. As with probability and possibility, a ranking function is characterized by its behavior on singletons in finite spaces: $\kappa(U) = \min_{u \in U} \kappa(u)$. To ensure that (Rk2) holds, it must be the case that $\min_{w \in W} \kappa(w) = 0$.

Ranking functions can be viewed as possibility measures in a straightforward way. Given a ranking function κ, define the possibility measure $Poss_\kappa$ as follows: for all $U \in 2^W$,

$$Poss_\kappa(U) = \begin{cases} \dfrac{1}{1 + \kappa(U)} & \text{if } \kappa(U) \neq \infty \\ 0 & \text{if } \kappa(U) = \infty \end{cases} \tag{3.2}$$

Proposition 3.2.4. *Let κ be a ranking function. The function $Poss_\kappa$ as defined by expression* (3.2) *is a possibility measure.*

3.2.5 Preferential Structures

Sometimes, we do not want or cannot assign numbers concerning the likelihood of events/facts. In that case, we want or can only compare the relative likelihood of different events/facts. For that purpose, preferential structures are adequate formalisms for representing uncertainty.

Definition 3.2.10 (Preferential structure). A *preferential structure* is a tuple $(W, \mathcal{O})$ where

- W is a non-empty set;
- $\mathcal{O}$ is a function assigning to each $w \in W$ a pair $(W_w, \succeq_w)$ where $W_w \subseteq W$ and $\succeq_w$ is a partial preorder on W_w (*i.e.* a reflexive and transitive relation on W_w). □

Intuitively, we have $u \succeq_w v$ when, from the point of view of w, u is at least as likely as v. Given this interpretation, the fact that $\succeq_w$ is assumed to be a partial preorder is easy to justify. Transitivity just says that if u is at least as likely as v, and v is at least as likely as w, then u is at least as likely as w; reflexivity just says that world w is at least as likely as itself. The fact that $\succeq_w$ is partial allows for agents who are not able to compare two worlds in likelihood. We extend the definition of $\succeq_w$ to *sets* in two different ways as follows:

$$\begin{array}{lll} U \succeq^s_w V & \text{iff} & \text{for all } v \in V - U, \text{ there is } u \in U \text{ such that } u \succ v \\ & & \text{and it is not the case that } x \succ u \text{ for any } x \in U \\ U \succeq^e_w V & \text{iff} & \text{for all } v \in V, \text{ there is some } u \in U \text{ such that } u \succeq_w v \end{array}$$

Finally, if $\succeq$ is any partial order, we write $u \succ v$ as a shorthand for $u \succeq v$ and $v \not\succeq u$.

3.2.6 Justifying Probability Numbers

If belief is represented in terms of probabilities, then it is important to explain what the numbers represent, where they come from, and why the property of Finite Additivity is appropriate.

Objective interpretation: Principle of Indifference and relative-frequency. The classical approach to applying probability, which goes back to the 17^{th} century and 18^{th} centuries, is to reduce a situation to a number of elementary outcomes. Then, we apply the following principle:

Principle of Indifference: all elementary outcomes are equally likely.

Applying the principle of indifference, if there are n elementary outcomes, the probability of each one is $\frac{1}{n}$. Clearly, this definition satisfies *P1* and *P2*. However, How do we determine the elementary outcomes ? If a coin is biased, what are the equally likely

outcomes ? Another interpretation of the probability numbers is that they represent relative frequencies. The probability that a coin has *bias* 0.6 (where the bias of a coin is the probability that it lands heads) is that it lands heads roughly 60 % of the time when it is tossed sufficiently often. In that case, probability is an *objective* property of a situation.

Subjective interpretation: Dutch book. Another interpretation suggests that the numbers reflect subjective assessments of likelihood: this is the *subjective viewpoint*. Consider the following bet:

$$(U,\alpha) := \begin{cases} \text{if } U \text{ happens then the agent wins } 100(1-\alpha) \text{ euros,} \\ \text{if } \overline{U} \text{ happens then the agent loses } 100\alpha \text{ euros.} \end{cases}$$

The bet $(\overline{U}, 1-\alpha)$ is called the *complementary* bet to (U,α):

$$(\overline{U},\alpha) := \begin{cases} \text{if } \overline{U} \text{ happens then the agent wins } 100\alpha \text{ euros,} \\ \text{if } U \text{ happens then the agent loses } 100(1-\alpha) \text{ euros.} \end{cases}$$

Whether the bet (U,α) is *at least as good as* the bet $(\overline{U}, 1-\alpha)$, written $(U,\alpha) \succeq (\overline{U}, 1-\alpha)$, clearly depends on α. Assume that the agent has a preference order $\succeq$ between sets of bets of the form $\{(U_1,\alpha_1),\ldots,(U_k,\alpha_k)\}$ and $\{(\overline{U}_1, 1-\alpha_1),\ldots,(\overline{U}_k, 1-\alpha_k)\}$. Let us consider the following rationality postulates:

If B_1 is guaranteed to give at least as much money as B_2, then $B_1 \succeq B_2$,
if B_1 is guaranteed to give more money than B_2, then $B_1 \succ B_2$ (RAT1)

If $B_1 \succ B_2$ and $B_2 \succ B_3$, then $B_1 \succ B_3$ (RAT2)

Either $(U,\alpha) \succeq (\overline{U}, 1-\alpha)$ or $(\overline{U}, 1-\alpha) \succeq (U,\alpha)$ (RAT3)

If $(U_i,\alpha_i) \succeq (V_i,\beta_i)$ for all $i = 1,\ldots,k$,
then $\{(U_1,\alpha_1),\ldots,(U_k,\alpha_k)\} \succeq \{(V_1,\beta_1),\ldots,(V_k,\beta_k)\}$. (RAT4)

In (RAT1), "guaranteed to give at least as much money" means that no matter what happens, the agent does at least as well with B_1 as with B_2. For example, if $B1 = (U,\alpha)$ and $B_2 = (V,\beta)$, then this means that $\alpha \leq \beta$.

Theorem 3.2.5. *If an agent satisfies* (RAT1)–(RAT4)*, then for each subset U of W, a number α_U exists such that $(U,\alpha) \succeq (\overline{U}, 1-\alpha)$ for all $\alpha < \alpha_U$ and $(\overline{U}, 1-\alpha) \succeq (U,\alpha)$ for all $\alpha > \alpha_U$. Moreover, the function defined by $\mu(U) = \alpha_U$ is a probability measure.*

So, if the agent is certain that U is *not* the case, then α_U should be 0, and if the agent is certain that U is the case, then α_U should be 1. That is, if the agent is certain that U, then for any $\alpha > 0$, it should be the case that $(U,\alpha) \succeq (\overline{U}, 1-\alpha)$, since she feels that with (U,α) she is guaranteed to win $100(1-\alpha)$ euros, while with $(\overline{U}, 1-\alpha)$ she is guaranteed to lose the same amount.

Theorem 3.2.5 entails that if U_1 and U_2 are disjoint subsets of W and we have that $\alpha_{U_1 \cup U_2} \neq \alpha_{U_1} + \alpha_{U_2}$, then there is a set B_1 of bets such that the agent prefers B_1 to the complementary set B_2, yet the agent is guaranteed to lose money with B_1 and guaranteed to win money with B_2 (thus contradicting (RAT1)). Such a collection of bets B_1 is called in the literature a *Dutch book* (in the sense of "bookmaker").

3.2.7 Choosing a Representation

Choosing the right formalism to represent uncertainty depends on the real-world situation we model and on the modeling assumptions that we follow (see Section 4.1 for more details on the various modeling assumptions).

- Probability has the advantage of being well understood. It is a powerful tool; many technical results have been proved that facilitate its use, and a number of arguments suggest that, under certain assumptions, probability is the only "rational" way to represent uncertainty. Choosing a single probability space for modeling situations is more suitable when the modeler follows a *perfect external approach*.

- Belief functions and sets of probability measures have many of the advantages of probability but may be more appropriate in a setting where there is uncertainty about the likelihood. Choosing a set of probability measures or belief functions for modeling situations is more suitable when the modeler follows an *imperfect external approach*.

- Partial preorders on possible worlds may be also more appropriate in setting where no quantitative information is available.

3.3 Reasoning about Uncertainty

All the representations of uncertainty that we have considered in the previous section are defined by classes of models. They can also be viewed as the semantics of specific logics (to be defined). So, on top of these classes of models, we can define logical language in order to *reason* about the uncertainty that these model represent. This overall procedure corresponds to the third approach for defining a logic that we have identified in Section 1.1 and we are going to follow this approach in this section in order to define logics for reasoning about uncertainty.

3.3.1 Plausibility Measures: an Abstract Framework

First, we define more precisely the classes of models/structures that will constitute the semantics of various logics for reasoning about uncertainty that we are going to define. We are going to show that they are all in fact specific instances of the general framework based on plausibility measures introduced by Friedman and Halpern (2001).

Definition 3.3.1 (Probability, lower probability, belief, ranking, possibility, preferential structures). A *pointed probability* (resp. *lower probability*, *belief*, *ranking*, *possibility*, *preferential*) *structure* is a tuple $S = (W, \mathcal{X}, \pi, w)$ where

- W is a non-empty set and $w \in W$;
- $\mathcal{X}(w) := (W_w, \mathcal{F}_w, \mu_w)$ for each $w \in W$ is a tuple such that
  - $W_w \subseteq W$ is a non-empty set;
  - $\mathcal{F}_w$ is an algebra over W_w;
  - μ_w is a probability measure (resp. a lower probability measure, a belief function, a ranking function, a possibility measure, a partial preorder) on $\mathcal{F}_w$;
- $\pi : W \to \mathcal{C}_{\mathsf{PL}}$ is a function called the *valuation function*.

A *pointed probability* (resp. *lower probability*, *belief*, *ranking*, *possibility*, *preferential*) *structure* is *measurable* when for all $w \in W$, $\mathcal{F}_w = 2^{W_w}$. We denote by $\mathcal{S}^{prob}$ (resp. $\mathcal{S}^{bel}$, $\mathcal{S}^{low}$, $\mathcal{S}^{rank}$, $\mathcal{S}^{poss}$ and $\mathcal{S}^{pref}$) the class of all pointed measurable probability (resp. lower probability, belief, ranking, possibility, preferential) structures. □

Plausibility measures. The basic idea under plausibility measure is straightforward. A probability measure maps sets in an algebra $\mathcal{F}$ over a set W of worlds to $[0; 1]$. A *plausibility measure* is more general: it maps sets in $\mathcal{F}$ to some arbitrary partially ordered set.

In the rest of the lecture notes, D is a non-empty set *partially ordered* by a relation $\leq$ (so that $\leq$ is reflexive, transitive and anti-symmetric). We further assume that D contains two special elements $\top$ and $\bot$ such that for all $d \in D$, $\bot \leq d \leq \top$. As usual, we define the ordering $<$ by taking $d_1 < d_2$ if and only if $d_1 \leq d_2$ and $d_1 \neq d_2$.

Definition 3.3.2 (Plausibility measure and structure). A *(qualitative) plausibility measure* is a function $Pl : \mathcal{F} \to D$ satisfying (Pl1)–(Pl3) (resp. (Pl1)–(Pl5)):

$$Pl(\emptyset) = \bot \tag{Pl1}$$

$$Pl(W) = \top \tag{Pl2}$$

$$\text{If } U \subseteq V, \text{ then } Pl(U) \leq Pl(V) \tag{Pl3}$$

$$\begin{array}{l}\text{If } A, B \text{ and } C \text{ are pairwise disjoint sets,}\\ Pl(A \cup B) > Pl(C) \text{ and } Pl(A \cup C) > Pl(B) \text{ imply } Pl(A) > Pl(B \cup C)\end{array} \tag{Pl4}$$

$$\text{If } Pl(A) = Pl(B) = \bot, \text{ then } Pl(A \cup B) = \bot. \tag{Pl5}$$

A *pointed qualitative plausibility structure* is a tuple $S = (W, \mathcal{X}, \pi, w)$ where

- W is a non-empty set and $w \in W$;
- $\mathcal{X}(w) := (W_w, \mathcal{F}_w, Pl_w)$ for each $w \in W$ is a tuple such that
  - $W_w \subseteq W$ is a non-empty set;

- $\mathcal{F}_w$ is an algebra over W_w;
- Pl_w is a qualitative plausibility measure on $\mathcal{F}_w$;

- $\pi : W \to \mathcal{C}_{\mathsf{PL}}$ is a function called the *valuation function.*

A *pointed qualitative plausibility structure* is *measurable* when for all $w \in W$, $\mathcal{F}_w = 2^{W_w}$. We denote by $\mathcal{S}^{qual}$ the class of all pointed measurable qualitative plausibility structures. A *simple qualitative plausibility structure* is a tuple $S = (W, Pl, \pi)$ where Pl is a qualitative plausibility measure on the algebra 2^W over W. □

Proposition 3.3.1 below shows that plausibility measures are abstract enough to embed the various formalisms for representing uncertainty that we have encountered in these lecture notes.

Proposition 3.3.1. *Probability measures, lower probability measures, belief functions, possibility measures, ranking functions and preferential relations are qualitative plausibility measures.*

Proof. For all except ranking functions, $D = [0; 1]$, $\bot = 0$, $\top = 1$, and $\leq_D$ is the standard ordering on the reals. For ranking functions, $D = \mathbb{N}^*$, $\bot = \infty$, $\top = 0$, and the ordering $\leq_{\mathbb{N}^*}$ is the opposite of the standard ordering on $\mathbb{N}^*$. □

3.3.2 Logics for Quantitative Reasoning

We consider a generic logical language for reasoning *quantitatively* about uncertainty. It contains likelihood terms of the form $\ell(\varphi) > b$ which reads as 'the agents assigns a likelihood greater than b to φ' or of the form $2\ell(\varphi) + \ell(\psi) > 0.5$ which reads as 'the sum of twice the likelihood of φ with the likelihood of ψ is greater than 0.5'.

Definition 3.3.3 (Quantitative language $\mathcal{L}_{\mathsf{Quant}}$). The language $\mathcal{L}_{\mathsf{Quant}}$ is defined inductively by the following grammar in BNF:

$$\mathcal{L}_{\mathsf{Quant}} : \quad \varphi \quad ::= \quad p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid (a_1\ell(\varphi) + \ldots + a_k\ell(\varphi) > b)$$

where $p \in PROP$ and $a_1, \ldots, a_k, b \in \mathbb{R}$. We use the following abbreviations:

$$\ell(\varphi) - \ell(\psi) > b := \ell(\varphi) + (-1)\ell(\psi) > b \qquad \ell(\varphi) > \ell(\psi) := \ell(\varphi) - \ell(\psi) > 0$$
$$\ell(\varphi) < \ell(\psi) := \ell(\psi) - \ell(\varphi) > 0 \qquad \ell(\varphi) \leq b := \neg(\ell(\varphi) > b)$$
$$\ell(\varphi) \geq b := -\ell(\varphi) \leq -b \qquad \ell(\varphi) = b := (\ell(\varphi \geq b) \wedge (\ell(\varphi) \leq b) \qquad \square$$

This language $\mathcal{L}_{\mathsf{Quant}}$ is very generic and can be given various semantics, depending on the class of (quantitative) models that we consider: probabilistic, sets of probabilities, possibilistic, belief functions, *etc.*

Definition 3.3.4 (Satisfaction relations of $\mathcal{L}_{\mathsf{Quant}}$ for $\mathcal{S}^{prob}$, $\mathcal{S}^{low}$, $\mathcal{S}^{bel}$ and $\mathcal{S}^{poss}$). The *satisfaction relation* $\models \subseteq \mathcal{S}^{prob} \times \mathcal{L}_{\mathsf{Quant}}$ is defined inductively as follows. Let $(S, w) \in \mathcal{S}^{prob}$ and $\varphi, \varphi_1, \ldots, \varphi_k \in \mathcal{L}_{\mathsf{Quant}}$.

$$\begin{array}{lll} S, w \models p & \text{iff} & \pi(w)(p) = T \\ S, w \models \neg\varphi & \text{iff} & \text{it is not the case that } S, w \models \varphi \\ S, w \models \varphi \wedge \psi & \text{iff} & S, w \models \varphi \text{ and } S, w \models \psi \\ S, w \models a_1\ell(\varphi_1) + \ldots + a_k\ell(\varphi_l) > b & \text{iff} & a_1\mu_w([\![\varphi_1]\!]) + \ldots + a_k\mu_w([\![\varphi_k]\!]) > b \end{array}$$

where $[\![\varphi]\!] := \{w \in W_w : S, w \models \varphi\}$. The satisfaction relations for $\mathcal{S}^{low}$, $\mathcal{S}^{rank}$, $\mathcal{S}^{poss}$ and $\mathcal{S}^{bel}$ are defined identically. □

The set of validities of the logics $(\mathcal{L}_{\mathsf{Quant}}, \mathcal{S}^{prob}, \models)$, $(\mathcal{L}_{\mathsf{Quant}}, \mathcal{S}^{low}, \models)$, $(\mathcal{L}_{\mathsf{Quant}}, \mathcal{S}^{rank}, \models)$, $(\mathcal{L}_{\mathsf{Quant}}, \mathcal{S}^{poss}, \models)$ and $(\mathcal{L}_{\mathsf{Quant}}, \mathcal{S}^{bel}, \models)$ can be axiomatized and the respective proof systems are given in (Halpern, 2003, Chap 7). These axiomatizations can then be used to *reason* about uncertainty in *quantitative* terms.

3.3.3 Logics for Qualitative Reasoning

For qualitative reasoning, we cannot resort to numbers anymore. Hence, we define a qualitative logical language $\mathcal{L}_{\mathsf{Qual}}$ which allows for the expression of properties about the relative likelihood of statements (expressed via formulas of this logical language).

Definition 3.3.5 (Qualitative language $\mathcal{L}_{\mathsf{Qual}}$). The language $\mathcal{L}_{\mathsf{Qual}}$ is defined inductively by the following grammar in BNF:

$$\mathcal{L}_{\mathsf{Qual}} : \quad \varphi \quad ::= \quad p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid (\ell(\varphi) \geq \ell(\varphi))$$

We use the same abbreviations as in Definition 3.3.3 (when they exist). □

This language $\mathcal{L}_{\mathsf{Qual}}$ is very generic and can be given various semantics, depending on the class of models that we consider: preferential, ranking, plausibility, possibility, *etc.*

Definition 3.3.6 (Satisfaction relation of $\mathcal{L}_{\mathsf{Qual}}$ for $\mathcal{S}^{pref}$, $\mathcal{S}^{qual}$, $\mathcal{S}^{rank}$ and $\mathcal{S}^{poss}$). The *satisfaction relation* $\models \subseteq \mathcal{S}^{pref} \times \mathcal{L}_{\mathsf{Qual}}$ is defined inductively as follows. Let $(S, w) \in \mathcal{S}^{pref}$ and $\varphi, \psi \in \mathcal{L}_{\mathsf{Qual}}$.

$$\begin{array}{lll} S, w \models p & \text{iff} & \pi(w)(p) = T \\ S, w \models \neg\varphi & \text{iff} & \text{it is not the case that } S, w \models \varphi \\ S, w \models \varphi \wedge \psi & \text{iff} & S, w \models \varphi \text{ and } S, w \models \psi \\ S, w \models \ell(\varphi) \geq \ell(\psi) & \text{iff} & [\![\varphi]\!] \succeq^e_w [\![\psi]\!] \end{array}$$

where $[\![\varphi]\!] := \{w \in W_w : S, w \models \varphi\}$. The last truth condition for $(S, w) \in \mathcal{S}^{qual}$, $(S, w) \in \mathcal{S}^{rank}$ and $(S, w) \in \mathcal{S}^{poss}$ is defined respectively as follows:

$$\begin{array}{lll} S, w \models \ell(\varphi) \geq \ell(\psi) & \text{iff} & Pl_w([\![\varphi]\!]) \geq Pl_w([\![\psi]\!]) \\ S, w \models \ell(\varphi) \geq \ell(\psi) & \text{iff} & \kappa_w([\![\varphi]\!]) \leq \kappa_w([\![\psi]\!]) \\ S, w \models \ell(\varphi) \geq \ell(\psi) & \text{iff} & Poss_w([\![\varphi]\!]) \geq Poss_w([\![\psi]\!]) \end{array}$$

The other Boolean truth conditions are the same as for $\mathcal{S}^{pref}$. □

Another semantics can be given if we use $\succeq_w^s$ instead of $\succeq_w^e$ in the truth condition for $\ell(\varphi) \geq \ell(\psi)$. Like for logics for quantitative reasoning of the previous section, the set of validities of the logics $(\mathcal{L}_{\mathsf{Qual}}, \mathcal{S}^{pref}, \models)$, $(\mathcal{L}_{\mathsf{Qual}}, \mathcal{S}^{qual}, \models)$, $(\mathcal{L}_{\mathsf{Qual}}, \mathcal{S}^{rank}, \models)$ and $(\mathcal{L}_{\mathsf{Qual}}, \mathcal{S}^{poss}, \models)$ can be axiomatized (Halpern, 2003, Chap 7). These axiomatizations can then be used to *reason* about uncertainty in *qualitative* terms.

3.4 Updating Uncertainty

In this section, we adopt the following assumptions. Assumption 1 will be removed in Chapter 6 when we deal with belief revision.

Assumption 1 What the agent is told is true and it initially considers the actual world possible.

Assumption 2 The way an agent obtains the new information does not itself give the agent information.

3.4.1 Conditioning Probabilities

From *un*conditional to conditional probability. How should a probability measure μ be updated to a new probability $\mu_{|U}$ that takes the new information that the actual world is in U into account ?

First, if the agent believes that U is true, then it seems reasonable to require that all the worlds in $\overline{U}$ are impossible:

$$\mu_{|U}(\overline{U}) = 0 \tag{3.3}$$

Second, if all that the agent has learned is U, then the relative likelihood of worlds in U should remain unchanged (because of assumption 3). That is, if $V_1, V_2 \subseteq U$ with $\mu(V_2) > 0$, then

$$\frac{\mu(V_1)}{\mu(V_2)} = \frac{\mu_{|U}(V_1)}{\mu_{|U}(V_2)}. \tag{3.4}$$

Equations (3.3) and (3.4) completely determine $\mu_{|U}$ if $\mu(U) > 0$:

Proposition 3.4.1. *If $\mu(U) > 0$ and $\mu_{|U}$ is a probability measure on W satisfying* (3.3) *and* (3.4)*, then*

$$\mu_{|U}(V) = \frac{\mu(V \cap U)}{\mu(U)} \tag{3.5}$$

We often write $\mu(V \mid U)$ rather than $\mu_{|U}(V)$. The function $\mu_{|U}$ is called a *conditional probability (measure)* and $\mu(V \mid U)$ is read "the probability of V *given* (or *conditional on*) U." Sometimes, $\mu(U)$ is called the *unconditional* probability of U.

From conditional to *un*conditional probability. Conditional probability is defined only if $\mu(U) \neq 0$. Worlds of probability 0 are somehow problematic from a conceptual point of view. Are they really impossible ? How unlikely does a world have to be before it is assigned probability 0 ? Should a world ever be assigned probability 0 ? If there are worlds with probability 0 that are not truly impossible, then what does it mean to condition on sets with probability 0 ?

This leads us to define the notion of *conditional* probability. It is more primitive and basic than unconditional probability and it will allow us to conditionalize on sets U of probability 0.

Definition 3.4.1 (Popper algebra). A *Popper algebra* over W is a set $\mathcal{F} \times \mathcal{F}'$ of subsets of $W \times W$ such that:

1. $\mathcal{F}$ is an algebra over W;
2. $\mathcal{F}'$ is a non-empty subset of $\mathcal{F}$;
3. $\mathcal{F}'$ is closed under supersets in $\mathcal{F}$, *i.e.* if $V \in \mathcal{F}'$, $V \subseteq V'$, and $V' \in \mathcal{F}$, then $V' \in \mathcal{F}'$. □

Note that in the definition of a Popper algebra, $\mathcal{F}'$ need not be an algebra.

Definition 3.4.2 (Conditional probability space). A *conditional probability space* is a tuple $(W, \mathcal{F}, \mathcal{F}', \mu)$ such that $\mathcal{F} \times \mathcal{F}'$ is a Popper algebra over W and $\mu : \mathcal{F} \times \mathcal{F}' \to [0; 1]$ satisfies the following conditions:

$\mu(U \mid U) = 1$	if $U \in \mathcal{F}'$	(CP1)
$\mu(V_1 \cup V_2 \mid U) = \mu(V_1 \mid U) + \mu(V_2 \mid U)$	if $V_1 \cap V_2 = \emptyset, V_1, V_2 \in \mathcal{F}$, $U \in \mathcal{F}'$	(CP2)
$\mu(V \mid U) = \mu(V \cap U \mid U)$	if $U \in \mathcal{F}'$, $V \in \mathcal{F}$	(CP3)
$\mu(U_1 \mid U_3) = \mu(U_1 \mid U_2) \times \mu(U_2 \mid U_3)$	if $U_1 \subseteq U_2 \subseteq U_3$, $U_2, U_3 \in \mathcal{F}'$, $U_1 \in \mathcal{F}$	(CP4)

From a conditional probability measure, we obtain an *un*conditional probability measure by conditioning on W. Conversely, given an *un*conditional probability measure μ, we can define naturally a conditional probability measure by considering Equation (3.5).[1] This conditional probability will satisfy Conditions (CP1) – (CP4).

However, conditional probability measures are more primitive because there are conditional probability measures that are extensions of unconditional probability measures μ such that $\mathcal{F}'$ includes some sets U for which $\mu(U) = 0$ (for example, see the *non-standard probability measures* of (Halpern, 2003, p. 76)).

[1] If an unconditional probability measure is identified with a conditional probability measure defined on $\mathcal{F} \times \{W\}$, then the conditional probabilities defined by Equation (3.5) are *extensions* of μ to $\mathcal{F} \times \mathcal{F}'$ in the sense of "extensions" as they are defined in Section 3.2.1.1.

Justifying probabilistic conditioning. Probabilistic conditioning can be justified in much the same way that probability is justified.

Objective interpretation: Principle of Indifference and relative-frequency. If it seems reasonable to apply the principle of indifference to W and then U is observed or learned, it seems equally reasonable to apply the principle of indifference again to $W \cap U$. This results in taking all the elements of $W \cap U$ to be equally likely and assigning all the elements in $\overline{W \cap U}$ probability 0, which is exactly what (3.5) says. In the *relative-frequency* interpretation, $\mu(V \mid U)$ can be viewed as the fraction of times that V occurs of the times that U occurs. Again, Equation (3.5) holds.

Subjective interpretation: Dutch book. Let $(V \mid U, \alpha)$ denote the following bet:

> If U happens and if V also happens, then I win $100(1-\alpha)$ euros, while if $\overline{V}$ happens, then I lose 100α euros. If U does not happen, then the bet is called off (I do not win or lose anything).

As before, suppose that the agent has to choose between bets of the form $(V \mid U, \alpha)$ and $(\overline{V} \mid U, 1-\alpha)$. For worlds in $\overline{U}$, both bets are called off, so they are equivalent. Then,

Theorem 3.4.1. *If an agent satisfies* (RAT1)–(RAT4)*, then for all $U, V \subseteq W$ such that $\alpha_U > 0$, there is a number $\alpha_{V|U}$ such that $(V \mid U, \alpha) \succeq (\overline{V} \mid U, 1-\alpha)$ for all $\alpha < \alpha_{V|U}$ and $(\overline{V} \mid U, 1-\alpha) \succeq (V \mid U, \alpha)$ for all $\alpha > \alpha_{V|U}$. Moreover, $\alpha_{V|U} = \dfrac{\alpha_{V \cap U}}{\alpha_U}$.*

We assume implicitly that conditioning by U amounts to consider not just the agent's current beliefs regarding V if U *were* to occur, but also how the agent would change his beliefs regarding V if U *actually* did occur.

Bayes' Rule. One of the most important results in probability theory is Bayes' Rule, even if its proof is straightfoward. It relates $\mu(V \mid U)$ and $\mu(U \mid V)$.

Proposition 3.4.2 (Bayes' Rule). *If $\mu(U), \mu(V) > 0$, then*

$$\mu(V \mid U) = \frac{\mu(U \mid V)\mu(V)}{\mu(U)}. \qquad \text{(Bayes' Rule)}$$

Proof. The proof just consists of simple algebraic manipulation. Observe that

$$\frac{\mu(U \mid V)\mu(V)}{\mu(U)} = \frac{\mu(V \cap U)\mu(V)}{\mu(U)\mu(V)} = \frac{\mu(V \cap U)}{\mu(U)} = \mu(V \mid U). \qquad \square$$

3.4.2 Conditioning Sets of Probabilities, Inner and Outer Measures

Suppose an agent's uncertainty is defined in terms of a set $\mathcal{P}$ of probability measures. If the agent observes U, the obvious thing to do is to condition each member of $\mathcal{P}$ on U. We have, however, to remove the probability measures $\mu \in \mathcal{P}$ for which $\mu(U) = 0$.

Definition 3.4.3 (Conditional sets of probabilities). Let $\mathcal{P}$ be a set of probability measures and let $U \subseteq W$. We define the *conditional set of* $\mathcal{P}$ *by* U as follows:

$$\mathcal{P} \mid U := \left\{ \mu_{|U} \mid \mu \in \mathcal{P}, \mu(U) > 0 \right\}. \qquad \square$$

Given the connection between lower/upper probability measures and inner/outer measures shown of 3.2.1, we define the conditional inner/outer measures as follows:

Definition 3.4.4 (Conditional inner and outer measures). Let μ be a probability measure on $\mathcal{F}'$ and let $\mathcal{P}_\mu$ consist of all the extensions of μ to $\mathcal{F}$: $\mathcal{P}_\mu := \{\mu' \mid \mu' \text{ is a probability measure on } \mathcal{F} \text{ and } \mu'(U) = \mu(U) \text{ for all } U \in \mathcal{F}'\}$. We define the *conditional inner and outer measures* as follows: for all $U, V \in \mathcal{F}$ such that $\mu^*(U) > 0$,

$$\mu_*(V \mid U) := (\mathcal{P}_\mu \mid U)_* (V)$$

$$\mu^*(V \mid U) := (\mathcal{P}_\mu \mid U)^* (V) \qquad \square$$

The following theorem provides a more constructive definition of conditional inner and outer measures.

Theorem 3.4.2. *Let μ be a probability measure. Suppose that $\mu^*(U) > 0$. Then,*

$$\mu_*(V \mid U) = \begin{cases} \dfrac{\mu_*(V \cap U)}{\mu_*(V \cap U) + \mu^*(\overline{V} \cap U)} & \text{if } \mu^*(\overline{V} \cap U) > 0 \\ 1 & \text{if } \mu^*(\overline{V} \cap U) = 0 \end{cases}$$

$$\mu^*(V \mid U) = \begin{cases} \dfrac{\mu^*(V \cap U)}{\mu^*(V \cap U) + \mu_*(\overline{V} \cap U)} & \text{if } \mu^*(V \cap U) > 0 \\ 0 & \text{if } \mu^*(V \cap U) = 0 \end{cases}$$

Example 3.4.1 (Three prisoners puzzle). The three-prisoners puzzle is as follows:

> Of three prisoners, a, b and c, two are to be executed, but a does not know which. He therefore says to the jailer, "Since either b or c is certainly going to be executed, you will give me no information about my own chances if you give me he name of one man, either b or c, who is going to be executed". Accepting this argument, the jailer truthfully replies, "b will be executed". After this annoucement, what is the probability that a will live ? Should it be $\frac{1}{2}$, or $\frac{1}{3}$ as before ?

We represent the problem as follows. A possible situation is a pair (x, y) where $x, y \in \{a, b, c\}$. Intuitively, a pair (x, y) represents a situation where x is pardoned and the jailer says that y will be executed in response to a's question. This yields the set of possible worlds $W := \{(a, b), (a, c), (b, c), c, b)\}$. Then, we define the following events:

- $says_b := \{(a,b),(c,b)\}$ corresponds to the event where the jailer says that b will be executed;
- $live_a := \{(a,b),(a,c)\}$ corresponds to the event where a is pardoned and lives;
- $live_b := \{(b,c)\}$ corresponds to the event where b lives;
- $live_c := \{(c,b)\}$ corresponds to the event where c lives.

According to the principle of indifference, each prisoner is equally likely to be pardoned:

$$\mu(lives_a) = \mu(lives_b) = \mu(lives_c) = \frac{1}{3}.$$

Let $\mathcal{F}'$ consist of all the sets that can be formed by taking unions of $lives_a, lives_b, lives_c$: $\mathcal{F}' := \{\emptyset, \{(b,c)\}, \{(c,b)\}, \{(b,c),(c,b)\}, \{(a,b),(a,c)\}, \{(a,b),(a,c),(b,c)\}, \{(a,b),(a,c),(c,b)\}, W\}$. Note that $says_b = \{(a,b),(c,b)\} \notin \mathcal{F}'$. Likewise, the set $\{(a,b)\}$ corresponding to the event where a lives and the jailer says that b will be executed does *not* belong to $\mathcal{F}'$. What could be the probability of $\{(a,b)\}$?

- *Case 1*: a assumes that the jailer applies the principle of indifference in choosing between b and c if a is pardoned. In that case, we have

  $$\mu(\{(a,b)\}) = \mu(\{(a,c)\} = \frac{\mu(\{(a,b),(a,c)\})}{2} = \frac{1}{6}.$$

  With this assumption,

  $$\mu(lives_a \mid says_b) = \frac{\mu(lives_a \cap says_b)}{\mu(says_b)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

  The intuitive answer – that the jailer's answer gives a no information – is correct if the jailer applies the principle of indifference.

- *Case 2*: a assumes that the jailer does *not* apply the principle of indifference. In other words, a does *not* know what strategy the jailer is using to answer (and is not willing to place a probability on these strategies).

  We consider the set of probability measures $\mathcal{P}_J := \{\mu_\alpha \mid \alpha \in [0;1]\}$, where

  $$\mu_\alpha(lives_a) = \mu_\alpha(lives_b) = \mu_\alpha(lives_c) = \frac{1}{3} \qquad \mu_\alpha(says_b \mid lives_a) = \alpha$$

  So, for example, if $\alpha = 0$, then if a is pardoned, the jailer will definitely say c. Thus, if the jailer actually says b, then a knows that he is definitely not pardoned, that is, $\mu_0(lives_a \mid says_b) = 0$. Similarly, if $\alpha = 1$, then a knows that if either he or c is pardoned, then the jailer will say b, while if b is pardoned, the jailer will say c. Given that the jailer says b, from a's point of view, the one pardoned is equally likely to be him or c; thus, $\mu_1(lives_a \mid says_b) = \frac{1}{2}$.

Also, we consider the probability measure μ_J on $\mathcal{F}'$ that agrees with each of the measure in $\mathcal{P}_J$. Then, we have that

$$(\mu_J)_*(lives_a \cap says_b) = (\mu_J)_*(\{(a,b)\}) = (\mathcal{P}_J)_*(\{(a,b)\}) = 0$$

$$(\mu_J)^*(lives_a \cap says_b) = (\mu_J)^*(\{(a,b)\}) = (\mathcal{P}_J)^*(\{(a,b)\}) = \frac{1}{3}$$

$$\begin{aligned}(\mu_J)^*(\overline{lives_a} \cap says_b) &= (\mu_J)^*(\{(c,b)\}) = (\mu_J)^*(\{(c,b)\}) \\ &= (\mathcal{P}_J)_*(\{(c,b)\}) = (\mathcal{P}_J)^*(\{(c,b)\}) = \frac{1}{3}\end{aligned}$$

Thanks to these results and the expressions of Theorem 3.4.2, we obtain

$$(\mu_J)_*(lives_a \mid says_b) = \frac{(\mu_J)_*(lives_a \cap says_b)}{(\mu_J)_*(lives_a \cap says_b) + (\mu_J)^*(\overline{lives_a} \cap says_b)} = 0$$

$$(\mu_J)^*(lives_a \mid says_b) = \frac{(\mu_J)^*(lives_a \cap says_b)}{(\mu_J)^*(lives_a \cap says_b) + (\mu_J)_*(\overline{lives_a} \cap says_b)} = \frac{1}{2}.$$

So, when prisoner a does not know the strategy of the jailer concerning the announcement, his prior point probability of $\frac{1}{3}$ "diffuses" to an interval $\left[0; \frac{1}{2}\right]$. □

3.4.3 Conditioning Belief Functions

Recall from Theorem 3.2.2 that given a belief function Bel, the set $\mathcal{P}_{Bel} = \{\mu \mid \mu(U) \geq Bel(U) \text{ for all } U \subseteq W\}$ of probability measures is such that $Bel = (\mathcal{P}_{Bel})_*$ and $Plaus = (\mathcal{P}_{Bel})^*$. The association of Bel with $\mathcal{P}_{Bel}$ can be used to define a notion of conditional belief in terms of conditioning on sets of probability measures.

Definition 3.4.5 (Conditional belief function). Given a belief function Bel defined on W and a set U such that $Plaus(U) > 0$, define functions $Bel_{|U} : 2^W \to [0;1]$ and $Plaus_{|U} : 2^W \to [0;1]$ as follows:

$$Bel_{|U}(V) := (\mathcal{P}_{Bel} \mid U)_*(V)$$

$$Plaus_{|U}(V) := (\mathcal{P}_{Bel} \mid U)^*(V)$$ □

Given the close relationship between beliefs and inner measures, the following analogue of Theorem 3.4.2 should not come as a surprise. This theorem provides an alternative and more constructive definition of conditional belief functions.

Theorem 3.4.3. *Let Bel be a belief function and Plaus its associated plausibility function. Suppose that* $Plaus(U) > 0$. *Then,*

$$Bel(V \mid U) = \begin{cases} \dfrac{Bel(V \cap U)}{Bel(V \cap U) + Plaus(\overline{V} \cap U)} & \text{if } Plaus(\overline{V} \cap U) > 0 \\ 1 & \text{if } Plaus(\overline{V} \cap U) = 0 \end{cases}$$

$$Plaus(V \mid U) = \begin{cases} \dfrac{Plaus(V \cap U)}{Plaus(V \cap U) + Bel(\overline{V} \cap U)} & \text{if } Plaus(V \cap U) > 0 \\ 0 & \text{if } Plaus(V \cap U) = 0 \end{cases}$$

Moreover, $Bel_{|U}$ *is a belief function and* $Plaus_{|U}$ *is its corresponding plausibility function.*

3.4.4 Conditioning Possibility Measures and Ranking Functions

In this section, we simply give the definitions of conditional possibility measures and conditional ranking functions. Just as the definition of a conditional probability measure is motivated by the postulates (CP1)–(CP4) of Definition 3.4.2, the definition of a conditional possibility measure and of a conditional ranking function given in Equations (3.6) and (3.7) respectively can be motivated by postulates similar to the postulates of (CP1)–(CP4) of Definition 3.4.2 (see (Halpern, 2003, p. 95–97) for more details).

Definition 3.4.6 (Conditional possibility measure). Let *Poss* be a possibility measure. The *conditional possibility measure* $Poss_{|U}$ is defined by: for all $U \subseteq W$ such that $Poss(U) > 0$,

$$Poss_{|U}(V) = \begin{cases} Poss(V \cap U) & \text{if } Poss(V \cap U) < Poss(U), \\ 1 & \text{if } Poss(V \cap U) = Poss(U). \end{cases} \tag{3.6}$$

Definition 3.4.7 (Conditional ranking function). Let κ be a ranking function. The *conditional ranking function* $\kappa_{|U}$ is defined by: for all $U \subseteq W$ such that $\kappa(U) \neq \infty$,

$$\kappa(V \mid U) = \kappa(V \cap U) - \kappa(U) \tag{3.7}$$

3.4.5 Jeffrey's Rule

Jeffrey's Rule applies to situations where the agent might have some uncertainty about the incoming information.

Example 3.4.2. Suppose that an object is either red, blue, green or violet. An agent initially ascribes probability $\frac{1}{5}$ to each of *red*, *blue*, and *green*, and probability $\frac{2}{5}$ to *violet*. Then the agent gets a quick glimpse of the object in a dimly lit room. As a result of this glimpse, he believes that the object is probably a darker color, although he is not

sure. He thus ascribes probability 0.7 to it being green or blue and probability 0.3 to it being red or violet. How should he update his initial probability measure based on this observation ?

We represent the agent's observation as follows:

$$0.7\{blue, green\}; 0.3\{red, violet\}$$

The example suggest that an appropriate way of updating the agent's initial probability measure μ is to consider the linear combination:

$$\mu' := 0.7\mu_{|\{blue,green\}} + 0.3\mu_{|\{red,violet\}}.$$

As expected, $\mu(\{blue, green\}) = 0.7$ and $\mu'(\{red, violet\}) = 0.3$. Moreover, $\mu'(red) = 0.1$, $\mu'(violet) = 0.2$ and $\mu'(blue) = \mu'(green) = 0.35$. Thus, μ' gives the two sets about which the agent information – $\{blue, green\}$ and $\{red, violet\}$ – the expected probabilities. Within each of these sets, the relative probability of the outcomes remains the same as before conditioning. □

More generally, suppose that $U_1, \ldots, U_n$ is a partition of W (see Section A.4 of the Appendix for the definition of a partition). An *observation* over W is an expression of the form $\alpha_1 U_1; \ldots; \alpha_n U_n$, where $\alpha_1 + \ldots + \alpha_n = 1$. This is to be interpreted as an observation that leads the agent to believe U_j with probability α_j, for $j = 1, \ldots, n$. This suggests that $\mu_{|\alpha_1 U_1;\ldots;\alpha_n U_n}$, the probability measure resulting from the update, should have the following property for $j = 1, \ldots, n$:

$$\mu_{|\alpha_1 U_1;\ldots;\alpha_n U_n}(V) = \alpha_j \frac{\mu(V)}{\mu(U_j)} \text{ if } V \subseteq U_j \text{ and } \mu(U_j) > 0. \tag{J}$$

An observation is *consistent* with a probability measure μ if it does not give positive probability to a set that was initially thought to have probability 0: formally, if $\alpha_j > 0$, then $\mu(U_j) > 0$.

Proposition 3.4.3 (Jeffrey's Rule). *Let μ be a probability measure and let $\alpha_1 U_1; \ldots; \alpha_n U_n$ be an observation on W which is consistent with μ. If $\mu_{|\alpha_1 U_1;\ldots;\alpha_n U_n}$ is a probability measure satisfying Condition* (J)*, then*

$$\mu_{|\alpha_1 U_1;\ldots;\alpha_n U_n} = \alpha_1 \mu(V \mid U_1) + \ldots + \alpha_n \mu(V \mid U_n). \tag{Jeffrey's Rule}$$

Note that the usual notion of probabilistic conditioning is just a special case of Jeffrey's rule: $\mu_{|U} = \mu_{|1U;0\overline{U}}$. However, unlike Jeffrey's rule, probabilistic conditioning *is* commutative: if $\mu(U_1 \cap U_2) \neq \emptyset$, then

$$\left(\mu_{|U_1}\right)_{|U_2} = \left(\mu_{|U_2}\right)_{|U_1} = \mu_{|U_1 \cap U_2}$$

This does not hold for Jeffrey's Rule: observing $o_1 = 0.7\{blue, green\}; 0.3\{red, violet\}$ and then observing $o_2 = 0.3\{blue, green\}; 0.7\{red, violet\}$ does not yield the same result

as observing first o_2 and then o_1: $\left(\mu_{|o_1}\right)_{|o_2} \neq \left(\mu_{|o_2}\right)_{|o_1}$. The last observation has always priority over the previous observations for Jeffrey's Rule.

There are straightforward analogues of Jeffrey's Rule for sets of probabilities, belief functions, possibility measures and ranking functions.

- *Belief function*:

$$Bel_{|\alpha_1 U_1;\ldots;\alpha_n U_n} = \alpha_1 Bel_{|U_1} + \ldots + \alpha_n Bel_{|U_n}$$

- *Possibility Measures*: note that + and × of probabilities become *max* and *min*.

$$\begin{aligned} &Poss_{|\alpha_1 U_1;\ldots;\alpha_n U_n}(V) \\ &\quad = max\left\{min\{\alpha_1, Poss(V \mid U_1)\}, \ldots, min\{\alpha_n, Poss(V \mid U_n)\}\right\}. \end{aligned}$$

- *Ranking functions*: note that + becomes *min* and the role of 1 is played by 0.

$$\kappa_{|\alpha_1 U_1;\ldots;\alpha_n U_n}(V) = min\{\alpha_1 + \kappa(V \mid U_1), \ldots, \alpha_n + \kappa(V \mid U_n)\}.$$

3.5 Further Reading

This Chapter is based on Chapters 2 and 3 of (Halpern, 2003). The representation and management of uncertainty is a vast area of research. We mention for example the series of handbooks (Gabbay and Smets, 1998).

Chapter 4

Reasoning with Others about Uncertainty

"Chuangtse and Hueitse had strolled onto the bridge over the Hao, when the former observed, "See how the small fish are darting about ! That is the happiness of the fish." "You are not a fish yourself," said Hueitse. "How can you know the happiness of the fish?" "And you not being I," retorted Chuangtse, "how can you know that I do not know ?""

– Chuangtse, c. 300 B.C.

4.1 Introduction

Modeling a situation involving multiple agents depends very much on the modeling point of view. Indeed, the models built to represent the situation will be quite different depending on whether the modeler is an agent involved in the situation or not. To illustrate this point, let us consider the following example. Assume that the agents Yann and Alice are in a room and that there is a coin in a box that both cannot see because the box is closed. Now, assume that Alice cheats, opens the box and looks at the coin. Moreover, assume that Yann does not suspect anything about it and that Alice knows it (Yann might be inattentive or out of the room for a while). How can we represent this resulting situation? On the one hand, if the modeler is an external observer (different from Yann and Alice) knowing everything that has happened, then in the model that this external observer builds Alice knows whether the coin is heads or tails up. On the other hand, if the modeler is Yann then in the model that Yann builds Alice does *not* know whether the coin is heads or tails up. As we see in this example, the intuitive interpretation of a model really makes sense only when one knows the modeling point of view.

The importance of specifying a modeling point of view is also stressed at a great extent in Newtonian mechanics in physics where physicists must always specify which *frame of reference* they consider when they want to study a natural phenomenon. And just as for epistemic situations, the representation of this phenomenon depends very much on this frame of reference. For example, assume somebody drops a ball from the top of the high mast of a ship sailing nearby a harbor. Then, viewed from the frame of

reference of the ship, the trajectory of the ball will be a straight line. But viewed from the frame of reference of the harbor, the trajectory will be a parabola (the more rapidly the ship sails and the higher the mast is, the more curved the parabola will be).

Given an epistemic situation, assume that we want to model the beliefs of the agents $AGTS = \{1, \ldots, n\}$ and possibly the actual state of the world. What kinds of modeling points of view are there? For a start, we can distinguish whether or not the modeler is one of these agents $AGTS$ under scrutiny.

1. First, consider the case where the modeler is *one* of the agents $AGTS$. In what follows, we call this modeler-agent agent Y (like You). The models she builds could be seen as models she has 'in her mind'. They represent the way she perceives the surrounding world. In that case, agent Y is involved in the situation, she is considered on a par by the other agents and interacts with them. So she should be represented in the formalism and her models should deal not only with the other agents' beliefs but also with the other agents' beliefs about her own beliefs. This is an internal and subjective point of view, the situation is modeled from the inside. Therefore, for this very reason her beliefs might be erroneous. Hence the models she builds might also be erroneous. We call this agent point of view the *internal* point of view.

2. Second, consider the case where the modeler is *not* one of the agents $AGTS$. The modeler is thus an observer external to the situation. She is not involved in the situation and she does not exist for the agents, or at least she is not taken into consideration in their representation of the world. So she should not be represented in the formalism and particularly the agents' beliefs about her own beliefs should also not be represented. The models that this modeler builds are supposed to represent the situation 'from above', from an external and objective point of view. There are then two other sub-possibilities depending on whether or not the modeler has a perfect knowledge of the situation.

   (a) In case the modeler has a perfect knowledge of the situation, then everything that is true in the model that she builds is true in reality and vice versa, everything that is true in reality is also true in the model. This thesis was already introduced by Baltag and Moss (2004). Basically, the models built by the modeler are perfectly correct. The modeler has access to the minds of the agents and knows perfectly not only what they believe but also what the actual state of the world is. This is a kind of 'divine' point of view and we call it the *perfect external* point of view.

   (b) In case the modeler does not have a perfect knowledge of the situation then, like the internal point of view but unlike the perfect external point of view, the models built might be erroneous. The models could also be correct but then, typically, the modeler would be uncertain about which is the actual world (in that sense, she would not have a perfect knowledge of the situation).

	the modeler is uncertain about the situation	the modeler is one of the agents
internal approach	•	•
imperfect external approach	•	
perfect external approach		

Figure 4.1: Essential differences between the internal and external approaches

> What the modeler knows can be obtained for example by observing what the agents say and do, by asking them questions... We call this point of view the *imperfect external* point of view.

Because we proceeded by successive dichotomies, we claim that the internal, the perfect external and the imperfect external points of view are the only three logically possible points of view when we want to model epistemic situations. From now on we will call these modeling approaches the internal, the external and the imperfect external approaches; their differences are summarized in Figure 4.1.[1] The fields of application of these three approaches are different. The internal and imperfect external approaches have rather applications in artificial intelligence where agents/robots acting in the world need to have a formal representation of the surrounding world and to cope with uncertain information. The internal approach has also applications in cognitive psychology where the aim is to model the cognition of one agent (possibly in a multi-agent setting). The perfect external approach has rather applications in game theory (Battigalli and Bonanno, 1999), social psychology (distributed cognition) or distributed systems (Fagin et al., 1995) for example. Indeed, in these fields we need to model situations accurately from an external point of view in order to explain and predict what happens in these situations.

The modeling point of view is definitely not the only important factor to specify when one wants to model epistemic situations: the second important factor is obviously our *object of study*, i.e. *what* we actually model. Typically, it is the actual state of the world and the beliefs of the agents $AGTS$ about each other. But this could also perfectly be their beliefs about other agents $j_1, \ldots, j_m$ or the beliefs of only *some* of these agents $AGTS$ (about *all* the agents $AGTS$) for instance. Therefore, to proceed methodically and properly (and similarly as in physics), when one wants to model epistemic situations one should ideally specify from the start a combination of these two factors. Indeed, each combination gives rise to a particular kind of formalism. However some

[1] In (Nagel, 1986), the internal and external points of view are studied from a broader philosophical perspective and not just for their need in representing agents' beliefs. Nagel mainly deals there with the issues of how these views can be combined and if they can possibly be integrated. He does so by tracing the manifestations of these issues in a number of philosophical topics: the metaphysics of mind, the theory of knowledge, free will, and ethics. He argues that giving a complete account of reality (as in philosophy of mind) or of all reasons for actions (as in ethics) in objective terms *only* is not always possible.

combinations might turn out to be equivalent to others: for example, if the object of study is the epistemic state of a single agent Y (in a single or a multi-agent setting), then the perfect external approach for this object of study amounts to the internal approach where the modeler-agent is Y herself and the object of study is the actual state of the world (and possibly the other agents' beliefs about each other in a multi-agent setting). This example suggests that the internal approach is somehow reducible to the perfect external approach if we specify appropriate objects of study. But because the corresponding object of study in the external approach of a given object of study in the internal approach might be quite convoluted in some cases we prefer to keep the natural and intuitive distinction between the internal and the perfect external approaches.

All this said, in the rest of the chapter, we will only follow and consider the perfect external approach. The logical framework we will study to deal with uncertainty in a multi-agent setting is Dynamic Epistemic Logic (DEL). It is built on top of epistemic logic to which it adds dynamics and events. Epistemic logic will be the topic of Section 4.2. Then, in Section 4.3, actions and events will enter into the picture and we will introduce the logical framework of DEL.[2]

4.2 Representing and Reasoning about Uncertainty: Epistemic Logic

Epistemic logic is a modal logic that is concerned with the logical study of the notions of knowledge and belief. It is thereby concerned with understanding the process of *reasoning* about knowledge and belief: which principles relating the notions of knowledge and belief are intuitively plausible ? As epistemology, it stems from the Greek word $\epsilon\pi\iota\sigma\tau\eta\mu\eta$ or 'episteme' meaning knowledge. But epistemology is more concerned with analyzing the very *nature* of knowledge (addressing questions such as "What is the definition of knowledge?" or "How is knowledge acquired?"). In fact, epistemic logic grew out of epistemology in the middle ages thanks to the efforts of Burley and Ockham (Boh, 1993). But the formal work, based on modal logic, that inaugurated contemporary research into epistemic logic dates back only to 1962 and is due to Hintikka (Hintikka, 1962). It then sparked in the 1960's discussions about the principles of knowledge and belief and many axioms for these notions were proposed and discussed (Lenzen, 1978). For example, the interaction axioms $K\varphi \rightarrow B\varphi$ and $B\varphi \rightarrow KB\varphi$ are often considered to be intuitive principles: if agent a knows φ then (s)he also believes φ, or if agent a believes φ, then (s)he knows that (s)he believes φ. More recently, these kinds of philosophical theories were taken up by researchers in economics (Battigalli and Bonanno, 1999), artificial intelligence and theoretical computer science (Fagin et al., 1995; Meyer and van der Hoek, 1995) where reasoning about knowledge is a central topic. Due to the new setting in which epistemic logic was used, new perspectives and new features such

[2] A distinction is sometimes made between events and actions, an action being a specific type of event performed by an agent. In the sequel, we will not make this distinction and we will use alternatively the term action or event.

as computability issues were then added to the agenda of epistemic logic.

4.2.1 Syntax and Semantics

Syntax. The epistemic language is an extension of the basic modal language of Definition 1.3.4 with a *common knowledge* operator C_A and a *distributed knowledge* operator D_A. These new operators are discussed after the following definition.

Definition 4.2.1 (Epistemic language $\mathcal{L}_{\mathsf{EL}}$). The *epistemic language* $\mathcal{L}_{\mathsf{EL}}$ is defined inductively as follows:

$$\mathcal{L}_{\mathsf{EL}} : \varphi \quad ::= \quad p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_j\varphi \mid C_A\varphi \mid D_A\varphi$$

where $p \in PROP$, $j \in AGTS$ and $A \subseteq AGTS$. The formula $\langle K_j\rangle\varphi$ is an abbreviation for $\neg\Box_j\neg\varphi$, $E_A\varphi$ is an abbreviation for $\bigwedge_{j\in A} K_j\varphi$ and $C\varphi$ an abbreviation for $C_{AGTS}\varphi$. □

For example, if $A = \{1, 2, 3\}$, then $\bigwedge_{j\in A} K_j(p \wedge q) = K_1(p \wedge q) \wedge K_2(p \wedge q) \wedge K_3(p \wedge q)$.

Group notions: general, common and distributed knowledge. In a multi-agent setting there are three important epistemic concepts: general belief (or knowledge), distributed belief (or knowledge) and common belief (or knowledge). The notion of common belief (or knowledge) was first studied by Lewis in the context of conventions (Lewis, 1969). It was then applied to distributed systems (Fagin et al., 1995) and to game theory (Aumann, 1976), where it allows to express that the rationality of the players, the rules of the game and the set of players are commonly known.

General knowledge. General belief of φ means that everybody in the group of agents $AGTS$ believes that φ. Formally this corresponds to the following formula:

$$E\varphi := \bigwedge_{j\in AGTS} K_j\varphi. \tag{4.1}$$

Common knowledge. Common belief of φ means that everybody believes φ but also that everybody believes that everybody believes φ, that everybody believes that everybody believes that everybody believes φ, and so on *ad infinitum*. Formally, this corresponds to the following formula

$$C\varphi := E\varphi \wedge EE\varphi \wedge EEE\varphi \wedge \ldots \tag{4.2}$$

As we do not allow infinite conjunction the notion of common knowledge will have to be introduced as a primitive in our language.

Before defining the language with this new operator, we are going to give an example introduced by Lewis (1969) that illustrates the difference between these two notions (here we exceptionally use the notion of knowledge instead of belief to

make things clearer). Lewis wanted to know what kind of knowledge is needed so that the statement p: "every driver must drive on the right" be a convention among a group of agents. In other words he wanted to know what kind of knowledge is needed so that everybody feels safe to drive on the right. Suppose there are only two agents i and j. Then everybody knowing p (formally Ep) is not enough. Indeed, it might still be possible that the agent i considers possible that the agent j does not know p (formally $\neg K_i K_j p$). In that case the agent i will not feel safe to drive on the right because he might consider that the agent j, not knowing p, could drive on the left. To avoid this problem, we could then assume that everybody knows that everybody knows that p (formally EEp). This is again not enough to ensure that everybody feels safe to drive on the right. Indeed, it might still be possible that agent i considers possible that agent j considers possible that agent i does not know p (formally $\neg K_i K_j K_i p$). In that case and from i's point of view, j considers possible that i, not knowing p, will drive on the left. So from i's point of view, j might drive on the left as well (by the same argument as above). So i will not feel safe to drive on the right. Reasoning by induction, Lewis showed that for any $k \in \mathbb{N}$, $Ep \wedge E^1 p \wedge \ldots \wedge E^k p$ is not enough for the drivers to feel safe to drive on the right. In fact what we need is an infinite conjunction. In other words, we need common knowledge of p: Cp.

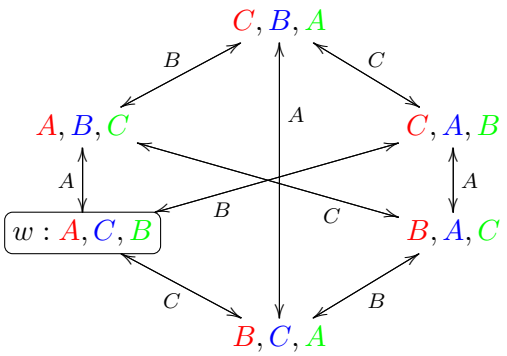
Distributed knowledge. Distributed knowledge of φ means that if the agents pulled their knowledge altogether, they would know that φ holds. In other words, the knowledge of φ is *distributed* among the agents. The formula $D_A\varphi$ reads as 'it is distributed knowledge among the set of agents A that φ holds'.

Semantics. Epistemic logic is a modal logic. So, what we call an *epistemic model* $\mathcal{M} = (W, R_1, \ldots, R_n, V)$ is just a Kripke model as used in modal logic. The possible worlds W are the relevant worlds needed to define such a representation and the valuation V specifies which propositional facts (such as 'it is raining') are true in these worlds. Finally the accessibility relations R_j can model either the notion of knowledge or the notion of belief. We set $w' \in R_j(w)$ in case the world w' is compatible with agent j's belief (respectively knowledge) in world w. Intuitively, a *pointed epistemic model* $(\mathcal{M}, w_a)$, where $w_a \in \mathcal{M}$, represents from an external point of view how the actual world w_a is perceived by the agents $AGTS$.

Definition 4.2.2 (Satisfaction relation). For every epistemic model $\mathcal{M}$, $w \in \mathcal{M}$ and $\varphi \in \mathcal{L}_{\mathsf{EL}}$, define

$$\mathcal{M}, w \models C_A\varphi \quad \text{iff} \quad \text{for all } v \in \left(\bigcup_{j\in A} R_j\right)^+ (w),\ \mathcal{M}, v \models \varphi$$

$$\mathcal{M}, w \models D_A\varphi \quad \text{iff} \quad \text{for all } v \in \bigcap_{j\in A} R_j(w),\ \mathcal{M}, v \models \varphi$$



A: "Ann has the red card"

C: "Claire has the blue card"

B: "Bob has the green card"

$\mathcal{M}, w \models (A \wedge \Box_A A) \wedge (C \wedge \Box_C C) \wedge (B \wedge \Box_B B)$

$\mathcal{M}, w \models \Box_A(B \vee B) \wedge \Box_A(C \vee C)$

$\mathcal{M}, w \models E(A \vee A \vee A) \wedge C(A \vee A \vee A).$

Figure 4.2: Card Example

where $\left(\bigcup_{j \in A} R_j\right)^+$ is the transitive closure of $\bigcup_{j \in A} R_j$: we have that $v \in \left(\bigcup_{j \in A} R_j\right)^+ (w)$ if, and only if, there are $w_0, \ldots, w_n \in \mathcal{M}$ and $j_1, \ldots, j_n \in A$ such that $w_0 = w, w_n = v$ and for all $i \in \{1, \ldots, n\}$, $w_{i-1} R_{j_i} w_i$. □

Despite the fact that the notion of common belief has to be introduced as a primitive in the language, we can notice in this definition that epistemic models do not have to be modified in order to give truth value to the common knowledge and distributed knowledge operators.

Example 4.2.1 (Card example). Ann (A), Bob (B) and Claire (C) play a card game with three cards: a green one, a red one and a blue one. Each of them has a single card but they do not know the cards of the other players. This example is depicted in Figure 4.2. □

The notion of knowledge might comply to some constraints (or axioms) such as $K_j\varphi \rightarrow K_j K_j \varphi$: if agent j knows something, she knows that she knows it. These constraints might affect the nature of the accessibility relations R_j which may then comply to some extra properties. So, we are now going to define some particular classes of epistemic models that all add some extra constraints on the accessibility relations R_j. These constraints are matched by particular axioms for the knowledge operator K_j.

Definition 4.2.3 (Properties of accessibility relations). We give in Figure 4.3 a list of properties of the accessibility relations. We also give, below each property, the axiom which *defines* the class of epistemic frames that fulfill this property. We choose, without any particular reason, to use the knowledge modality to write these conditions. □

serial:	$R(w) \neq \emptyset$
D:	$K\varphi \rightarrow \langle K\rangle\varphi$
transitive:	If $w' \in R(w)$ and $w'' \in R(w')$, then $w'' \in R(w)$
4:	$K\varphi \rightarrow KK\varphi$
Euclidean:	If $w' \in R(w)$ and $w'' \in R(w)$, then $w' \in R(w'')$
5:	$\neg K\varphi \rightarrow K\neg K\varphi$
reflexive:	$w \in R(w)$
T:	$K\varphi \rightarrow \varphi$
symetric:	If $w' \in R(w)$, then $w \in R(w')$
B:	$\varphi \rightarrow K\neg K\neg\varphi$
confluent:	If $w' \in R(w)$ and $w'' \in R(w)$, then there is v such that $v \in R(w')$ and $v \in R(w'')$
.2:	$\langle K\rangle K\varphi \rightarrow K\langle K\rangle\varphi$
weakly connected:	If $w' \in R(w)$ and $w'' \in R(w)$, then $w' = w''$ or $w' \in R(w'')$ or $w'' \in R(w')$
.3:	$\langle K\rangle\varphi \wedge \langle K\rangle\psi \rightarrow \langle K\rangle(\varphi \wedge \psi) \vee \langle K\rangle(\psi \wedge \langle K\rangle\varphi) \vee \langle K\rangle(\varphi \wedge \langle K\rangle\psi)$
semi-Euclidean:	If $w'' \in R(w)$ and $w \notin R(w'')$ and $w' \in R(w)$, then $w'' \in R(w')$
.3.2:	$(\langle K\rangle\varphi \wedge \langle K\rangle K\psi) \rightarrow K(\langle K\rangle\varphi \vee \psi)$
R1:	If $w'' \in R(w)$ and $w \neq w''$ and $w' \in R(w)$, then $w'' \in R(w')$
.4:	$(\varphi \wedge \langle K\rangle K\varphi) \rightarrow K\varphi$

Figure 4.3: List of properties of accessibility relations and corresponding axioms

Knowledge versus Belief In this chapter we use the same notation K for both knowledge and belief. Hence, depending on the context, $K\varphi$ will either read 'the agent Knows that φ holds' or 'the agent Believes that φ holds'. A crucial difference is that, unlike knowledge, beliefs can be *wrong*: the Truth axiom $K\varphi \to \varphi$ holds only for knowledge, but not necessarily for belief. In the next section, we are going to examine other axioms, some of them pertain more to the notion of knowledge whereas some others pertain more to the notion of belief.

4.2.2 Axiomatization

The axioms of an epistemic logic obviously display the way the agents reason. For example the axiom K together with the rule of inference MP entail that if I know φ ($K\varphi$) and I know that φ implies ψ ($K(\varphi \to \psi)$) then I know that ψ ($K\psi$). Stronger constraints can be added. The following proof systems are often used in the literature.

Definition 4.2.4 (Proof Systems for $\mathcal{L}_{\mathsf{EL}}$). We define the following proof systems for $\mathcal{L}_{\mathsf{EL}}$:

$$\begin{array}{rclrclrcl} \mathsf{KD45} & = & \mathsf{K}+\mathsf{D}+4+5 & \mathsf{S4.3} & = & \mathsf{S4}+.3 & \mathsf{S5} & = & \mathsf{S4}+5. \\ \mathsf{S4} & = & \mathsf{K}+\mathsf{T}+4 & \mathsf{S4.3.2} & = & \mathsf{S4}+.3.2 & & & \\ \mathsf{S4.2} & = & \mathsf{S4}+.2 & \mathsf{S4.4} & = & \mathsf{S4}+.4 & & & \end{array}$$

We denote by $\mathbb{L}_{\mathsf{EL}}$ the set of proof systems $\mathbb{L}_{\mathsf{EL}} := \{\mathsf{K}, \mathsf{KD45}, \mathsf{S4}, \mathsf{S4.2}, \mathsf{S4.3}, \mathsf{S4.3.2}, \mathsf{S4.4}, \mathsf{S5}\}$.

Moreover, for all $\mathcal{H} \in \mathbb{L}_{\mathsf{EL}}$, we define the proof system $\mathcal{H}^{\mathsf{C}}$ by adding the following axiom schemes and rules of inference to those of $\mathcal{H}$. For all $A \subseteq AGTS$,

$$\begin{array}{ll} \mathsf{Dis} & K_j\varphi \to D_A\varphi \\ \mathsf{E} & E_A\varphi \leftrightarrow \bigwedge_{j\in A} K_j\varphi \\ \mathsf{Mix} & C_A\varphi \to E_A(\varphi \wedge C_A\varphi) \\ \mathsf{Ind} & \text{if } \varphi \to E_A(\psi \wedge \varphi) \text{ then } \varphi \to C_A\psi \qquad \text{(Induction Rule)} \end{array}$$

□

The relative strength of the proof systems for knowledge is as follows:

$$\mathsf{S4} \subset \mathsf{S4.2} \subset \mathsf{S4.3} \subset \mathsf{S4.3.2} \subset \mathsf{S4.4} \subset \mathsf{S5}. \tag{4.3}$$

So, all the theorems of $\mathsf{S4.2}$ are also theorems of $\mathsf{S4.3}, \mathsf{S4.3.2}, \mathsf{S4.4}$ and $\mathsf{S5}$. Many philosophers claim that in the most general cases, the logic of knowledge is $\mathsf{S4.2}$ or $\mathsf{S4.3}$ (Lenzen, 1978; Stalnaker, 2006). Typically, in computer science, the logic of belief (*doxastic* logic) is taken to be $\mathsf{KD45}$ and the logic of knowledge (*epistemic* logic) is taken to be $\mathsf{S5}$, even if the logic $\mathsf{S5}$ is only suitable for situations where the agents do not have mistaken beliefs.

We discuss the most important axioms of Figure 4.3. Axioms T and $\mathsf{4}$ state that if the agent knows a proposition, then this proposition is true (axiom T for Truth), and if

the agent knows a proposition, then she knows that she knows it (axiom 4, also known as the "KK-principle"or "KK-thesis"). Axiom T is often considered to be the hallmark of knowledge and has not been subjected to any serious attack. In epistemology, axiom 4 tends to be accepted by internalists, but not by externalists (Hemp, 2006) (also see (Lenzen, 1978, Chap. 4)). Axiom 4 is nevertheless widely accepted by computer scientists (but also by many philosphers, including Plato, Aristotle, Saint Augustine, Spinoza and Shopenhauer, as Hintikka recalls (1962)). A more controversial axiom for the logic of knowledge is axiom 5: This axiom states that if the agent does not know a proposition, then she knows that she does not know it. This addition of 5 to S4 yields the logic S5. Most philosophers (including Hintikka) have attacked this axiom, since numerous examples from everyday life seem to invalidate it.[3] In general, axiom 5 is invalidated when the agent has mistaken beliefs which can be due for example to misperceptions, lies or other forms of deception. Axiom D states that the agent's beliefs are consistent. In combination with axiom K (where the knowledge operator is replaced by a belief operator), axiom D is in fact equivalent to a simpler axiom D' which conveys, maybe more explicitly, the fact that the agent's beliefs cannot be inconsistent ($B\bot$): $\neg B\bot$. The other intricate axioms .2, .3, .3.2 and .4 have been introduced by epistemic logicians such as Lenzen and Kutchera in the 1970s and presented for some of them as key axioms of epistemic logic. They can be characterized in terms of intuitive interaction axioms relating knowledge and beliefs (Aucher, 2015).

In all the theories of rational agency developed in artificial intelligence, the logic of belief is KD45. Note that all these agent theories follow the perfect external approach. This is at odds with their intention to implement their theories in machines. In that respect, an internal approach seems to be more appropriate since, in this context, the agent needs to reason from its own internal point of view. For the internal approach, the logic of belief is S5, as proved by Arlo-Costa (1999) (for the notion of *full belief*) and Aucher (2010).[4]

Definition 4.2.5 (Classes of Epistemic Models). For all $\mathcal{H} \in \mathbb{L}_{\mathsf{EL}}$, the class $\mathcal{C}_{\mathcal{H}}$ or $\mathcal{C}_{\mathcal{H}^{\mathsf{C}}}$ of $\mathcal{H}$–models or $\mathcal{H}^{\mathsf{C}}$–models is the class of epistemic models whose accessibility relations satisfy the properties listed in Figure 4.3 defined by the axioms of $\mathcal{H}$ or $\mathcal{H}^{\mathsf{C}}$. □

Theorem 4.2.1 (Soundness and Completeness). *For all $\mathcal{H} \in \mathbb{L}_{\mathsf{EL}}$, $\mathcal{H}$ is sound and strongly complete for $\mathcal{L}_{\mathsf{EL}}$ w.r.t. the class of $\mathcal{H}$–models, and $\mathcal{H}^{\mathsf{C}}$ is sound and strongly complete for $\mathcal{L}^{\mathsf{C}}_{\mathsf{EL}}$ w.r.t. the class of $\mathcal{H}^{\mathsf{C}}$–models.*

[3] For example, assume that a university professor believes (is certain) that one of her colleague's seminars is on Thursday (formally Bp). She is actually wrong because it is on Tuesday ($\neg p$). Therefore, she does not know that her colleague's seminar is on Tuesday ($\neg Kp$). If we assume that axiom 5 is valid then we should conclude that she knows that she does not know that her colleague's seminar is on Tuesday ($K\neg Kp$) (and therefore she also believes that she does not know it: $B\neg Kp$). This is obviously counterintuitive.

[4] In both philosophy and computer science, there is formalization of the internal point of view. Perhaps one of the dominant formalisms for this is Moore's auto-epistemic logic (R.C.Moore, 1984, 1995). In philosophy, there are models of full belief like the one offered by Levi (1997), which is also related to ideas in auto-epistemic logic. In Aucher (2010), I provide more details on the internal approach and its connection to the other modeling approaches, namely the imperfect and the perfect external approaches.

	$n = 1$	$n \geq 2$	with common knowledge
K, S4	PSPACE	PSPACE	EXPTIME
KD45	NP	PSPACE	EXPTIME
S5	NP	PSPACE	EXPTIME

Figure 4.4: Computational complexity of the satisfiability problem

4.2.3 Decidability

All the logics introduced are decidable. We list in Figure 4.4 the complexity of the *satisfiability problem* for each of them. All these results are due to Halpern and Moses (1992). Note that if the satisfiability problem for these logics becomes linear time if there are only finitely many propositional letters in the language. For $n \geq 2$, if we restrict to finite nesting, then the satisfiability problem is NP-complete for all the modal logics considered, but S4. If we then further restrict the language to having only finitely many primitive propositions, the complexity goes down to linear time in all cases (Halpern, 1995). The computational complexity of the *model checking problem* is in P in all cases.

4.3 Updating Uncertainty: Dynamic Epistemic Logic

Dynamic Epistemic Logic (DEL) is a formalism trying to model epistemic situations involving several agents, and changes that can occur to these situations after incoming information or more generally incoming action. The methodology of DEL is such that it splits the task of representing the agents' beliefs and knowledge into three parts:

1. One represents their beliefs about an initial situation thanks to an *epistemic model*;
2. One represents their beliefs about an event taking place in this situation thanks to an *event model*;
3. One represents the way the agents update their beliefs about the situation after (or during) the occurrence of the event thanks to a *product update*.

Typically, an informative event can be a public announcement to all the agents of a formula ψ: this public announcement and correlative update constitute the dynamic part. Note that epistemic events can be much more complex than simple public announcement, including hiding information for some of the agents cheating, *etc.* This complexity is dealt with in Section 4.3.2 introducing the notion of event model. In Section 4.3.1, we will first focus on public announcements to get an intuition of the main underlying ideas that occur in DEL.

4.3.1 Public Events: Public Annoucement Logic

We start by giving a concrete example where DEL can be used, to better understand what is going on. Then, we will present a sketchy formalization of the phenomenon called *Public Announcement Logic* (PAL)

Example 4.3.1 (Muddy children). We have two children, A and B, both dirty. A can see B but not himself, and B can see A but not herself. Let p be the proposition stating that A is dirty, and q be the proposition stating that B is not dirty.

1. We represent the initial situation by the pointed epistemic model $(\mathcal{N}, s)$ represented in Figure 4.5, where relations are equivalence relations. States s, t, u, v intuitively represent possible worlds, a proposition (for example p) satisfiable at one of these states intuitively means that in the possible world corresponding to this state, the intuitive interpretation of p (p is dirty) is true. The links between states labelled by agents (A or B) intuitively express a notion of indistinguishability for the agent at stake between two possible worlds. For example, the link between s and t labelled by A intuitively means that A can not distinguish the possible world s from t and vice versa. Indeed, A can not see himself, so he cannot distinguish between a world where he is dirty and one where he is not dirty. However, he can distinguish between worlds where B is dirty or not because he can see B. With this intuitive interpretation we are brought to assume that our relations between states are equivalence relations.

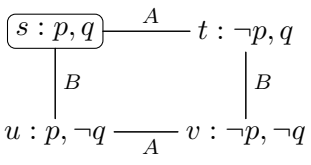


Figure 4.5: Initial pointed epistemic model $(\mathcal{N}, s)$

2. Now, suppose that their father comes and announces that at least one is dirty. Then we update the model and this yields the epistemic model of Figure 4.6. What we actually do is suppressing the worlds where the content of the announcement is not fulfilled. In our case this is the world where $\neg p$ and $\neg q$ are true. This suppression is what we call the update. We then get the model depicted. As a result of the announcement, both A and B do know that at least one of them is dirty. We can read this from the model.

3. Now suppose there is a second (and final) announcement that says that neither knows they are dirty (an announcement can express facts about the situation as well as epistemic facts about the knowledge held by the agents). We then update similarly the model by suppressing the worlds which do not satisfy the content of the announcement, or equivalently by keeping the worlds which do satisfy the

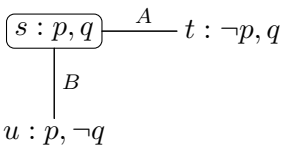


Figure 4.6: Updated epistemic model after the first announcement $p \vee q$

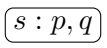


Figure 4.7: Updated epistemic model after the second announcement $(\neg K_A p \wedge \neg K_A \neg p) \wedge (\neg K_B q \wedge \neg K_B \neg q)$

announcement. This update process thus yields the pointed epistemic model represented below. By interpreting this model, we get that A and B both know that they are dirty, which seems to contradict the content of the announcement. However, if we assume that A and B are both perfect reasoners and that this is common knowledge among them, then this inference makes perfect sense. □

Public Announcement Logic (PAL) We present the syntax and semantic of Public Announcement Logic (PAL), which combines features of epistemic logic and propositional dynamic logic (Harel et al., 2000).

Definition 4.3.1 (Language $\mathcal{L}_{PAL}$). We define the language $\mathcal{L}_{PAL}$ inductively as follows:

$$\mathcal{L}_{PAL} : \varphi \quad ::= \quad p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_j \varphi \mid [\varphi!]\varphi$$

where $j \in AGTS$. The formula $\langle K_j \rangle \alpha$ is an abbreviation for $\neg\Box_j\neg\alpha$ and $\langle \psi! \rangle \varphi$ is an abbreviation for $\neg[\psi!]\neg\varphi$. The epistemic language is interpreted as in Definition 1.3.6, while the semantic clause for the new dynamic action modality is "forward looking" among models as follows:

$$\mathcal{M}, w \models [\psi!]\varphi \quad \text{iff} \quad \text{if } \mathcal{M}, w \models \psi \text{ then } \mathcal{M}^\psi, w \models \varphi$$

where $\mathcal{M}^\psi := (W^\psi, R_1^\psi, \ldots, R_n^\psi, V^\psi)$ with $W^\psi := \{w \in W : \mathcal{M}, w \models \psi\}$, $R_j^\psi := R_j \cap W^\psi \times W^\psi$ for all $j \in \{1, \ldots, n\}$ and $V^\psi(p) := V(p) \cap W^\psi$. □

The formula $[\psi!]\varphi$ intuitively means that after a truthful announcement of ψ, φ holds. The formula $\langle \psi! \rangle \varphi$ intuitively means that the announcement ψ is possible and after this announcement φ holds. A public announcement of a proposition ψ changes the current epistemic model like in Figure 4.8.

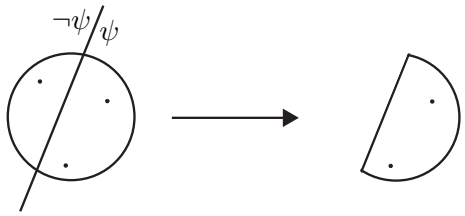


Figure 4.8: Eliminate all worlds which currently do not satisfy ψ

Theorem 4.3.1 (Soundness and Completeness). *The proof system $\mathcal{H}_{PAL}$ defined below is sound and strongly complete for $\mathcal{L}_{PAL}$ w.r.t. $\mathcal{C}_{\mathsf{ML}}$*

The Axioms and the rule of inference of $\mathcal{H}_{\mathsf{ML}}$	$(\mathcal{H}_{\mathsf{ML}})$
$[\psi!]p \leftrightarrow \psi \rightarrow p$, *for atomic facts* p	(Red1)
$[\psi!]\neg\varphi \leftrightarrow \psi \rightarrow \neg[\psi!]\varphi$	(Red2)
$[\psi!](\varphi \vee \chi) \leftrightarrow [\psi!]\varphi \wedge [\psi!]\chi$	(Red3)
$[\psi!]K_i\varphi \leftrightarrow (\psi \rightarrow K_i(\psi \rightarrow [\psi!]\varphi))$	(Red4)

Here is a typical calculation using the reduction axioms that shows that $[q!]Kq$ is a theorem of $\mathcal{H}_{PAL}$:

$[q!]Kq \leftrightarrow (q \rightarrow K(q \rightarrow [q!]q)$	(Red4)
$\leftrightarrow (q \rightarrow K(q \rightarrow (q \rightarrow q)))$	(Red1)
$\leftrightarrow (q \rightarrow K\top)$	$(\mathcal{H}_{\mathsf{ML}})$
$\leftrightarrow \top$	$(\mathcal{H}_{\mathsf{ML}})$

This states that after a public annoucement of q, the agent knows that q holds.

Example 4.3.2 (Muddy children). Here are some of the statements that hold in the muddy children puzzle formalized in PAL.

$$\mathcal{N}, s \models p \wedge q$$

'In the initial situation, A is dirty and B is dirty'.

$$\mathcal{N}, s \models (\neg K_A p \wedge \neg K_A \neg p) \wedge (\neg K_B q \wedge \neg K_B \neg q)$$

'In the initial situation, A does not know whether he is dirty and B neither'.

$$\mathcal{N}, s \models [p \vee q!](K_A(p \vee q) \wedge K_B(p \vee q))$$

'After the public announcement that at least one of the children A and B is dirty, both of then know that at least one of them is dirty'. However:

$$\mathcal{N}, s \models [p \vee q!]((\neg K_A p \wedge \neg K_A \neg p) \wedge (\neg K_B q \wedge \neg K_B \neg q))$$

'After the public announcement that at least one of the children A and B is dirty, they still do not know that they are dirty'. Moreover:

$$\mathcal{N}, s \models [p \vee q!][(\neg K_A p \wedge \neg K_A \neg p) \wedge (\neg K_B q \wedge \neg K_B \neg q)!](K_A p \wedge K_B q)$$

'After the successive public announcements that at least one of the children A and B is dirty and that they still do not know whether they are dirty, A and B then both know that they are dirty'.

In this last statement, we see at work an interesting feature of the update process: a formula is not necessarily true after being announced. That is what we technically call "self-persistence" and this problem arises for epistemic formulas (unlike propositional formulas). One must not confuse the announcement and the update induced by this announcement, which might cancel some of the information encoded in the announcement. □

PAL is decidable, its model checking problem is solvable in polynomial time and its satisfiability problem is PSPACE-complete (Lutz, 2006; Aucher and Schwarzentruber, 2013).

4.3.2 Arbitrary Events: Event Model and Product Update

In this section, we focus on items 2 and 3 of page 79, namely on how to represent events and on how to update an epistemic model with such a representation of events by means of a product update.

Representation of Events The language $\mathcal{L}_\alpha$ was introduced in (Baltag et al., 1999). The propositional letters p_ψ describing events are called *atomic events* and range over $PROP_\alpha = \{p_\psi : \psi \text{ ranges over } \mathcal{L}_{\mathsf{EL}}\}$. The reading of p_ψ is "an event of precondition ψ is occurring".

Definition 4.3.2 (Event Language $\mathcal{L}_\alpha$). We define the language $\mathcal{L}_\alpha$ inductively as follows:

$$\mathcal{L}_\alpha : \alpha \quad ::= \quad p_\psi \mid \neg\alpha \mid (\alpha \wedge \alpha) \mid K_j\alpha$$

where $\psi \in \mathcal{L}_{\mathsf{EL}}$ and $j \in AGTS$. The formula $\langle K_j\rangle\alpha$ is an abbreviation for $\neg K_j \neg\alpha$. □

A pointed event model $(\mathcal{E}, e)$ represents how the actual event represented by e is perceived by the agents. Intuitively, $f \in R_j(e)$ means that while the possible event represented by e is occurring, agent j considers possible that the possible event represented by f is actually occurring.

Definition 4.3.3 (Event Model). An *event model* is a tuple $\mathcal{E} = (W_\alpha, R_1, \ldots, R_m, I)$ where:

- W_α is a non-empty set of possible events,

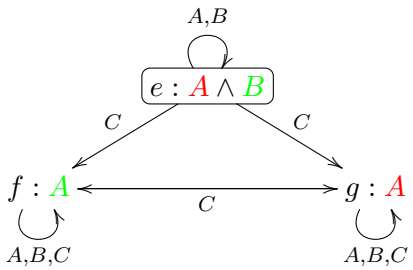


Figure 4.9: Players A and B show their cards to each other in front of player C

- $R_j \subseteq W_\alpha \times W_\alpha$ is an accessibility relation on W_α, for each $j \in AGTS$,
- $I : W_\alpha \to \mathcal{L}_{\mathsf{EL}}$ is a function assigning to each possible event a formula of $\mathcal{L}_{\mathsf{EL}}$. The function I is called the *precondition function*.

We write $e \in \mathcal{E}$ for $e \in W_\alpha$, and $(\mathcal{E}, e)$ is called a *pointed event model* (e often represents the actual event). We denote by $\mathcal{C}_\alpha$ the set of pointed event models. $R_j(e)$ denotes the set $\{f \in W_\alpha : R_j e f\}$. □

The truth conditions of the language $\mathcal{L}_\alpha$ are identical to the truth conditions of the language $\mathcal{L}_{\mathsf{EL}}$:

Definition 4.3.4 (Satisfaction Relation). Let $\mathcal{E}$ be an event model, $e \in \mathcal{E}$ and $\alpha \in \mathcal{L}_\alpha$. The *satisfaction relation* $\mathcal{E}, e \models \alpha$ is defined inductively as follows:

$$\begin{array}{lll} \mathcal{E}, e \models p_\psi & \text{iff} & I(e) = \psi \\ \mathcal{E}, e \models \neg\alpha & \text{iff} & \text{it is not the case that } \mathcal{E}, e \models \alpha \\ \mathcal{E}, e \models \alpha \wedge \beta & \text{iff} & \mathcal{E}, e \models \alpha \text{ and } \mathcal{E}, e \models \beta \\ \mathcal{E}, e \models K_j\alpha & \text{iff} & \text{for all } f \in R_j(e), \mathcal{E}, f \models \alpha \end{array}$$

□

Example 4.3.3 (Card Example). Let us resume Example 4.2.1 and assume that players A and B show their card to each other. As it turns out, C noticed that A showed her card to B but did not notice that B did so to A. Players A and B know this. This event is represented in the event model $(\mathcal{E}, e)$ of Figure 4.9. The boxed possible event e corresponds to the actual event 'players A and B show their red and green cards respectively to each other' (with precondition $A \wedge B$), f stands for the event 'player A shows her green card' (with precondition A) and g stands for the atomic event 'player A shows her red card' (with precondition A). The following statement holds in the example of Figure 4.9:

$$\begin{aligned} \mathcal{E}, e \models p_{A\wedge B} \wedge (\langle K_A\rangle p_{A\wedge B} \wedge K_A p_{A\wedge B}) \wedge (\langle K_B\rangle p_{A\wedge B} \wedge K_B p_{A\wedge B}) \\ \wedge (\langle K_C\rangle p_A \wedge \langle K_C\rangle p_A \wedge K_C\,(p_A \vee p_A)) \end{aligned} \tag{4.4}$$

It states that players A and B show their cards to each other, players A and B 'know' this and consider it possible, while player C considers possible that player A shows her

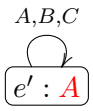


Figure 4.10: Pointed event model $(\mathcal{E}', e')$

green card and also considers possible that player A shows her red card, since he does not know her card. In fact, that is all that player C considers possible since he believes that either player A shows her red card or her green card. Another example of event model is given in Figure 4.10. This second example corresponds to the event whereby Players A shows her card publicly to everybody. The following statement holds in the example of Figure 4.10:

$$\mathcal{E}', e' \models p_A \wedge K_A p_A \wedge K_B p_A \wedge K_C p_A \wedge K_A K_A p_A \wedge K_A K_B p_A \wedge K_A K_C p_A \wedge K_B K_A p_A \\ \wedge K_B K_B p_A \wedge K_B K_C p_A \wedge K_C K_A p_A \wedge K_C K_B p_A \wedge K_C K_C p_A \wedge \dots$$

It states that player A shows her red card and that players A, B and C 'know' it, that players A, B and C 'know' that each of them 'know' it, *etc.* In other words, there is *common knowledge* among players A, B and C that player A shows her red card.

$$\mathcal{E}', e' \models p_A \wedge C p_A. \qquad \square$$

Update of the Initial Situation by the Event: Product Update The DEL product update of (Baltag et al., 1998) is defined as follows. This update yields a new $\mathcal{L}_{\mathsf{EL}}$-model $(\mathcal{M}, w) \otimes (\mathcal{E}, e)$ representing how the new situation which was previously represented by $(\mathcal{M}, w)$ is perceived by the agents after the occurrence of the event represented by $(\mathcal{E}, e)$.

Definition 4.3.5 (Product update). Let $\mathcal{M} = (W, R_1, \dots, R_m, I, w)$ be an epistemic model and let $\mathcal{E} = (W_\alpha, R_1, \dots, R_m, I, e)$ be an event model. The *product update of* $\mathcal{M}$ *and* $\mathcal{E}$ is the epistemic model $\mathcal{M} \otimes \mathcal{E} = (W^\otimes, R_1^\otimes, \dots, R_m^\otimes, I^\otimes)$ defined as follows: for all $v \in W$ and all $f \in W_\alpha$,

- $W^\otimes = \{(v, f) \in W \times W_\alpha : \mathcal{M}, v \models I(f)\}$,
- $R_j^\otimes(v, f) = \{(u, g) \in W^\otimes : u \in R_j(v) \text{ and } g \in R_j(f)\}$,
- $I^\otimes(v, f) = I(v)$. $\square$

Example 4.3.4. As a result of the event of Figure 4.9, the agents update their beliefs. We get the situation represented in the epistemic model $(\mathcal{M}, w) \otimes (\mathcal{E}, e)$ of Figure 4.11. In this $\mathcal{L}_{\mathsf{EL}}$–model, we have for example the following statement:

$$(\mathcal{M}, w) \otimes (\mathcal{E}, e) \models (B \wedge K_A B) \wedge K_C \neg K_A B.$$

It states that player A 'knows' that player B has the green card but player C believes that it is not the case. $\square$

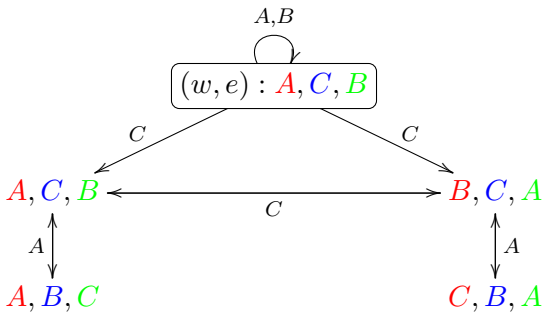


Figure 4.11: Pointed epistemic model $(\mathcal{M}, w) \otimes (\mathcal{E}, e)$

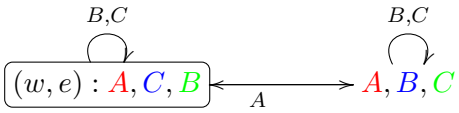


Figure 4.12: Pointed epistemic model $(\mathcal{M}, w) \otimes (\mathcal{E}', e')$

Example 4.3.5. The result of the event of Figure 4.10 whereby Ann shows her card publicly is represented in Figure 4.12. In this pointed epistemic model, the following statement holds:

$$(\mathcal{M}, w) \otimes (\mathcal{F}, e) \models C_{\{B,C\}}(A \wedge B \wedge C) \wedge \neg K_A(B \wedge C).$$

It states that there is common knowledge among B and C that they know the true state of the world (namely A has the red card, B has the green card and C has the blue card), but A does not know it. □

4.3.3 A General Language

Because of their limited perception of the surrounding world, human and artificial agents often need to reason on partial and incomplete descriptions of events and situations. For any agent, the behavior of other agents is often partially or completely unknown to them: the other agents may simply be out of sight for instance. For example, how can we be sure that an intruder does not know a certain piece of information after observing an exchange of messages in a group of agents if what we *only know* about him is that he was only able to read the messages limited to a certain vocabulary, or that he could only intercept and read the messages sent or received by a subgroup of agents? In general, we would be interested in expressing the following kind of formula: $[\alpha]\varphi$, whose intuitive reading would be "φ holds after the occurrence of an event such that what we *only know* about it is that it satisfies α". This formula α typically describes partially and incompletely the event occurring, although it could provide a full description of it as well.

Definition 4.3.6 (Language $\mathcal{L}_F$). The language $\mathcal{L}_F$ is defined inductively as follows:

$$\mathcal{L}_F : \varphi \quad ::= \quad p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_j\varphi \mid [\alpha]\varphi$$

where p ranges over $PROP$, α ranges over $\mathcal{L}_\alpha$ and j over $AGTS$. The formula $\langle\alpha\rangle\varphi$ is an abbreviation of the formula $\neg[\alpha]\neg\varphi$.

Let $(\mathcal{M}, w)$ be a pointed epistemic model. The truth conditions for the language $\mathcal{L}_F$ are defined as in Definition 1.3.6, except for the operator $[\alpha]\varphi$:

$$\mathcal{M}, w \models [\alpha]\varphi \quad \text{iff} \quad \begin{array}{l}\text{for all pointed event model } (\mathcal{E}, e) \text{ such that } \mathcal{E}, e \models \alpha, \\ \text{if } \mathcal{M}, w \models Pre(e) \text{ then } (\mathcal{M}, w) \otimes (\mathcal{E}, e) \models \varphi\end{array} \qquad \square$$

A sound and complete proof system for $\mathcal{L}_F$ can be found in (Aucher, 2012). The following proposition shows that the logic of public anouncement logic $\mathsf{L}_{PAL} := (\mathcal{L}_{PAL}, \mathcal{C}_{\mathsf{ML}}, \models)$ is at least as expressive as our general logic $\mathsf{L}_F := (\mathcal{L}_F, \mathcal{C}_{\mathsf{ML}}, \models)$, *i.e.* $\mathsf{L}_F \geq \mathsf{L}_{PAL}$ (see Definition 2.3.7).

Proposition 4.3.1. *Let $\psi \in \mathcal{L}_{\mathsf{EL}}$. Then, for all pointed epistemic models $(\mathcal{M}, w)$,*

$$\mathcal{M}, w \models [\psi!]\varphi \textit{ iff } \mathcal{M}, w \models [p_\psi \wedge Cp_\psi]\varphi$$

4.4 Further Reading

We only mention the main textbooks related to the content of this chapter. For epistemic logic, (Fagin et al., 1995) and (Meyer and van der Hoek, 1995) are the standard textbooks in computer science. See the survey of Gochet and Gribomont (2006) for a more interdisciplinary approach. Also, have a look at the seminal book of Hintikka (1962), the founder of epistemic logic. As for DEL, the reader is invited to consult the textbooks of van Ditmarsch et al. (2007) and especially van Benthem (2011). The scope of the book of van Benthem (2011) is very wide and interdisciplinary.

Finally, we stress that DEL is not the only framework dealing with communication among agents in MAS. Based on Searle's and Austin's *speech acts theory* stemming from analytic philosophy, the FIPA (Foundation for Intelligent Physical Agents) agent communication language was developed and further used in the Java Agent Development Environment (JADE, Bellifemine et al. (2007)).

Part III

Commonsense Reasoning

Introduction to Part III

"Il faut reconnaître à la pratique une logique qui n'est pas celle de la logique pour éviter de lui demander plus de logique qu'elle n'en peut donner et de se condamner soit à lui extorquer des incohérences, soit à lui imposer une cohérence forcée."

– Pierre Bourdieu, *Le sens pratique*, 1980

In everyday life, the way we represent and reason about the surrounding world and the way we revise and update our representation of the world plays an important role in our decision making process. As it turns out, the everyday life reasoning can be subtle and it requires a careful formal analysis. This has led researchers in artificial intelligence and computer science to develop logic-based theories that study and formalize belief change and the so-called "common sense reasoning". The rationale underlying the development of such theories is that it would ultimately help us understand our everyday life reasoning and the way we update our beliefs, and that the resulting work could subsequently lead to the development of tools that could be used, for example, by artificial agents to act autonomously in an uncertain and changing world. A number of theories have been proposed to capture different kinds of updates and the reasoning styles that they induce, using different formalisms and under various assumptions: default and non-monotonic logics (Makinson, 2005; Gabbay et al., 1998), belief revision theory (Gärdenfors, 1988), conditional logic (Nute and Cross, 2001).

In everyday life, two types of reasoning arise frequently: *default reasoning* and *counterfactual reasoning*. On the one hand, default reasoning involves leaping to conclusions and deals with the most 'normal' or 'typical' situations. In default reasoning, $\varphi \supset \psi$ is interpreted as 'typically or normally, if φ holds then ψ holds as well'. For example, if an agent sees a bird, she may conclude that it flies. However, not all birds fly: penguins and ostriches do not fly, nor do newborn birds, dead birds, or birds made of clay. Nevertheless, birds *typically* fly, and by default, in everyday life, we often reason with such abusive simplifications that are revised only after we receive more information. This explains informally why default reasoning is *non-monotonic*: adding new information may withdraw and invalidate some of our previous inferences. On the other hand, counterfactual reasoning involves reaching conclusions with assumptions that may be counter to fact. In legal cases it is often important to assign blame. A lawyer might well want to argue as follows: "I admit that my client was drunk and that it was raining. Nevertheless, if

the car's brakes had functioned properly, the car would not have hit Mr. Dupont. The car's manufacturer is at fault at least as much as my client".

In everyday life, other kinds of reasoning often occurs, specially in a changing environment. If we receive an incoming information which is coherent with our beliefs then we can just add it to them. But if the incoming information contradicts our beliefs then we have somehow to revise our beliefs, and as it turns out there is no obvious way to decide what should be our resulting beliefs. Solving this problem is the goal of the logic-based *belief revision theory* developed by Alchourrón, Gärdenfors and Makinson (to which we will refer by the term AGM) (Alchourrón et al., 1985; Gärdenfors, 1988; Gärdenfors and Rott, 1995). Their idea is to introduce 'rationality postulates' that specify which belief revision operations can be considered as being 'rational' or reasonable, and then to propose specific revision operations that fulfill these postulates. So, belief revision deals with the representation of mechanisms for revising our beliefs.

As we said, default reasoning, sometimes identified with *non-monotonic reasoning*, involves making default assumptions and reasoning with the most typical or "normal" situations. Even if the phenomena that are studied by default reasoning and belief revision seem to be rather different, we will see in Section 7.4 that they are in fact "two sides of the same coin", and they can be related via the so-called "Ramsey test".

In this part, we will provide a brief but concise overview of some of these logical frameworks for dealing with commonsense reasoning. Chapter 5 will deal with conditionals and counterfactuals as they have been mostly studied in philosophy. Chapter 6 will deal with default reasoning as it has been mostly studied in artificial intelligence. Chapter 7 will deal with belief revision. Default reasoning and belief revision will be related formally to each other by means of the so-called "Ramsey test" in Chapter 7.

Chapter 5

Conditionals

"If I were a swan, I'd be gone.
If I were a train, I'd be late.
And if I were a good man, I'd talk with you more often than I do.
If I were to sleep, I could dream.
If I were afraid, I could hide.
If I go insane, please don't put your wires in my brain.
If I were the moon, I'd be cool.
If I were a rule, I would bend.
If I were a good man, I'd understand the spaces between friends.
If I were alone, I would cry.
And if I were with you, I'd be home and dry.
And if I go insane, will you still let me join in with the game?
If I were a swan, I'd be gone.
If I were a train, I'd be late again.
If I were a good man, I'd talk to you more often than I do."

– Pink Floyd, *If*, 1970

5.1 Introduction

How to understand conditionals is an old issue in the history of logic. Disputes about it can be found in the Stoics and in the Middle Ages (Sanford, 2003). Generally speaking, conditionals relate some proposition (the *consequent*) to some other proposition (the *antecedent*) on which, in some sense, it depends. They are expressed in English by 'if' or cognate constructions. The grammar of conditionals imposes certain requirements on the tense (past, present, future) and mood (indicative, subjunctive) of the sentence expressing the antecedent and the consequent within it. However, not all sentences using 'if' are conditionals. Consider, for example, 'if I may say so, you have a nice ear-ring', '(Even) if he was plump, he could still run fast', or 'if you want a banana, there is one in the kitchen'. A rough and ready test for 'if φ, ψ' to be a conditional is that it can be rewritten equivalently as 'that φ implies that ψ'.

The connective $\rightarrow$ of propositional logic is usually called the *material implication* (or *material conditional*). The formula $\varphi \rightarrow \psi$ is logically equivalent to $\neg\varphi \vee \psi$ and it is true if, and only if, φ is false or ψ is true. Thus, we have $\psi \models \varphi \rightarrow \psi$ and $\neg\varphi \models \varphi \rightarrow \psi$. It is often told in a first course in logic that conditionals may be represented as the material implication $\rightarrow$. There are some obvious objections to this claim. Indeed, in that case, the following statements would be true, although they definitely appear to be false:

If Rennes is in the Netherlands then 2+2=4.

If Rennes is in France then World War II ended in 1945.

If World War II ended in 1941 then gold is an acid.

Some people argue nevertheless that these inferences are correct, because of some pragmatic rules of utterance introduced by Grice (1991) which state that we should always assert the strongest statement: in these three statements, the strongest statement is either the consequent or the negation of the antecedent (or both). Even if we accept the above argument in favor of material implication as a formalization of conditional, there are still some other arguments against it. Indeed, in the conditionals below, although the truth values of the antecedents and consequents of these conditional are the same, the intuitive truth values of the Conditional (1) is true whereas the truth value of the Conditional (2) is false.

If Shakespeare didn't write Hamlet, someone else did. (1)

If Shakespeare hadn't written Hamlet, someone else would have. (2)

Clearly, subjunctive conditionals, like statement (2), cannot be material because statement (2) is false although its antecedent is false. In response to this kind of example, philosophers distinguish between two kinds of conditionals: *subjunctive conditionals* or *counterfactuals* where the consequent is expressed using the word 'would' in the subjunctive form, and others called *indicative conditionals*. This said, sometimes when the consequent of the conditional is in the indicative *future* tense, it belongs to the category of counterfactuals.

5.2 The Problem

Formally, the syntax of conditional logics extends the syntax of propositional logic by the addition of the conditional connective $\varphi \supset \psi$ standing for "If φ, ψ":

Definition 5.2.1 (Conditional language $\mathcal{L}_{\mathsf{CL}}$). The language $\mathcal{L}_{\mathsf{CL}}$ is defined by the following grammar in Backus-Naur Form (BNF):

$$\mathcal{L}_{\mathsf{CL}}: \quad \varphi \quad ::= \quad p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid (\varphi \supset \varphi)$$

where $p \in PROP$. We use the same abbreviations as in Definition 1.3.1. To save parenthesis, we use the following ranking of binding strength: $\neg, \wedge, \vee, \supset, \rightarrow, \leftrightarrow$. □

The research problem and the challenge for conditional logics is to provide an appropriate semantics for the language $\mathcal{L}_{\mathsf{CL}}$ as well as a meaningful truth condition for the connective $\supset$. We are going to study different proposals for semantics of conditionals.

In Section 5.3, we will examine whether a truth-functional/extensional semantics can be given to conditionals. We will focus on the *material implication*. Then, since it turns out to be problematic, we will wonder in Section 5.4 whether it is possible to provide an intensional semantics directly based on the possible world semantics. We will first consider the notion of *strict conditional*. Then, since it is also problematic, we will introduce in Sections 5.5 and 5.6 other possible world semantics for conditionals based on the notions of selection functions and system of spheres. These two semantics have been proposed by Stalnaker (1968) and Lewis (1973) and are the most well-known and accepted semantics for conditional. Stalnaker claims that his definitions are both suited for indicative and subjunctive conditionals, whereas Lewis claims that his semntics is rather suited for subjunctive conditionals (counterfactuals). Finally, in Section 5.9, we will introduce six familiar conditional logics and in Section 5.10 we will provide proof systems for them (either a Tableau method or a Hilbert proof system).

5.3 Truth-functional Semantics: Material Implication

In logic and linguistic, the most common approach to specifying the meaning of a complex sentence is to specify the truth conditions of the complex sentence, in terms of the truth conditions of its parts. For example, if φ and ψ are two sentences such as "Adèle is in Rennes" and "Benoit is in Rennes". Our question will be: are the truth conditions of "If φ, ψ" of the simple, extensional, truth-functional kind, like those of "φ and ψ", "φ or ψ" and "It is not the case that φ" ? That is, do the truth values of φ and of ψ determine the truth value of "If φ, ψ"? Or are they non-truth-functional, like those of "φ because ψ", "φ before ψ", "the agent believes φ"? That is, are they such that the truth values of φ and ψ may, in some cases, leave open the truth value of "If φ, ψ" ? This kind of semantics is called *truth–functional* and is based on a *compositional* definition of well-formed formulas.

The truth-functional theory of the conditional was essential to Frege's new logic. It was taken up enthusiastically by Russell (who called it "material implication"), Wittgenstein in the Tractatus, and the logical positivists, and it is now found in every logic text. If "if" is truth-functional, this is the right truth function to assign to it: of the sixteen possible truth-functions of φ and ψ, it is the only serious candidate. Indeed, it is uncontroversial that when φ is true and ψ is false, "If φ, ψ" is false. Moreover, it is uncontroversial that "If φ, ψ" is sometimes true when φ and ψ are respectively (true, true), or (false, true), or (false, false). For example, "If it's a square, it has four sides", said of an unseen geometric figure, is true, whether the figure is a square, a rectangle or a triangle. Non-truth-functional accounts agree that "If φ, ψ" is false when φ is true and ψ is false; and they agree that the conditional is sometimes true for the other three combinations of truth-values for the components; but they deny that the conditional is always true in each of these three cases. Some agree with the truth-functionalist that

φ	ψ	$\varphi \to \psi$	$\neg\varphi \to \psi$	$\varphi \to \neg\psi$
T	T	T	T	F
T	F	F	T	T
F	T	T	T	T
F	F	T	F	T

Figure 5.1: Truth-Functional Interpretation

φ	ψ	$\varphi \supset \psi$	$\neg\varphi \supset \psi$	$\varphi \supset \neg\psi$
T	T	T	T/F	F
T	F	F	T/F	T
F	T	T/F	T	T/F
F	F	T/F	F	T/F

Figure 5.2: Non-Truth-Functional Interpretation

when φ and ψ are both true, "If φ, ψ" must be true. Some do not, demanding a further relation between the facts that φ and that ψ (see Read (1995)). In any case, all the non-truth-functionalists agree that when φ is false, "If φ, ψ" may be either true or false. For instance, if I say "If you touch that wire, you will get an electric shock" and you don't touch it, then was my remark true or false ? According to the non-truth-functionalist, it depends on whether the wire is live or dead, on whether you are insulated, *etc.* The best-known objection to the truth-functional account, one of the "paradoxes of material implication", is that the falsity of φ is sufficient for the truth of "If φ, ψ". In every possible situation in which φ is false, "$\varphi \to \psi$" is true. Can it be right that the falsity of "Rennes is in the Netherlands" entails the truth of "2+2=4"?

But even if we come to the conclusion that $\to$ does not match perfectly our natural-language "if", it comes close, and it has the virtues of simplicity and clarity. Natural language is sometimes vague and imprecise, and we cannot expect our theories to achieve better than approximate fit. Perhaps, in the interests of precision and clarity, in serious reasoning we should replace the elusive "if" with its neat, close relative, $\to$. This was no doubt Frege's attitude. Frege's primary concern was to construct a system of logic, formulated in an idealized language, which was adequate for mathematical reasoning. If "$\varphi \to \psi$" doesn't translate perfectly our natural-language "If φ, ψ", but plays its intended role, so much the worse for natural language. For the purpose of doing mathematics, Frege's judgement was probably correct. The main defects of $\to$ don't show up in mathematics. There are some peculiarities, but as long as we are aware of them, they can be lived with. And arguably, the gain in simplicity and clarity more than offsets the oddities.

The oddities are harder to tolerate when we consider conditional statements about matters dealing with everyday life. The difference is this: in everyday life, we often

accept and reject propositions with degrees of confidence less than certainty. The kind of statement "I think, but am not sure, that φ" plays no central role in mathematical thinking. In everyday life, we often use conditionals whose antecedent we think is likely to be false. We use them often, accepting some, rejecting others: "I think I won't need to get in touch, but if I do, I shall need a phone number". In fact, the way we update and infer information is quite different from the actual reasoning of mathematicians.

The definition of material implication has the unhappy consequence that all conditionals with unlikely antecedents are likely to be true. To think it likely that $\neg\varphi$ is to think it likely that a sufficient condition for the truth of $\varphi \to \psi$ obtains. Take someone who thinks that the left wing will not win the election ($\neg L$), and who rejects the thought that if they do win, they will double income tax (T). According to the definition of material implication this person has grossly inconsistent opinions. Indeed, to reject $L \to T$ is to accept $L \wedge \neg T$; for this is the only case in which $L \to T$ is false. How can someone accept $L \wedge \neg T$ yet reject L ? We would be intellectually disabled if we used mathematical reasoning in this kind of situation: we would not have the power to discriminate between believable and unbelievable conditionals whose antecedent we think is likely to be false.

5.4 Modal Semantics: Strict Conditional

Given our discussion in Section 5.3, a truth-functional/extensional semantics cannot be provided for conditionals. "We have seen that the truth value of a conditional is not always determined by the actual truth values of its antecedent and consequent, but perhaps it is determined by the truth values that its antecedent and consequent take in some other possible worlds. So, maybe we should look not only at the truth values of the antecedent and the consequent in the actual world, but also at their truth values in all possible worlds which have the same laws as does our own. When two worlds obey the same physical laws, we can say that each is a physical alternative of the other. The proposal, then, is that $\varphi \supset \psi$ is true if ψ is true at every physical alternative to the actual world at which φ is true. Suppose we say a proposition is physically necessary if and only if it is true at every physical alternative to the actual world, and suppose we express the claim that a proposition φ is physically necessary by $\Box\varphi$. Then, the proposal we are considering is that the following equivalence always holds:

$$\models_{\mathsf{S5}} (\varphi \supset \psi) \leftrightarrow \Box(\varphi \to \psi) \tag{5.1}$$

Another way of arriving at (5.1) is the following. English subjunctive conditionals are not truth-functional because they say more than that the antecedent is false or the consequent is true. The additional content is a claim that there is some sort of connection between the antecedent and the consequent. The kind of connection which seems to occur to people more readily in this context is a physical or causal connection. How can we represent this additional content in our formalization of English subjunctive conditionals ? One way is to interpret $\varphi \supset \psi$ as involving the claim that it is physically impossible that φ be true and ψ false. Once again, we come up with (5.1)." (Nute and Cross, 2001, p. 6–7)

Some More Problematic Inferences. It is easy enough to check that the following are all valid in classical logic and for strict conditionals:

$$\{\varphi \supset \psi, \psi \supset \chi\} \models \varphi \supset \chi \qquad \text{(Hypothetical Syllogism)}$$
$$\varphi \supset \psi \models \neg\psi \supset \neg\varphi \qquad \text{(Contraposition)}$$
$$\varphi \supset \psi \models (\varphi \wedge \chi) \supset \psi \qquad \text{(Monotonicity)}$$

But now consider the three following arguments of the same respective forms:

Hypothetical Syllogism: If the other candidates pull out, John will get the job. If John gets the job, the other candidates will be disappointed. Hence, if the other candidates pull out, they will be disappointed.

Contraposition: If we take the car then it will not break down en route. Hence, if the car does break down en route, we did not take it.

Monotonicity: If it does not rain tomorrow we will go to the cricket. Hence, if it does not rain tomorrow and I am killed in a car accident tonight then we will go to the cricket.

If the conditional was material, then these inferences would be valid, which they certainly do not appear to be, since they may have true premises and a false conclusion. Hence, we have a new set of objections against the conditional being material. (And since the conditionals are indicative, they tell just as much against one who claims only that indicative conditionals are material.)

5.5 Selection Functions

We introduce the selection function semantics introduced by Stalnaker (1968). It can be defined as a specific instantiation of the possible world semantics. Intuitively, $wR_\varphi v$ means that φ is true at world v, which is, *ceteris paribus*, the same as world w. *Ceteris paribus* is Latin and means 'other things being equal': the possible world v is different from the possible world w with respect to φ, all 'other things being equal'.

Definition 5.5.1 (Selection function model). A *selection function model* $\mathcal{M}$ is a tuple $\mathcal{M} := (W, \{R_\varphi : \varphi \in \mathcal{L}_{\mathsf{CL}}\}, V)$ where

- W is a non-empty set whose elements are called *possible worlds*;
- R_φ are binary relations over W;
- $V : PROP \times W \to \{T, F\}$ is a function called a *valuation*.

If $w \in W$ and $\varphi \in \mathcal{L}_{\mathsf{CL}}$, we write $wR_\varphi v$ or $R_\varphi wv$ for $(w, v) \in R_\varphi$, and $f_\varphi(w)$ denotes $\{v : v \in W \text{ and } wR_\varphi v\}$ and it is called a *selection function*. We abusively write $w \in \mathcal{M}$ for $w \in W$. The pair $(\mathcal{M}, w)$ is called a *pointed selection function model*. The class of all pointed selection function models is denoted $\mathcal{S}_{\mathsf{CL}}$. □

The intuitive interpretation of $f_\varphi(w)$ is that it is the world most like w in which φ is true or the set of worlds most like w or sufficiently like w in which φ is true.

Stalnaker (1968) suggested that we posit an absurd world λ at which every proposition is true. Alternatively, a selection function which picks a unique closest world is equivalent to a function which picks a set of worlds with at most one member. Then the empty set plays the same role as the absurd world.

Definition 5.5.2 (Satisfaction relation $\models_{\mathsf{CL}}$). The *satisfaction relation* $\models_{\mathsf{CL}} \subseteq \mathcal{S} \times \mathcal{L}_{\mathsf{CL}}$ is defined inductively as follows (we omit the subscript CL subsequently). Let $(\mathcal{M}, w) \in \mathcal{S}_{\mathsf{CL}}$ and $\varphi, \psi \in \mathcal{L}_{\mathsf{CL}}$. The truth conditions for the propositional letters and the connectives $\neg$ and $\wedge$ are the same as in Definition 1.3.3. As for the connective $\supset$, we have:

$$\mathcal{M}, w \models \varphi \supset \psi \text{ iff for all } v \in f_\varphi(w), \mathcal{M}, v \models \psi \tag{5.2}$$

Hence, the triple $\mathsf{CL} := (\mathcal{L}_{\mathsf{CL}}, \mathcal{S}_{\mathsf{CL}}, \models_{\mathsf{CL}})$ is a logic called the *basic conditional logic.* □

So, a conditional $\varphi \supset \psi$ is true in a world w just in case ψ is true in every world in $f_\varphi(w)$, *i.e.*, if ψ is true in the φ–world(s) most like or sufficiently like w. In other words, truth condition (5.2) is equivalent to the following truth condition, where for all $\varphi \in \mathcal{L}_{\mathsf{CL}}$, $[\![\varphi]\!]_{\mathcal{M}} := \{w \in W : \mathcal{M}, w \models \varphi\}$:

$$w \in [\![\varphi \supset \psi]\!]_{\mathcal{M}} \quad \text{iff} \quad f_\varphi(w) \subseteq [\![\psi]\!]_{\mathcal{M}}$$

5.6 Systems of Spheres

The founders of conditional logic (Stalnaker and Lewis) suggested that the worlds accessible to w via f_φ – that is, the worlds essentially the same as w, except that φ is true there – should be thought of as the worlds *most similar* to w at which φ is true. How to understand similarity in this context is a difficult question. It is clear, though, at least that similarity is something that comes by degrees.

A way of making the notion precise formally is as follows. We suppose that each world, w, comes with a system of 'spheres'. All the worlds in a sphere are more similar to w than any world outside that sphere. We may depict the idea as in Figure 5.3. All the worlds in S_0^w are more similar to w than the worlds in S_1^w that are not in S_0^w ($S_1^w - S_0^w$). all the worlds in S_1^w are more similar than the worlds in $S_2^w - S_1^w$, *etc.* Technically, for any world w there is a set of subsets of W, $\$_w := \{S_0^w, S_1^w, \ldots, S_n^w\}$ (for some n), such that $w \in S_0^w \subseteq S_1^w \subseteq \ldots \subseteq S_n^w = W$.

Definition 5.6.1 (System of spheres). A *system of spheres model* $\mathcal{M}$ is a tuple $\mathcal{M} := (W, \$, V)$ where

- W is a non-empty set whose elements are called *possible worlds.*
- $\$$ is a function $\$: W \to 2^{2^W}$ called a *system of spheres* that assigns to each possible world w a nested set $\$(w)$, denoted $\$_w$, of sets of worlds closed under unions and finite intersections (*i.e.* if $S, S' \in \$_w$ then $S \cap S' \in \$_w$ and $S \cup S' \in \$_w$).

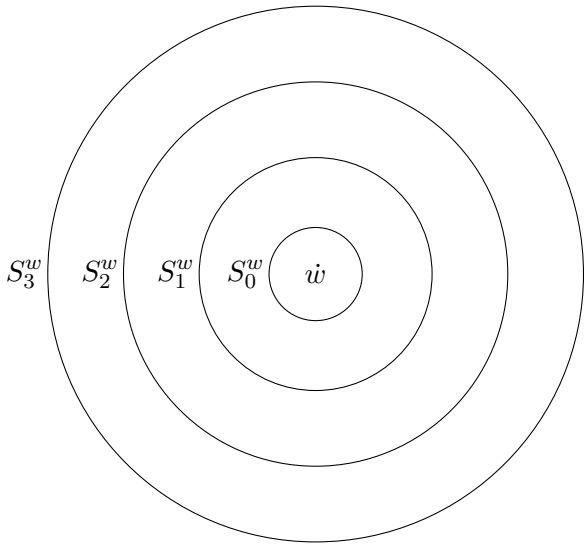


Figure 5.3: System of Spheres $\$_w$

- $V : PROP \times W \to \{T, F\}$ is a function called a *valuation*.

The pair $(\mathcal{M}, w)$ is called a *pointed system of spheres model*. The class of all pointed system of spheres models is denoted $\mathcal{V}$. □

Definition 5.6.2 (Satisfaction relation $\models_{\mathsf{V}}$). The *satisfaction relation* $\models_{\mathsf{V}} \subseteq \mathcal{V} \times \mathcal{L}_{\mathsf{CL}}$ is defined inductively as follows (we omit the subscript V subsequently). Let $(\mathcal{M}, w) \in \mathcal{V}$ and $\varphi, \psi \in \mathcal{L}_{\mathsf{CL}}$. The truth conditions for the propositional letters and the connectives $\neg$ and $\wedge$ are the same as in Definition 1.3.3. As for the connective $\supset$, we have:

$$\mathcal{M}, w \models \varphi \supset \psi \quad \text{iff} \quad \begin{array}{l} \bigcup \$_w \cap [\![\varphi]\!]_{\mathcal{M}} = \emptyset \text{ or there is } S \in \$_w \text{ such that} \\ S \cap [\![\varphi]\!]_{\mathcal{M}} \neq \emptyset \text{ and } S \subseteq [\![\varphi \to \psi]\!]_{\mathcal{M}}. \end{array}$$

where $[\![\varphi]\!]_{\mathcal{M}} = \{w \in W : \mathcal{M}, w \models \varphi\}$. Hence, the triple $\mathsf{V} := (\mathcal{L}_{\mathsf{CL}}, \mathcal{V}, \models_{\mathsf{V}})$ is a logic. □

5.7 From Systems of Spheres to Selection Functions

Given a system of sphere \$, we can define a selection function $f^{\$}$ associated to \$. The formal definition below is illustrated in Figure 5.4.

Definition 5.7.1. Let $\mathcal{M} = (W, \$, V)$ be a system of spheres model. The *selection function model* $\mathcal{M}^{\$} = (W^{\$}, \{R^{\$}_{\varphi} : \varphi \in \mathcal{L}_{\mathsf{CL}}\}, V^{\$})$ *associated to* $\mathcal{M}$ is defined as follows:

- $W^{\$} := W$ and $V^{\$} := V$;
- $f^{\$}_{\varphi}(w) := \begin{cases} [\![\varphi]\!]_{\mathcal{M}} \cap \min\limits_{\subseteq} \{S : S \in \$_w \text{ and } S \cap [\![\varphi]\!]_{\mathcal{M}} \neq \emptyset\} & \text{if } [\![\varphi]\!]_{\mathcal{M}} \neq \emptyset \\ \emptyset & \text{otherwise.} \end{cases}$

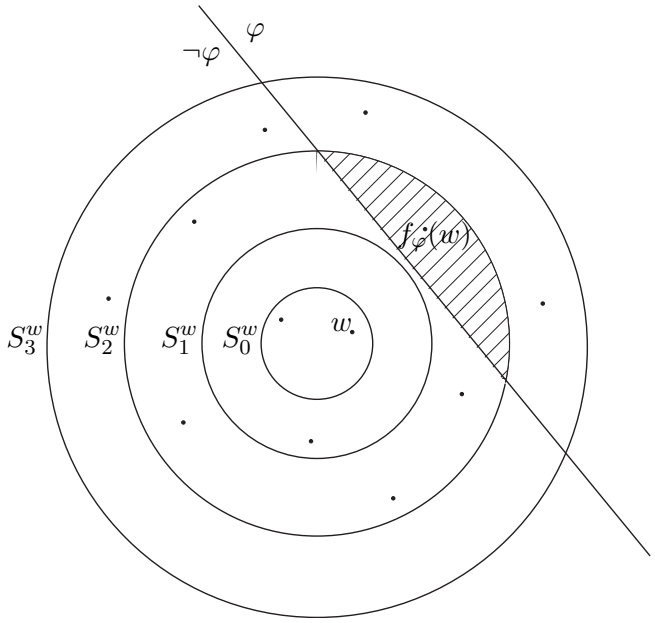


Figure 5.4: From System of Spheres $\$_w$ to Selection Function $f^{\$}_{\varphi}(w)$

We recall that we have $(w, v) \in R^{\$}_{\varphi}$ if, and only if, $v \in f^{\$}_{\varphi}(w)$. □

The sphere $\min_{\subseteq}\{S : S \in \$_w \text{ and } S \cap [\![\varphi]\!]_{\mathcal{M}} \neq \emptyset\}$ corresponds to the sphere S^w_2 in Figure 5.4. This sphere can be thought of as containing exactly those worlds at which the *ceteris paribus* clause ('all other things than φ being equal') is true.

Proposition 5.7.1. *Let $(\mathcal{M}, w)$ be a pointed system of spheres model and let $(\mathcal{M}^{\$}, w)$ be its associated pointed selection function model. Then, for all $\varphi \in \mathcal{L}_{CL}$,*

$$\mathcal{M}, w \models \varphi \quad \textit{iff} \quad \mathcal{M}^{\$}, w \models \varphi.$$

5.8 Other Semantics

Adams (1975) offers a probabilistic alternative to possible worlds semantics for conditionals. Adams assumes that conditionals have probabilities but do not have truth values. The formal language is restricted so the conditional operator $\supset$ does not occur within the scope of another conditional operator or within the scope of any truth-functional operator. The probabilistic semantics, then, only applies to *first-degree* conditionals. The probability of a conditional $\varphi \supset \psi$ is just the corresponding standard conditional probability:

$$\mu(\varphi \supset \psi) := \frac{\mu(\varphi \wedge \psi)}{\mu(\varphi)}$$

where $\mu(\varphi) \neq 0$. Adam's proposal is based on a notion of *probabilistic entailement*. A set of sentences and conditionals Γ *probabilistically entails* a conclusion φ if, and only if, for any real number $\epsilon > 0$ there is a real number $\delta > 0$ such that if $\mu(\psi) > 1 - \delta$ for each $\psi \in \Gamma$, then $\mu(\varphi) > 1 - \epsilon$. An argument is *probabilistically sound* if its premises probabilistically entail its conclusion. Adams (1975) shows that probabilistic soundness is equivalent to validity in the first-degree fragment of VC.

5.9 Some Familiar Conditional Logics

In this section, we are going to present six familiar conditional logics: CL^+, V, VW and VC (Lewis, 1973), SS (Pollock, 1976) and $\mathsf{C2}$ (Stalnaker, 1968).

5.9.1 Selection Functions Semantics

Since no constraint is imposed on the relations f_φ (*i.e.* R_φ), Stalnaker's conditional logic CL is the analogue for conditional logics of the basic modal logic ML. The following familiar constraints are often imposed. Below, $(\mathcal{M}, w)$ is any pointed selection function model, $\varphi, \psi \in \mathcal{L}_{\mathsf{CL}}$ and $[\![\varphi]\!]_{\mathcal{M}} := \{w \in W : \mathcal{M}, w \models \varphi\}$:

$$f_\varphi(w) \subseteq [\![\varphi]\!]_{\mathcal{M}} \tag{ID}$$
$$\text{If } w \in [\![\varphi]\!]_{\mathcal{M}}, \text{ then } w \in f_\varphi(w) \tag{MP}$$
$$\text{If } w \in [\![\varphi]\!]_{\mathcal{M}}, \text{ then } f_\varphi(w) = \{w\} \tag{CS}$$
$$\text{If } f_\varphi(w) = \emptyset, \text{ then } f_\psi(w) \cap [\![\varphi]\!]_{\mathcal{M}} = \emptyset \tag{MOD}$$
$$f_{\varphi \vee \psi}(w) \subseteq f_\varphi(w) \cup f_\psi(w) \tag{CA}$$
$$\text{If } f_\varphi(w) \cap [\![\psi]\!]_{\mathcal{M}} \neq \emptyset, \text{ then } f_{\varphi \wedge \psi}(w) \subseteq f_\varphi(w) \tag{CV}$$
$$\text{If } f_\varphi(w) \subseteq [\![\psi]\!]_{\mathcal{M}} \text{ and } f_\psi(w) \subseteq [\![\varphi]\!]_{\mathcal{M}}, \text{ then } f_\varphi(w) = f_\psi(w) \tag{CSO}$$
$$f_\varphi(w) \text{ is a singleton.} \tag{CEM}$$

Then, we define the five conditional logics V, VW, VC, SS and $\mathsf{C2}$. For all $\mathsf{L} \in \left\{\mathsf{CL}, \mathsf{CL}^+, \mathsf{V}, \mathsf{VW}, \mathsf{VC}, \mathsf{SS}, \mathsf{C2}\right\}$, we define the logic L as the triple $(\mathcal{L}_{\mathsf{CL}}, \mathcal{S}_{\mathsf{L}}, \models_{\mathsf{CL}})$, where $\mathcal{S}_{\mathsf{L}}$ denotes the classes of pointed conditional models as they are defined in Figure 5.5.

5.9.2 Systems of Spheres Semantics

The following familiar constraints are often imposed on systems of spheres. Below, $(\mathcal{M}, w)$ is a pointed system of sphere models, $\varphi, \psi \in \mathcal{L}_{\mathsf{CL}}$ and $[\![\varphi]\!] = \{w \in W : \mathcal{M}, w \models \varphi\}$:

$$\text{If } S \in \$_w, \text{ then } w \in S \tag{Weak Centering}$$
$$\{w\} \in \$_w \tag{Strong Centering}$$
$$\text{If } \bigcup \$_w \cap [\![\varphi]\!] \neq \emptyset, \text{ then there is a singleton } S \in \$_w \text{ such that } S \subseteq [\![\varphi]\!] \tag{CEM'}$$

Sub classes of $\mathcal{S}_{\mathsf{CL}}$	Semantic conditions
$\mathcal{S}_{\mathsf{CL}}$	None
$\mathcal{S}_{\mathsf{CL}^+}$	(ID), (MP)
$\mathcal{S}_{\mathsf{V}}$	(ID), (MOD), (CSO), (CV)
$\mathcal{S}_{\mathsf{VW}}$	(ID), (MP), (MOD), (CSO), (CV)
$\mathcal{S}_{\mathsf{VC}}$	(ID), (MP), (MOD), (CSO), (CV), (CS)
$\mathcal{S}_{\mathsf{SS}}$	(ID), (MP), (MOD), (CSO), (CA), (CS)
$\mathcal{S}_{\mathsf{C2}}$	(ID), (MP), (MOD), (CSO), (CV), (CEM)

Figure 5.5: Semantic Conditions for Selection Function Models

Sub classes of $\mathcal{V}$	Semantic conditions
$\mathcal{V}_{\mathsf{V}'}$	None
$\mathcal{V}_{\mathsf{VW}'}$	(Weak Centering)
$\mathcal{V}_{\mathsf{VC}'}$	(Strong Centering)
$\mathcal{V}_{\mathsf{C2}'}$	(CEM')

Figure 5.6: Semantic Conditions for System of Spheres Models

Then, we define the five conditional logics V', VW', VC' and C2'. For all $\mathsf{L}' \in \{\mathsf{V}', \mathsf{VW}', \mathsf{VC}', \mathsf{C2}'\}$, we define the logic L' as the triple $(\mathcal{L}_{\mathsf{CL}}, \mathcal{V}_{\mathsf{L}'}, \models_{\mathsf{CL}})$, where $\mathcal{V}_{\mathsf{L}'}$ denotes the classes of pointed conditional models defined in Figure 5.6. Note that, since Condition (CV) is validated by every system of spheres and (CV) is not a theorem of SS, there is no class of system of spheres that characterizes SS.

Proposition 5.9.1. *$L \in \{V, VW, VC, C2\}$. Then, L and L' define the same set of validities. That is, $\{\varphi \in \mathcal{L}_{CL} : \models_{L} \varphi\} = \{\varphi \in \mathcal{L}_{CL} : \models_{L'} \varphi\}$.*

5.10 Proof Systems for Conditionals

We consider two types of proof systems for our conditional logics. We provide sound and complete tableau methods for CL and CL^+ and sound and complete Hilbert proof systems for the logics $\{\mathsf{V}, \mathsf{VW}, \mathsf{VC}, \mathsf{SS}, \mathsf{C2}\}$. We will discuss the intuitive interpretation of the axioms and inference rules of the Hilbert systems.

5.10.1 Tableaux Methods

Definition 5.10.1 (Tableau rules of CL and CL^+). The tableau rules of CL are those of propositional logic of Figure 2.2 as well as $\supset$ and $\neg \supset$ below. The tableau rules of CL^+ are those of propositional logic of Figure 2.2 as well as the rules $\supset$, MP and $\neg \supset^+$

below. For rule MP, ℓ is a prefix occurring on the branch and φ is the antecedent of a conditional or negated conditional at a node on the branch.

$$\frac{(\ell\ \varphi \supset \psi) \quad (R_\varphi\ \ell\ \ell')}{(\ell'\ \psi)}\ \supset \qquad \frac{(\ell\ \neg(\varphi \supset \psi))}{\begin{array}{c}(R_\varphi\ \ell\ \ell')\\(\ell'\ \neg\psi)\end{array}}\ \neg\supset$$

$$\frac{}{\begin{array}{c|c}(\ell\ \neg\varphi) & \begin{array}{c}(\ell\ \varphi)\\(R_\varphi\ \ell\ \ell)\end{array}\end{array}}\ MP \qquad \frac{(\ell\ \neg(\varphi \supset \psi))}{\begin{array}{c}(R_\varphi\ \ell\ \ell')\\(\ell'\ \varphi)\\(\ell'\ \neg\psi)\end{array}}\ \neg\supset^{+}$$

We recall that the tableau tree for a formula is constructed as shown in Algorithm 2.2.1 of Figure 2.1. □

Theorem 5.10.1 (Soundness and completeness). *Let $\varphi \in \mathcal{L}_{\mathsf{CL}}$. Then, φ is satisfiable if, and only if, the tableau for φ is open.*

Example 5.10.1. Here is an example tableau proving $\vdash_{\mathsf{CL}} (p \supset q) \to (p \supset (q \vee r))$.

$$\begin{array}{c}(\ell\ \neg((p \supset q) \to (p \supset (q \vee r))))\\ \downarrow\\ (\ell\ (p \supset q))\\(\ell\ \neg(p \supset (q \vee r)))\\ \downarrow\\ (\ell'\ \neg(q \vee r))\\(R_p\ \ell\ \ell')\\ \downarrow\\ (\ell'\ \neg q)\\(\ell'\ \neg r)\\ \downarrow\\ (\ell'\ q)\\ \times\end{array}$$

□

5.10.2 Hilbert Systems

In order to axiomatize the validities of the logics $\{\mathsf{V}, \mathsf{VW}, \mathsf{VC}, \mathsf{SS}, \mathsf{C2}\}$, we introduce the following axioms and inference rules:

From $\varphi \leftrightarrow \psi$ infer $(\chi \supset \varphi) \leftrightarrow (\chi \supset \psi)$ (RCEC)
From $(\varphi_1 \wedge \ldots \wedge \varphi_n) \supset \psi$ infer $[(\chi \supset \varphi_1) \wedge \ldots \wedge (\chi \supset \varphi_n)] \rightarrow (\chi \supset \psi), n \geq 0$ (RCK)

$\varphi \supset \varphi$ (ID)
$(\varphi \supset \psi) \rightarrow (\varphi \rightarrow \psi)$ (MP)
$(\neg\varphi \supset \varphi) \rightarrow (\psi \supset \varphi)$ (MOD)
$[(\varphi \supset \psi) \wedge (\psi \supset \varphi)] \rightarrow [(\varphi \supset \chi) \leftrightarrow (\chi \supset \varphi)]$ (CSO)
$[(\varphi \supset \psi) \rightarrow \neg(\varphi \supset \neg\chi)] \rightarrow [(\varphi \wedge \chi) \supset \psi]$ (CV)
$(\varphi \wedge \psi) \rightarrow (\varphi \supset \psi)$ (CS)
$[(\varphi \supset \psi) \wedge (\chi \supset \psi)] \rightarrow [(\varphi \vee \chi) \supset \psi]$ (CA)
$(\varphi \supset \psi) \vee (\varphi \supset \neg\psi)$ (CEM)

Intuitive interpretation of the axioms and inference rules. (RCEC) says that we can substitute one of two provably equivalent sentences with the other in the consequent of a conditional. (MP) is so named because it supports a detachment rule for conditionals as a derived inference rule that is simlar to *Modus Ponens*: from φ and $\varphi \supset \psi$, infer ψ. (RCK) also supports a similar kind of detachment. A modal sentence $\supset \varphi$, which is read as 'φ is necessarily true', is often defined by the equivalence $\Box\varphi \leftrightarrow (\neg\varphi \supset \varphi)$. Using this defined modal operator, (MOD) becomes $\Box\varphi \rightarrow (\psi \supset \varphi)$. It reads as 'if φ is necessary, then φ would be true no matter what ψ might be'. It weakens the theorem of propositional logic $\varphi \rightarrow (\psi \rightarrow \varphi)$. (CSO) insures that conditionally (counterfactually) equivalent sentences have the same (counterfactual) consequents. (CEM) is an acronym for 'Conditional Excluded Middle' and its interpretation is clear.

Definition 5.10.2 (Proof systems). For each $\mathsf{L} \in \left\{\mathsf{CL}, \mathsf{CL}^+, \mathsf{V}, \mathsf{VW}, \mathsf{VC}, \mathsf{SS}, \mathsf{C2}\right\}$, we define the proof system $\mathcal{H}_\mathsf{L}$ by adding to the inference rules (RCEC) and (RCK) the corresponding axioms of the semantic conditions given in Figure 5.5. □

Theorem 5.10.2 (Soundness and completeness). *Let $\mathsf{L} \in \{\mathsf{CL}, \mathsf{CL}^+, \mathsf{V}, \mathsf{VW}, \mathsf{VC}, \mathsf{SS}, \mathsf{C2}\}$. Then, the proof system $\mathcal{H}_\mathsf{L}$ is sound and complete for $\mathcal{L}_{\mathsf{CL}}$ w.r.t. the class $\mathcal{S}_\mathsf{L}$ (and the class $\mathcal{V}_{\mathsf{L}'}$ if $\mathsf{L} \in \{\mathsf{V}, \mathsf{VW}, \mathsf{VC}, \mathsf{C2}\}$).*

The proof system C2 is the strongest of the five proof systems since every theorem of V, VW, VC, or SS is a theorem of C2. The axiom (CEM) is not a theorem of any of these weaker systems. VC is the next stronger system, containing all all theorems of VW and SS. (CS) is not a theorem of VW and (CV) is not a theorem of SS; thus neither VW nor SS is stronger than the other. V is weaker than VW. (MP) is generally considered a

necessary feature of any logic of commonsense conditionals, which means that V is too weak for this purpose. Lewis proposes V as a possible logic for conditional obligation. Reading $\varphi \supset \psi$ as 'If φ were the case, ψ ought to be the case', we would not want to adopt (MP) since what ought to be the case, too often is not.

Two derived inference rules valid for all five systems are of interest.

$$\text{From } \varphi \leftrightarrow \psi, \text{ infer } (\varphi \supset \chi) \leftrightarrow (\psi \supset \chi) \tag{RCEA}$$

$$\text{From } \varphi \rightarrow \psi, \text{ infer } \varphi \supset \psi \tag{RCE}$$

The next three important theses are theorems of none of these five systems. They correspond to the inferences (Hypothetical Syllogism), (Contraposition) and (Monotonicity) studied in Section 5.4.

$$[(\varphi \supset \psi) \wedge (\psi \supset \chi)] \rightarrow (\varphi \supset \chi) \tag{HS}$$

$$(\varphi \supset \psi) \rightarrow (\neg\psi \supset \neg\varphi) \tag{Contra}$$

$$(\varphi \supset \psi) \rightarrow [(\varphi \wedge \chi) \supset \psi] \tag{Mon}$$

These three theses do not appear to be related to each other at all closely. However, from the point of view of our five conditional logics they are *equivalent*: if any one of the three is added to V, VW, VC, SS or C2 as an axiom, then the other two are derivable as theorems. Counterexamples to these three theses can be found in Section 5.4.

5.11 Further Reading

This chapter is based on (Nute and Cross, 2001), (Priest, 2011, Chap. 5) and (Edgington, 2014) (note that (Edgington, 2014) is a revised chapter of (Goble, 2001)). Nute and Cross (2001) give a good survey of conditional logics. Stalnaker (1992) provides as well a short and readable survey of the philosophical issues involved in conditionals. Section 5.8 is from Nute (1994). The interface between probabilities and conditionals is still nowadays an active area of research.

Chapter 6

Default Reasoning

"If that were so, and it seems most probable, only a man who had lost his wits would have run from the house instead of towards it. If the gipsy's evidence may be taken as true, he ran with cries for help in the direction where help was least likely to be."

– Sir Arthur Conan Doyle, *The hound of the Baskerville*, 1902.

6.1 Introduction

Default reasoning deals with the kind of reasoning that we perform in everyday life when we have incomplete information about a situation. In that case, we reason assuming that we deal with the most 'normal' or most 'typical' situation. Hence, we often make inferences that are *defeasible* or *nonmonotonic*, in the sense that they can be defeated or blocked when we come to know or believe more information about the situation at stake. For example, below, Inference (6.2) defeats (6.1), and Inference (6.5) defeats (6.4) which itself defeats (6.3).

$$bird \supset fly \tag{6.1}$$
$$bird \wedge penguin \supset \neg fly \tag{6.2}$$

$$student \supset \neg taxPayer \tag{6.3}$$
$$student \wedge employed \supset taxPayer \tag{6.4}$$
$$student \wedge employed \wedge parent \supset \neg taxPayer \tag{6.5}$$

As Moore tells us:

> "By default reasoning, we mean drawing plausible inferences from less than conclusive evidence in the absence of any information to the contrary. The examples about birds being able to fly are of this type. [...] Default reasoning is nonmonotonic because, to use a term from philosophy, it is *defeasible*.

> Its conclusions are tentative, so, given better information, they may be withdrawn." (Moore, 1983, p. 273–274)

Generally speaking, in default and non-monotonic reasoning, the rule of monotonicity of propositional logic, *i.e.* from $\Gamma \models \varphi$ infer $\Gamma \cup \{\psi\} \models \varphi$, is no longer valid. Default inferences can be considered as specific kinds of conditionals. As it turns out, the general framework of plausibility measures introduced Section 3.3.1 will also allow us to recover the definition of Lewis' counterfactuals.

6.2 Logics for Defaults

We introduce a logic for reasoning about default statements. As usual in logic, it is defined in three parts: (1) the language, (2) the class of models and (3) the satisfaction relation.

Definition 6.2.1 (Language for defaults $\mathcal{L}_{\mathsf{DEF}}$). The language for defaults $\mathcal{L}_{\mathsf{DEF}}$ is defined by $\mathcal{L}_{\mathsf{DEF}} := \{\varphi \supset \psi : \varphi, \psi \in \mathcal{L}_{\mathsf{PL}}\}$. □

The formula $\varphi \supset \psi$ can be read as "if φ (is the case) then typically ψ (is the case)", "if φ, then normally ψ", "if φ, then by default ψ", and "if φ, then ψ is very likely". Thus, the default statement "birds typically fly" is represented as $bird \supset fly$. As we shall see in Section 6.5, $\mathcal{L}_{\mathsf{DEF}}$ can also be used for counterfactual reasoning, in which case $\varphi \supset \psi$ is interpreted as "if φ were true, then ψ would be true".

Example 6.2.1. Consider the following set of formulas:

$$\Gamma := \{bird \supset fly, penguin \supset \neg fly, penguin \supset bird\}$$

If $\supset$ is interpreted in Γ as the material implication $\rightarrow$ of propositional logic, then we can also derive that $penguin \supset fly$ from Γ, which is obviously counterintuitive. □

Numerous semantics have been proposed for default statements, such as preferential structures (Kraus et al., 1990), ϵ-semantics (Adams, 1975), and the possibilistic structures (Dubois and Prade, 1991) and κ-ranking (Spohn, 1988b,a) of Sections 3.2.3 and 3.2.4. All these semantics are in fact special instances of the general framework based on *plausibility measures* introduced by Friedman and Halpern (2001) and defined in Section 3.3.1. Theorem 6.3.1 will show that these alternative semantics define in fact the same set of validities.

Definition 6.2.2 (Simple structures). A *simple qualitative plausibility* (resp. *conditional probability*, *probability*, *ranking*, *possibility*, *preferential*) *structure* is a tuple $S = (W, Pl, \pi)$ where

- W is a non-empty set;
- Pl is a qualitative plausibility measure on 2^W (resp. a conditional probability measure on $2^W \times (2^W - \emptyset)$, a probability measure on 2^W, a ranking function on 2^W, a possibility measure on 2^W, a partial preorder on 2^W);

- $\pi : W \to \mathcal{C}_{\mathsf{PL}}$ is a function called the *valuation function.* □

In other words, a *simple* probability structure (resp. ranking structure, possibility structure, preferential structure) is a probability structure (resp. ranking structure, possibility structure, preferential structure) as defined in Definition 3.3.1 such that for all $w \in W$, $W_w = W$ and $\mathcal{F}_w = 2^W$.

Definition 6.2.3 (Satisfaction relation of $\mathcal{L}_{\mathsf{DEF}}$ for simple structures). Let S be a simple structure of Definition 6.2.2. Let $\varphi, \psi \in \mathcal{L}_{\mathsf{PL}}$. We define $S \models \varphi \supset \psi$ as follows:

- if $S = (W, Pl, \pi)$ is a simple qualitative plausibility structure, then

$$S \models \varphi \supset \psi \quad \text{iff} \quad \text{either } Pl([\![\psi]\!]) = \bot \text{ or } Pl([\![\varphi \wedge \psi]\!]) > Pl([\![\varphi \wedge \neg\psi]\!])$$

- if $S = (W, \mu, \pi)$ is a simple conditional probability structure, then

$$S \models \varphi \supset \psi \quad \text{iff} \quad \mu\left([\![\psi]\!] : [\![\varphi]\!]\right) = 1$$

- if $S = (W, Poss, \pi)$ is a simple possibility structure, then

$$S \models \varphi \supset \psi \quad \text{iff} \quad \text{either } Poss([\![\varphi]\!]) = 0 \text{ or } Poss([\![\varphi \wedge \psi]\!]) > Poss([\![\varphi \wedge \neg\psi]\!])$$

- if $S = (W, \kappa, \pi)$ is a simple ranking structure, then

$$S \models \varphi \supset \psi \quad \text{iff} \quad \text{either } \kappa([\![\varphi]\!]) = \infty \text{ or } \kappa([\![\varphi \wedge \psi]\!]) < \kappa([\![\varphi \wedge \neg\psi]\!])$$

- if $S = (W, \succ, \pi)$ is a simple preferential structure, then

$$S \models \varphi \supset \psi \quad \text{iff} \quad \text{either } [\![\varphi]\!] = \emptyset \text{ or } [\![\varphi \wedge \psi]\!] \succ^s [\![\varphi \wedge \neg\psi]\!]$$

where $[\![\varphi]\!] = \{w \in W : \pi(w) \models \varphi\}$. If Γ is a set of formulas of $\mathcal{L}_{\mathsf{DEF}}$ (possibly infinite), we write $S \models \Gamma$ when $S \models \varphi$ for all $\varphi \in \Gamma$. If moreover $\varphi \in \mathcal{L}_{\mathsf{DEF}}$, we write $\Gamma \models_{\mathcal{C}^{qual}} \varphi$ (resp. $\Gamma \models_{\mathcal{C}^{cond}} \varphi$, $\Gamma \models_{\mathcal{C}^{poss}} \varphi$, $\Gamma \models_{\mathcal{C}^{rank}} \varphi$, $\Gamma \models_{\mathcal{C}^{pref}} \varphi$) when for all simple qualitative plausibility (resp. conditional probability, possibility, ranking, preferential) structures S, if $S \models \Gamma$, then $S \models \varphi$. □

6.3 Proof System P

In this section, we consider an axiomatic characterization of default reasoning. There has in fact been some disagreement in the literature as to what properties $\supset$ should have. However, there seems to be some consensus on the following set of six core properties, which make up the axiom system P.

Definition 6.3.1 (System P). The proof system P for $\mathcal{L}_{\mathsf{DEF}}$ is defined by the following axiom and inference rules.

If $\vdash_{\mathsf{PL}} \varphi \leftrightarrow \varphi'$, then from $\varphi \supset \psi$ infer $\varphi' \supset \psi$	(Left Logical Equivalence, LLE)
If $\vdash_{\mathsf{PL}} \psi \rightarrow \psi'$, then from $\varphi \supset \psi$ infer $\varphi \supset \psi'$	(Right Weakening, RW)
$\varphi \supset \varphi$	(Reflexivity, REF)
From $\varphi \supset \psi_1$ and $\varphi \supset \psi_2$ infer $\varphi \supset \psi_1 \wedge \psi_2$	(AND)
From $\varphi_1 \supset \psi$ and $\varphi_2 \supset \psi$ infer $\varphi_1 \vee \varphi_2 \supset \psi$	(OR)
From $\varphi \supset \psi_1$ and $\varphi \supset \psi_2$ infer $\varphi \wedge \psi_2 \supset \psi_1$.	(Cautious Monotony, CM)

If $\Gamma \subseteq \mathcal{L}_{\mathsf{DEF}}$ and $\gamma \in \mathcal{L}_{\mathsf{DEF}}$, then we write $\Gamma \vdash_{\mathsf{P}} \gamma$ when there is a proof of γ from Γ in P (see Definition 1.4.1 for more details). □

The first three properties (Left Logical Equivalence, LLE), (Right Weakening, RW) and (Reflexivity, REF) of P seem noncontroversial. If φ and φ' are logically equivalent, then surely if ψ follows by default from φ, then it should also follow by default from φ'. Similarly, if ψ follows from φ by default, and ψ logically implies ψ', then surely ψ' should follow from φ by default as well. Finally, reflexivity just says that φ follows from itself.

The latter three properties get more into the heart of default reasoning. The (AND) rule says that defaults are closed under conjunction. For example, if an agent sees a bird, she may want to conclude that it flies. She may also want to conclude that it has wings. The (AND) rule allows her to put these conclusions together and conclude that, by default, birds both fly and have wings.

The (OR) rule corresponds to reasoning by cases. If red birds typically fly ($(red \wedge bird) \supset fly$) and nonred birds typically fly ($(\neg red \wedge bird) \supset fly$), then birds typically fly, no matter what color they are ($bird \supset fly$). Note that the (OR) rule actually gives only $(red \wedge bird) \vee (\neg red \wedge bird)) \supset fly$ here. The conclusion $bird \supset fly$ requires (Left Logical Equivalence, LLE), using the fact that $bird \leftrightarrow ((red \wedge bird) \vee (\neg red \wedge bird))$.

To understand (Cautious Monotony, CM), note that one of the most important properties of the material conditional is that it is *monotonic*. Getting extra information never results in conclusions being withdrawn. For example, if $\varphi \rightarrow \psi$ is true under some interpretation, then so is $\varphi \wedge \varphi' \rightarrow \psi$, no matter what φ' is. On the other hand, default reasoning is not always monotonic. From $bird \supset fly$ it does not follow that $bird \wedge penguin \supset fly$. Discovering that a bird is a penguin should cause the retraction of the conclusion that it flies. (Cautious Monotony, CM) captures one instance when monotonicity seems reasonable. If both ψ_1 and ψ_2 follow from φ by default, then discovering ψ_2 should not cause the retraction of ψ_1. For example, if birds typically fly and birds typically have wings, then it seems reasonable to conclude that birds that have wings typically fly.

All the propertis of P hold if $\supset$ is interpreted as the material implication. However, this interpretation leads to unwarranted conclusions, as Example 6.2.1 shows.

Theorem 6.3.1 (Soundness and completeness). *The proof system* P *is sound and complete for* $\mathcal{L}_{DEF}$ *w.r.t.* $\mathcal{C}^{qual}, \mathcal{C}^{cond}, \mathcal{C}^{poss}$, $\mathcal{C}^{rank}$ *and* $\mathcal{C}^{pref}$. *That is, if* Γ *is a finite set of formulas in* $\mathcal{L}_{DEF}$ *and* $\varphi \supset \psi \in \mathcal{L}_{DEF}$, *then*

$$\begin{array}{rll} \Gamma \vdash_{\mathsf{P}} \varphi \supset \psi & \textit{iff} & \Gamma \models_{\mathcal{C}^{qual}} \varphi \supset \psi \\ & \textit{iff} & \Gamma \models_{\mathcal{C}^{cond}} \varphi \supset \psi \\ & \textit{iff} & \Gamma \models_{\mathcal{C}^{poss}} \varphi \supset \psi \\ & \textit{iff} & \Gamma \models_{\mathcal{C}^{rank}} \varphi \supset \psi \\ & \textit{iff} & \Gamma \models_{\mathcal{C}^{pref}} \varphi \supset \psi \end{array}$$

Theorem 6.3.1 tells us that the conditional probabilistic, possibilistic, ranking and preferential semantics all have in common that they define the same set of validities axiomatized by the same proof system P (originally introduced by Kraus et al. (1990)). This remarkable fact is explained by Friedman and Halpern (2001); Halpern (2003). The proof system P is sound as long as the semantics satisfies (Pl4) and (Pl5). The proof system P is complete as long as the semantics is *rich*, *i.e.* for all $\varphi_1, \ldots, \varphi_k$ with $k > 1$ of pairwise mutually exclusive and satisfiable propositional formulas, there is a structure of the semantics such that

$$Pl([\![\varphi_1]\!]) > Pl([\![\varphi_2]\!]) > \ldots > Pl([\![\varphi_k]\!]) = \bot \qquad \text{(Rich)}$$

6.4 From Defaults to Conditionals

$\mathcal{L}_{\mathsf{DEF}}$ is a rather weak language. For example, although it can express the fact that a certain default holds, it cannot express the fact that a certain default does *not* hold, since $\mathcal{L}_{\mathsf{DEF}}$ does not allow negated default. There is no great difficulty extending the language to allow negated and nested default. This yields the language $\mathcal{L}_{\mathsf{CL}}$ of Definition 5.2.1. Its semantics is defined on the class of preferential structures as follows.

Definition 6.4.1 (Satisfaction relation of $\mathcal{L}_{\mathsf{CL}}$ for $\mathcal{S}^{pref}$, $\mathcal{S}^{qual}$, $\mathcal{S}^{rank}$ and $\mathcal{S}^{poss}$). The *satisfaction relation* $\models \subseteq \mathcal{S}^{pref} \times \mathcal{L}_{\mathsf{CL}}$ is defined inductively as follows. Let $(S, w) \in \mathcal{S}^{pref}$ and $\varphi, \psi \in \mathcal{L}_{\mathsf{CL}}$.

$$\begin{array}{lll} S, w \models p & \text{iff} & \pi(w)(p) = T \\ S, w \models \neg\varphi & \text{iff} & \text{it is not the case that } S, w \models \varphi \\ S, w \models \varphi \wedge \psi & \text{iff} & S, w \models \varphi \text{ and } S, w \models \psi \\ S, w \models \varphi \supset \psi & \text{iff} & \text{either } [\![\varphi]\!] = \emptyset \text{ or } [\![\varphi \wedge \psi]\!] \succeq^s_w [\![\varphi \wedge \neg\psi]\!] \end{array}$$

where $[\![\varphi]\!] := \{w \in W_w : S, w \models \varphi\}$. Similar definitions can be provided for $\mathcal{S}^{qual}$, $\mathcal{S}^{rank}$ and $\mathcal{S}^{poss}$: it suffices to index Pl, κ and $Poss$ by w in the clauses for the truth conditions of $\varphi \supset \psi$ of Definition 6.2.2. □

It should be clear from the definitions that formulas in $\mathcal{L}_{\mathsf{CL}}$ can be expressed in $\mathcal{L}_{\mathsf{Qual}}$ of Section 3.3.3. In fact the corresponding logics are equally expressive:

Proposition 6.4.1. *For all* $X \in \{qual, pref, rank, poss\}$, *the logics* $(\mathcal{L}_{CL}, \mathcal{S}^X, \models)$ *and* $(\mathcal{L}_{Qual}, \mathcal{S}^X, \models)$ *are equally expressive.*

Proof. The result follows from Propositions 8.6.1 and 8.6.2 of Halpern (2003). □

Definition 6.4.2 (System Cond). The Hilbert proof system Cond for $\mathcal{L}_{\mathsf{CL}}$ is defined by the following axiom and inference rules.

$$\begin{aligned}
&\text{All axioms and inference rules of } \mathcal{H}_{\mathsf{PL}} && \text{(Prop)}\\
&\varphi \supset \varphi && \text{(C1)}\\
&((\varphi \supset \psi_1) \wedge (\varphi \supset \psi_2)) \rightarrow (\varphi \supset (\psi_1 \wedge \psi_2)) && \text{(C2)}\\
&((\varphi_1 \supset \psi) \wedge (\varphi_2 \supset \psi)) \rightarrow ((\varphi_1 \vee \varphi_2) \supset \psi) && \text{(C3)}\\
&((\varphi \supset \psi_1) \wedge (\varphi \supset \psi_2)) \rightarrow ((\varphi \wedge \psi_2) \supset \psi_1) && \text{(C4)}\\
&\text{From } \varphi \leftrightarrow \varphi' \text{ infer } (\varphi \supset \psi) \rightarrow (\varphi' \supset \psi) && \text{(RC1)}\\
&\text{From } \psi \rightarrow \psi' \text{ infer } (\varphi \supset \psi) \rightarrow (\varphi \supset \psi') && \text{(RC2)}
\end{aligned}$$

□

The proof system Cond can be viewed as a generalization of system P. For example, the richer language allows the (AND) to be replaced by the axiom (C2). Similarly, (C1), (C3), (C4), (RC1) and (RC2) are the analogues of (Reflexivity, REF), (OR), (Cautious Monotony, CM), (Left Logical Equivalence, LLE) and (Right Weakening, RW) respectively.

Theorem 6.4.1 (Soundness and completeness). *The proof system Cond is sound and complete for $\mathcal{L}_{\mathsf{CL}}$ w.r.t. $\mathcal{S}^{qual}$ and $\mathcal{S}^{pref}$.*

6.5 From Conditionals to Counterfactuals

The language $\mathcal{L}_{\mathsf{CL}}$ can be used to reason about conditionals and counterfactuals as well as defaults. In Chapter 5, it was employed to reason about conditionals and counterfactuals whereas in this chapter it was used for default reasoning. The general framework based on plausibility measures is more general than the selection functions and systems of spheres of the previous Chapter 5. So, it is natural to wonder which constraints need to be added to this framework to recover at least the logic for counterfactuals based on the sphere semantics or the selection functions. That is what we will investigate in this section.

Definition 6.5.1. A *pointed counterfactual preferential* (resp. *ranking*, *plausibility*) *structure* is a pointed preferential (resp. ranking, plausibility) structure $S = (W, \mathcal{X}, \pi, w)$ (as defined in Definition 3.3.1) that satisfies the following condition:

$$\begin{aligned}
&w \in W_w \text{ and } w \succ_w v \text{ for all } v \in W_w \text{ such that } v \neq w && \text{(Counterfact}^{\succ}\text{)}\\
&w \in W_w \text{ and } \kappa_w(w) < \kappa_w(W_w - \{w\}) && \text{(Counterfact}^{\kappa}\text{)}\\
&w \in W_w \text{ and } Pl_w(w) > Pl_w(W_w - \{w\}) && \text{(Counterfact}^{Pl}\text{)}
\end{aligned}$$

where $u \succ_w v$ is an abbreviation for $u \succeq_w v$ and $v \not\succeq_w u$. □

"Note that in a counterfactual preferential structure, W_w is not the set of worlds the agent considers possible. W_w in general includes worlds that the agent knows perfectly well to be impossible. For example, suppose that in the actual world w the lawyer's client was drunk and it was raining. The lawyer wants to make the case that, even if his client hadn't been drunk and it had been sunny, the car would have hit the cow. [...] Thus, to evaluate the lawyer's claim, the worlds $w \in W_w$ that are closest to w where it is sunny and the client is sober and driving his car must be considered. But these are worlds that are currently known to be impossible. This means that the interpretation of W_w in preferential structures depends on whether the structure is used for default reasoning or counterfactual reasoning." (Halpern, 2003, p. 316)

In a pointed counterfactual preferential structure, the preorder $\succ_w$ represents a notion of similarity between possible worlds. We state $w_1 \succ_w w_2$ when w_1 is at least as close to w as w_2 (or at least as similar to w as w_2). With this interpretation in mind, w should be at least as close to itself than any other world: that is the intuitive interpretation of conditions (Counterfact$^{\succ}$), (Counterfact$^{\kappa}$) and (Counterfact^{Pl}). The additional property that corresponds to these conditions is:

$$\varphi \to (\psi \leftrightarrow (\varphi \supset \psi)) \qquad \text{(Counterfact)}$$

Axiom (Counterfact) reads as 'if φ is already true in the actual world, then the counterfactual $\varphi \supset \psi$ is true if, and only if, ψ is also true in the actual world'.

Theorem 6.5.1 (Soundness and completeness). *The proof system* $\mathsf{CFL} := \mathsf{Cond} + \{Counterfact\}$ *is sound and complete for the language* $\mathcal{L}_{\mathsf{CL}}$ *w.r.t. the class of pointed counterfactual preferential (as well as ranking and plausibility) structures.*

The truth conditions for $\varphi \supset \psi$ in the preferential, ranking and plausibility semantics roughly tells us that $\varphi \wedge \psi$ is more likely than $\varphi \wedge \neg\psi$. However, this does not seem to accord very much with the truth conditions for conditionals that we saw in Chapter 5, based on selection functions or systems of spheres. We are going to show now that an equivalent formulation can be found for the preferential semantics which is in fact very close to the semantics for counterfactuals based on selection functions proposed by Stalnaker (1968). If (S, w) is a pointed counterfactual preference structure, we define:

$$f_{\varphi}(w) := \{v \in [\![\varphi]\!] : \text{for all } u \in [\![\varphi]\!] - \{v\}, u \not\succeq_w v\} \tag{6.6}$$

where $[\![\varphi]\!] := \{w \in W_w : S, w \models \varphi\}$.

The notation $f_{\varphi}(w)$ has to be interpreted as 'the worlds in $[\![\varphi]\!]$ the most similar to w'. Under a 'default' reading, it should be read as 'the most normal/typical worlds in $[\![\varphi]\!]$'. The notation $f_{\varphi}(w)$ reminds very much the selection function introduced by Stalnaker for providing a semantics to counterfactuals. In fact, this choice of notation is meaningful from an intuitive point of view, since we have the following result:

Proposition 6.5.1. *Let* (S, w) *be a pointed preferential structure and let* $\varphi, \psi \in \mathcal{L}_{\mathsf{CL}}$. *Then,*

$$S, w \models \varphi \supset \psi \quad \textit{iff} \quad f_{\varphi}(w) \subseteq [\![\psi]\!]_S \tag{6.7}$$

where $[\![\psi]\!]_S := \{w \in S : S, w \models \psi\}$.

Hence, we can also provide a semantics for counterfactuals based on preferential, ranking and plausibility structures. It remains to investigate which specific conditions of Figure 5.5 the selection function of Expression (6.6) verifies in order to determine exactly to which conditional/counterfactual logic of the previous chapter corresponds CFL.

6.6 Further reading

This chapter is based on Chapter 8 of (Halpern, 2003). I also recommend the book of Priest (2011) for an overview on non-classical logics (the proof systems considered in this book are all tableau systems), and the book edited by Goble (2001) which introduces in a concise way various sub-areas of philosophical logic. In artificial intelligence, other very different kinds of formalisms have been introduced to deal with default and nonmonotonic reasoning, such as circumscription and Reiter's default logic. We refer the reader for example to Goble (2001) or Gabbay et al. (1998) for more information.

Chapter 7

Belief Revision

"To attain knowledge, add things every day, To attain wisdom, remove things every day."

– Lao Tzu, Tao-te Ching, ch. 48

7.1 Introduction

Originally, belief revision theory was developed by Alchouron, Gärdenfors and Makinson (AGM) with a strong syntactic stance. In AGM belief revision theory, the beliefs of the agent are represented by a *belief set* $\mathcal{K}$, *i.e.* a set of propositional formulas that is closed under logical consequence. These propositional formulas represent the beliefs of the agent. AGM distinguishes three types of belief change: expansion, revision and contraction. The expansion of $\mathcal{K}$ with a propositional formula φ, written $\mathcal{K}+\varphi$, consists of adding φ to $\mathcal{K}$ and taking all the logical consequences. Note that this might yield inconsistency. The revision of $\mathcal{K}$ with φ, written $\mathcal{K}*\varphi$, consists of adding φ to $\mathcal{K}$, but in order that the resulting set be consistent, some formulas are removed from $\mathcal{K}$. Finally, the contraction of $\mathcal{K}$ with φ, written here $\mathcal{K} \stackrel{\circ}{=} \varphi$, consists in removing φ from $\mathcal{K}$, but in order that the resulting set be consistent, some other formulas are also removed. Of course there are some connections between these operations. From a contraction operation, one can define a revision operation thanks to the Levi identity:

$$\mathcal{K} * \varphi := (\mathcal{K} \stackrel{\circ}{=} \neg\varphi) + \varphi.$$

And from a revision operation, one can define a contraction operation thanks to the Harper identity:

$$\mathcal{K} \stackrel{\circ}{=} \varphi := \mathcal{K} \cap (\mathcal{K} * \neg\varphi).$$

In this chapter, we will focus on the revision and the expansion operation. Assumption 1 stated at the beginning of Section 3.4 is therefore no longer valid.

$\mathcal{K} + \varphi$ is a belief set	($\mathcal{K}$+1)
$\varphi \in \mathcal{K} + \varphi$	($\mathcal{K}$+2)
$\mathcal{K} \subseteq \mathcal{K} + \varphi$	($\mathcal{K}$+3)
If $\varphi \in \mathcal{K}$ then $\mathcal{K} = \mathcal{K} + \varphi$	($\mathcal{K}$+4)
$\mathcal{K} + \varphi$ is the smallest set satisfying ($\mathcal{K}$+1)–($\mathcal{K}$+4).	($\mathcal{K}$+5)

Figure 7.1: AGM belief *expansion* postulates

7.2 Expansion

In this section, we assume that the set of propositional letters $PROP$ is finite, and in this paragraph, all the formulas belong to the propositional language $\mathcal{L}_{\mathsf{PL}}$ defined over $PROP$. Let $Cn(.)$ be the classical consequence operation, i.e. for a set of propositional formulas Γ, $Cn(\Gamma) := \{\varphi : \Gamma \vdash_{\mathsf{PL}} \varphi\}$. We can now define formally a belief set.

Definition 7.2.1 (Belief set). A *belief set* $\mathcal{K}$ is a set of propositional formulas of $\mathcal{L}_{\mathsf{PL}}$ such that $Cn(\mathcal{K}) = \mathcal{K}$. We denote by $\mathcal{K}_\perp$ the unique inconsistent belief set consisting of all propositional formulas. □

Classically, in AGM theory, we start by proposing rationality postulates that belief change operations must fulfill. These postulates make precise our intuitions about these operations and what we mean by rational change. The rationality postulates for the expansion operation + proposed by Gärdenfors (1988) are given in Figure 7.1.

($\mathcal{K}$+1) tells us that the expansion operation + is a function from pairs of belief set and formula to belief sets. This entails that we can iterate the expansion operation. ($\mathcal{K}$+2) tells us that when the agent expands her belief set by φ then as a result φ is one of her beliefs. All the other postulates refer to some kind of minimal change. ($\mathcal{K}$+3) tells us that when the agent expands by φ she does not throw away any of her former beliefs. ($\mathcal{K}$+4) tells us that if the agent already believes φ then expanding by φ should not change her beliefs: the change made to add φ to the belief set is minimal.

The following (representation) theorem tells us that these postulates actually determine a *unique* expansion operation on belief sets.

Theorem 7.2.1. *A function + satisfies ($\mathcal{K}$+1)–($\mathcal{K}$+5) if, and only if, for each belief set $\mathcal{K}$ and formula $\varphi \in \mathcal{L}_{PL}$, $\mathcal{K} + \varphi = Cn(\mathcal{K} \cup \{\varphi\})$.*

So from now on, we define the expansion operation + by $\mathcal{K} + \varphi = Cn(\mathcal{K} \cup \{\varphi\})$. So far our approach to expansion was syntactically driven. Now we are going to give a semantical approach to expansion and set some links between these two approaches.

We use the possible world semantics. First we consider the set $\mathcal{W}$ consisting of all the (logically) possible worlds. A possible world w can be viewed as an interpretation,

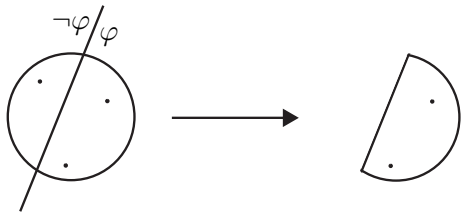


Figure 7.2: AGM expansion by φ

i.e. a function from $PROP$ to $\{\top, \bot\}$ which specifies which propositional letters (such as 'it is raining') are true in this world w (see Definition 1.3.2). For a propositional formula φ, we write $w \models \varphi$ when φ is true at w in the usual sense (See Definition 1.3.3) Then a formula φ is true in a set W of possible worlds, written $W \models \varphi$, if, and only if, for all $w \in W$, $w \models \varphi$. Besides, because $PROP$ is finite, $\mathcal{W}$ is also finite. We can then represent the agent's epistemic state by a subset W of $\mathcal{W}$ (which is consequently finite as well). Intuitively, W is the smallest set of possible worlds in which the agent believes that the actual world is located.

There is actually a very close correspondence between belief sets and sets of possible worlds.

Definition 7.2.2 (Belief set associated to a set of possible worlds). Let W be a finite set of possible worlds. We define the *belief set* $\mathcal{K}_W$ *associated to* W by $\mathcal{K}_W = \{\varphi \in \mathcal{L}_{\mathsf{PL}} : W \models \varphi\}$. Let $\mathcal{K}$ be a belief set. We define the set of possible worlds $W_\mathcal{K}$ associated to $\mathcal{K}$ by $W_\mathcal{K} = \{w \in W : w \models \varphi \text{ for all } \varphi \in \mathcal{K}\}$. Then,

$$W \models \varphi \text{ iff } \varphi \in \mathcal{K}_W \qquad\qquad \varphi \in \mathcal{K} \text{ iff } W_\mathcal{K} \models \varphi. \qquad \square$$

We define the semantic counterpart of the expansion operation defined previously.

Definition 7.2.3 (Expansion). Let W be a finite set of possible worlds and $\varphi \in \mathcal{L}_{\mathsf{PL}}$. The *expansion* of W by φ, written $W + \varphi$, is defined as follows.

$$W + \varphi = \{w \in W : w \models \varphi\}. \qquad \square$$

This semantic counterpart of the expansion is described graphically in Figure 7.2. The initial model W is on the left of the arrow and the expanded model $W + \varphi$ is on the right of the arrow. The dots represent possible worlds and the straight line separates the worlds satisfying φ from the worlds satisfying $\neg\varphi$.

Finally, we show that these two definitions of expansion, syntactic and semantic, are in fact equivalent.

Theorem 7.2.2. *For all belief sets* $\mathcal{K}$ *and all finite sets of possible worlds* W*,*

$$\psi \in \mathcal{K} + \varphi \textit{ iff } W_\mathcal{K} + \varphi \models \psi \qquad\qquad W + \varphi \models \psi \textit{ iff } \psi \in \mathcal{K}_W + \varphi.$$

$\mathcal{K} * \varphi$ is a belief set	$(\mathcal{K} * 1)$
$\varphi \in \mathcal{K} * \varphi$	$(\mathcal{K} * 2)$
$\mathcal{K} * \varphi \subseteq \mathcal{K} + \varphi$	$(\mathcal{K} * 3)$
If $\neg\varphi \notin \mathcal{K}$ then $\mathcal{K} + \varphi \subseteq \mathcal{K} * \varphi$	$(\mathcal{K} * 4)$
$\mathcal{K} * \varphi = \mathcal{K}_\perp$ iff φ is unsatisfiable	$(\mathcal{K} * 5)$
If $\vdash_{\mathsf{PL}} \varphi \leftrightarrow \varphi'$ then $\mathcal{K} * \varphi = \mathcal{K} * \varphi'$	$(\mathcal{K} * 6)$
$\mathcal{K} * (\varphi \wedge \varphi') \subseteq (\mathcal{K} * \varphi) + \varphi'$	$(\mathcal{K} * 7)$
If $\neg\varphi' \notin \mathcal{K} * \varphi$ then $(\mathcal{K} * \varphi) + \varphi' \subseteq \mathcal{K} * (\varphi \wedge \varphi')$	$(\mathcal{K} * 8)$

Figure 7.3: AGM belief *revision* postulates

7.3 Revision

In this section, all formulas are propositional formulas. Just as for expansion, Gärdenfors and his colleagues proposed rationality postulates for revision operations. These postulates make precise what we mean by rational change, and more precisely rational revision. They are given in Figure 7.3. We will not provide intuitive motivations for these postulates (even if some of them have been criticized), see (Gärdenfors, 1988) for details. However, note that these postulates do not characterize a unique revision operation, unlike the postulates for expansion.

Before going on, let us reconsider how we represent the agent's epistemic state. So far we have proposed two equivalent formalisms: belief set and (finite) set of possible worlds. As we said, a belief set is an infinite set of formulas closed under logical consequence. However, this cannot be handled easily by computers because of its infinitude. We would like to have a more compact and finite representation of the agent's epistemic state. For that, we follow the approach of Katsuno and Mendelzon (1992).

As argued by Katsuno and Mendelzon, because $PROP$ is finite, a belief set $\mathcal{K}$ can be equivalently represented by a mere propositional formula ψ. This formula is also called a belief base. Then $\varphi \in \mathcal{K}$ if and only if $\varphi \in Cn(\psi)$. Besides, one can easily show that $\chi \in \mathcal{K} + \varphi$ if and only if $\chi \in Cn(\psi \wedge \varphi)$. So in this approach, the expansion of the belief base ψ by φ is the belief base $\psi \wedge \varphi$, which is possibly an inconsistent formula. Now, given a belief base ψ and a formula φ, $\psi \circ \varphi$ denotes the revision of ψ by φ. But in this case, $\psi \circ \varphi$ is supposed to be consistent if φ is. Given a revision operation $*$ on belief sets, one can define a corresponding operation $\circ$ on belief bases as follows: $\vdash_{\mathsf{PL}} \psi \circ \varphi \rightarrow \chi$ if, and only if, $\chi \in Cn(\psi) * \varphi$. Thanks to this correspondence, Katsuno and Mendelzon set some rationality postulates for this revision operation $\circ$ on belief bases which are equivalent to the AGM rationality postulates for the revision operation $*$ on belief sets.

Lemma 7.3.1. *Let * be a revision operation on belief sets and $\circ$ its corresponding operation on belief bases. Then * satisfies the 8 AGM postulates $(\mathcal{K} * 1)$–$(\mathcal{K} * 8)$ if, and*

only if, $\circ$ *satisfies the postulates* (R1)–(R6) *below:*

$$\vdash_{PL} \psi \circ \varphi \rightarrow \varphi \tag{R1}$$

if $\psi \wedge \varphi$ *is satisfiable, then* $\vdash_{PL} \psi \circ \varphi \leftrightarrow \psi \wedge \varphi$ (R2)

If φ *is satisfiable, then* $\psi \circ \varphi$ *is also satisfiable* (R3)

If $\vdash_{PL} \psi \leftrightarrow \psi'$ *and* $\vdash_{PL} \varphi \leftrightarrow \varphi'$*, then* $\vdash_{PL} \psi \circ \varphi \leftrightarrow \psi' \circ \varphi'$ (R4)

$$\vdash_{PL} (\psi \circ \varphi) \wedge \varphi' \rightarrow \psi \circ (\varphi \wedge \varphi') \tag{R5}$$

If $(\psi \circ \varphi) \wedge \varphi'$ *is satisfiable, then* $\vdash_{PL} \psi \circ (\varphi \wedge \varphi') \rightarrow (\psi \circ \varphi) \wedge \varphi'$ (R6)

So far our approach to revision was syntactically driven. Now we are going to give a semantical approach to revision and then set some links between the two approaches.

Notation 7.3.1. $[\![\psi]\!]$ denotes the set of all logically possible worlds (also called models in that case) that make ψ true, i.e. $[\![\psi]\!] = \{w \in \mathcal{W} : w \models \psi\}$.

Definition 7.3.1 (Faithful assignment). A *pre-order* $\leq$ over $\mathcal{W}$ is a reflexive and transitive relation on $\mathcal{W}$. A preorder is *total* if for every $w, w' \in \mathcal{W}$, either $w \leq w'$ or $w' \leq w$. Consider a function that assigns to each propositional formula ψ a preorder $\leq_\psi$ over $\mathcal{W}$. We say this assignment is *faithful* if the following three conditions hold:

1. If $w, w' \in [\![\psi]\!]$, then $w <_\psi w'$ does not hold;
2. If $w \in [\![\psi]\!]$ and $w' \notin [\![\psi]\!]$, then $w <_\psi w'$ holds;
3. If $\vdash_{PL} \psi \leftrightarrow \psi'$, then $\leq_\psi = \leq_{\psi'}$. □

Intuitively, $w \leq_\psi w'$ means that the possible world w is closer to ψ than w'.

Definition 7.3.2. Let $\mathcal{M}$ be a subset of $\mathcal{W}$. A possible world w is *minimal* in $\mathcal{M}$ with respect to $\leq_\psi$ if $w \in \mathcal{M}$ and there is no $w' \in \mathcal{M}$ such that $w' <_\psi w$. Let

$$min(\mathcal{M}, \leq_\psi) = \{w : w \text{ is minimal in } \mathcal{M} \text{ with respect to } \leq_\psi\}$$ □

The following representation theorem sets some connections between the semantic approach and the syntactic one.

Theorem 7.3.1. *Revision operation* $\circ$ *satisfies postulates* (R1)–(R6) *iff there exists a faithful assignment that maps each belief base* ψ *to a total preorder* $\leq_\psi$ *such that*

$$[\![\psi \circ \varphi]\!] = min([\![\varphi]\!], \leq_\psi).$$

This semantic revision process is described in Figure 7.4. In this figure, the dots represent possible worlds and the diagonal line separates the worlds satisfying φ from the worlds satisfying $\neg\varphi$. The worlds in the inner circle are the worlds that satisfy ψ and thus correspond to $[\![\psi]\!]$. The other circles represent the ordering $\leq_\psi$: if $w <_\psi w'$ then w is within a smaller circle than w' and if $w =_\psi w'$ then w and w' are in between the same successive circles. So the further a world is from the inner circle, the further it is from ψ. The worlds in the hatched part are then the worlds that satisfy φ and which are the closest to ψ. Therefore they represent $[\![\psi \circ \varphi]\!] = min([\![\varphi]\!], \leq_\psi)$.

Grove (1988) proposed another semantic approach based on a system of spheres. But one can show that his framework can be recast in the one just described.

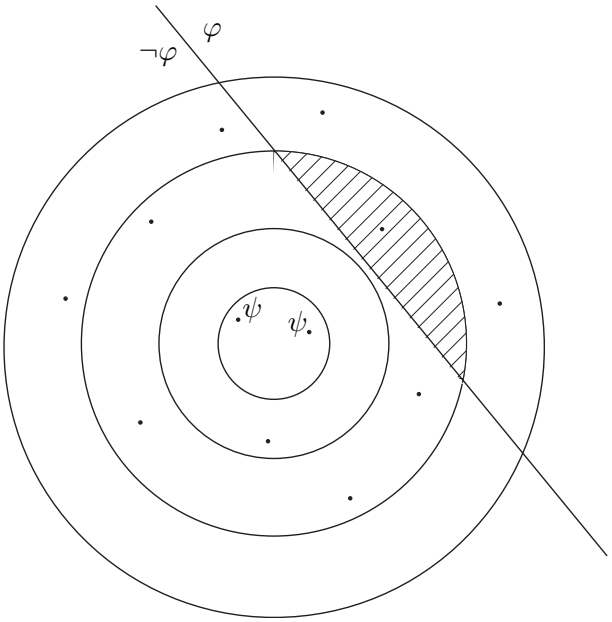


Figure 7.4: AGM belief revision by φ

7.4 "Two Sides of the Same Coin"

A well-known result connects closely non-monotonic reasoning (default reasoning) with belief revision. This lead Gardenförs and Makinson to claim that they are "two sides of the same coin" (Gärdenfors, 1991; Makinson and Gärdenfors, 1989). But in fact, this connection goes back to a footnote of Ramsey (1929) who introduced the so-called "Ramsey test" for providing semantics to conditionals.[1]

Theorem 7.4.1.

- *Suppose that a revision operation $\circ$ satisfies $(\mathcal{K}*1)$ - $(\mathcal{K}*8)$. Fix a belief set $\mathcal{K}$, and define a relation $\supset$ on propositional formulas by taking $\varphi \supset \psi$ to hold iff $\psi \in \mathcal{K}*\varphi$. Then, $\supset$ satisfies all the properties of P as well as Rational Monotonicity:*

  $$\text{if } \varphi \supset \psi_1 \text{ and not } \varphi \supset \neg\psi_2, \text{ then } \varphi \wedge \psi_2 \supset \psi_1 \qquad \text{(Rational Monotonicity)}$$

  Moreover, $\varphi \supset \bot$ if, and only if, φ is not satisfiable.

[1] Here is Ramsey's footnote: "If two people are arguing 'If p, then q?' and are both in doubt as to p, they are adding p hypothetically to their stock of knowledge and arguing on that basis about q; so that in a sense 'If p, q' and 'If p, $\neg q$' are contradictories. We can say that they are fixing their degree of belief in q given p. If p turns out false, these degrees of belief are rendered void. If either party believes not p for certain, the question ceases to mean anything to him except as a question about what follows from certain laws or hypotheses."(Ramsey, 1929, 154–155)

- *Conversely, suppose that $\supset$ is a relation on formulas that satisfies the properties of P and Rational Monotonicity, and $\varphi \supset \bot$ if, and only if, φ is not satisfiable. Let $\mathcal{K} = \{\psi \in \mathcal{L}_{PL} : \top \supset \psi\}$. Then, $\mathcal{K}$ is a belief set. Moreover, if $*$ is defined by taking $\mathcal{K} * \varphi = \{\psi : \varphi \supset \psi\}$, then $(\mathcal{K} * 1)$–$(\mathcal{K} * 8)$ hold for $\mathcal{K}$ and $*$.*

7.5 Further Reading

This Chapter is based on the book of Gärdenfors (1988). The Chapter of Gärdenfors and Rott (1995) is still fine for an introductory entry. Also, see the survey article of Fermé and Hansson (2011) for an overview of the work in belief revision theory in the last 25 years and for pointers to the literature.

Part IV

Appendix

Chapter A

Set Theory: Basic Notions and Notations

In this appendix, we recall some basic notions and notations of set theory that are used throughout the lecture notes. Set theory was developed by mathematicians to be able to talk about collections of objects. It has turned out to be an invaluable tool for defining some of the more complicated mathematical structures. As such, we use in these lecture notes some very basic set theory, mostly as a convenient notation for some of our constructions. This appendix is a series of definitions and examples, with some informal explanations.

A.1 Sets and Elements

A *set* is just a collection of mathematical objects. These objects can be numbers, letters, tuples or even other sets. For example, $\mathbb{R}$ is the set of all real numbers, $\mathbb{N}$ is the set of all natural number 0, 1, 2,... and $\mathbb{Z}$ is the set of all integers (both positive and negative) ..., -2, -1, 0, 1, 2,...

The objects in these collections are called *elements*. Given a set A and an object x, we use the notation $x \in A$ to denote that x is an element of A, and $x \notin A$ to denote that x is not an element of A. When $x \in A$, we say that x *is an element of* A or that A *contains* x. For example, $0,33 \in \mathbb{R}$ but $0,33 \notin \mathbb{N}$.

A.2 Defining Sets

We have three basic ways of writing down a set. We can list all of the elements, write down some defining property for its elements, or write its elements as the values taken by some expression. When we *define* a set, we often use the notation := instead of =.

A.2.1 Lists of Elements

The most basic way to define a set is as a list of its elements placed between curly braces. So $\{1, 2, 3\}$ is the set whose only elements are 1, 2, and 3. The set $\{1, 2, 3, \ldots, 100\}$ has elements consisting of all integers from 1 to 100, and $\{1, 2, 3, \ldots\}$ has elements consisting

of all of the positive integers. One particular set of interest, is the empty set. This is the set with no elements and it is denoted $\{\}$ or $\emptyset$.

A.2.2 Elements Satisfying a Condition

To express more complicated sets, we can take the set of all elements of some other set B satisfying some property P, written as $A := \{x \in B : P(x)\}$. This is the set of all elements $x \in B$ so that the property $P(x)$ holds. The above set A would contain exactly the elements y so that $y \in B$ and $P(y)$ holds. For example,

$$\{x \in \mathbb{R} : x > 1\}$$

is the set of all real numbers bigger than 1.

Intervals are specific sets of real numbers which can easily be defined by a property. If $a, b \in \mathbb{R}$ with $a \leq b$, then we define:

$$\begin{aligned}
[a; b] &:= \{x \in \mathbb{R} : a \leq x \leq b\} \\
[a; b[&:= \{x \in \mathbb{R} : a \leq x < b\} \\
]a; b] &:= \{x \in \mathbb{R} : a < x \leq b\} \\
]a; b[&:= \{x \in \mathbb{R} : a < x < b\}
\end{aligned}$$

A.2.3 Elements of a Given Form

Suppose that we want to express the set of all even numbers. That is the set of all integers n so that n is equal to $2m$ for some other integer m. We could of course write this as

$$\{n \in \mathbb{Z} : \text{there exists } m \in \mathbb{Z} \text{ so that } n = 2m\}.$$

On the other hand, this notation is somewhat cumbersome. It is often useful to have a way to write the set of all objects that can be produced in some way. Thus, for the set of even integers we use the alternative notation:

$$\{2m : m \in Z\}.$$

In general we use the notation

$$\{f(x, y, z) : x \in A, y \in B, z \in C\}$$

to denote the set of all things that can be written as $f(x, y, z)$ for some elements x, y, z of the sets A, B, C, respectively. So for example,

$$\{n^2 + m^2 : n, m \in \mathbb{Z}\}$$

is the set of all numbers that can be written as the sum of the squares of two integers, and

$$\{\{n, x\} : n \in \mathbb{Z}, x \in \mathbb{R}\}$$

is the set of sets that contain as elements one integer and one real number.

A.3 Equality and Subsets

Two sets are considered to be equal if they contain exactly the same elements. In other words two sets, A and B, are equal if and only if the elements of one are exactly the same as the elements of the other. So A equals B if for all x, $x \in A$ if, and only if, $x \in B$. This definition has some important consequences. In particular, it means that the sets $\{1, 2, 3\}$, $\{3, 1, 2\}$, and $\{1, 1, 2, 2, 2, 3\}$ each contain only the elements 1, 2, and 3 and are thus all the same set: $\{1, 2, 3\} = \{3, 1, 2\} = \{1, 1, 2, 2, 2, 3\}$.

If one cares about the order and multiplicity of elements, one will often consider tuples instead of sets. A *tuple* is a sequence of elements for which both the order and the multiplicities matter. A tuple is written as a list of elements between parentheses. For example $(1, 2, 3)$, which is distinct from $(3, 1, 2)$.

Another useful concept is that of subsets. We say that A is a *subset* of B, denoted $A \subseteq B$, if every element of A is also an element of B. So, for example, $\{1, 3\}$ is a subset of $\{1, 2, 3\}$ (formally, $\{1, 3\} \subseteq \{1, 2, 3\}$), but $\{1, 2, 5\}$ is not because $5 \notin \{1, 2, 3\}$. If W is a set, we denote by 2^W or $\mathcal{P}(W)$ the set of all subsets of W.

A.4 Operations: Union, Intersection, Difference

Another important way to create sets is to define them in terms of simpler sets. Here are a few simple operations that can be used to do this.

- Given two sets A and B the *union*, denoted $A \cup B$ is the set of all elements contained in either set. Namely,

$$A \cup B := \{x : x \in A \text{ or } x \in B\}.$$

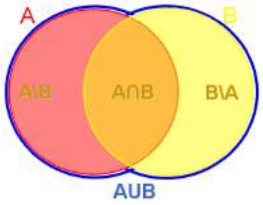
- Given two sets A and B the *intersection*, denoted $A \cap B$ is the set of all elements contained in both sets. Namely,

$$A \cap B := \{x : x \in A \text{ and } x \in B\}.$$

- Given two sets A and B the *difference* of A and B, denoted $A \backslash B$ is the set of elements that are in A but not B. Namely

$$A \backslash B := \{x : x \in A \text{ and } x \notin B\}.$$

  If W is a given set, the *complementation* of A (in W), denoted $\overline{A}$, is $W \backslash A$.



Intersection and union are *commutative* and *associative*: for all sets A, B, C,

$$A \cap B = B \cap A \qquad A \cup B = B \cup A$$
$$A \cap (B \cap C) = (A \cap B) \cap C \qquad A \cup (B \cup C) = (A \cup B) \cup C$$

Intersection and union are *distributive* unto each other: for all sets A, B, C,

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \qquad (A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \qquad (A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$

Complementation is *idempotent* and obeys to the *De Morgan laws*: for all sets A, B,

$$\overline{\overline{A}} = A$$
$$\overline{A \cap B} = \overline{A} \cup \overline{B} \qquad \overline{A \cup B} = \overline{A} \cap \overline{B}$$

If $A_1, \ldots, A_n$ are n sets (with $n \geq 2$) then $\bigcup_{i=1}^{n} A_i := A_1 \cup \ldots \cup A_n$ and $\bigcap_{i=1}^{n} A_i := A_1 \cap \ldots \cap A_n$.

A *partition* of A is a set of subsets $\{A_1, \ldots, A_n\}$ of A such that:

1. $\bigcup_{i=1}^{n} A_i = A$;
2. $A_i \cap A_j = \emptyset$ for all $i, j \in \{1, \ldots, n\}$ such that $i \neq j$.

For example, $\{\{1, 2\}, \{3, 4, 5\}, \{6\}\}$ is a partition of $\{1, 2, 3, 4, 5, 6\}$.

A.5 Functions

A *function* (sometimes also called a *mapping*) is a rule for assigning to each element x in a set A, another element $f(x)$ in some other set B. This f would be a function taking or 'mapping' elements of A to elements of B denoted $f : A \to B$. In this case, A is called the *domain* of f and B is called the *codomain*. When not all elements of A are mapped to elements of B, f is called a *partial function*; otherwise, f is (sometimes) called a *total function*.

- The function f is *surjective* or *onto* if every element of the codomain can be written as some value of f. Or, in other words, if for every $y \in B$ there exists an $x \in A$ so that $f(x) = y$.
- The function f is *injective* or *one-to-one* if distinct elements of the domain get mapped to distinct elements of the codomain. In other words, f is injective when $f(x) = f(y)$ only when $x = y$.
- If f is both injective and surjective, it is *bijective*. This means that for each element of B that there is one and only one element of A that f maps to it. In other words, f is bijective if for every $y \in B$ there is a unique $x \in A$ so that $f(x) = y$.

A.6 Relations and Cartesian Product

If A and B are two sets, the *cartesian product* of A and B, denoted $A \times B$, is the set of all tuples (a, b) where a and b range over A and B respectively. Formally:

$$A \times B := \{(a, b) : a \in A, b \in B\}.$$

For example, $\{1, 2, 3\} \times \{a, b\} := \{(1, a), (1, b), (2, a), (2, b), (3, a), (3, b)\}$. A *binary relation* R over a set A is a subset of the cartesian product $A \times A$. For example, if $A := \{w, v, u\}$, then $R := \{(w, w), (w, v), (v, u)\}$ is a binary relation over A. If R is a binary relation over a set A, we write wRv or Rwv for $(w, v) \in R$, and $R(w)$ denotes $\{v \in A : wRv\}$.

We can extend these notions to an arbitrary arity. If $A_1, \ldots, A_n$ are n sets (with $n \geq 2$), then the *cartesian product* of $A_1, \ldots, A_n$, denoted $A_1 \times \ldots \times A_n$, is the set of all tuples $(a_1, \ldots, a_n)$ where $a_1, \ldots, a_n$ range over $A_1, \ldots, A_n$ respectively. Formally:

$$A_1 \times \ldots \times A_n := \{(a_1, \ldots, a_n) : a_1, \ldots a_n \in A\}.$$

A *n-ary relation* R over a set A is a subset of the cartesian product $\underbrace{A \times \ldots \times A}_{\text{n times}}$. The Cartesian product $\underbrace{A \times \ldots \times A}_{\text{n times}}$ is also denoted A^n. For example, $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$.

A.7 Further Reading

Almost every mathematical textbook recalls the basics of set theory and its usual notations. We refer the reader to textbooks on discrete mathematics, such as (Conradie and Goranko, 2015; Rosen, 2012), which usually contain more informal explanations than the others on this topic.

Bibliography

Abiteboul, S., Hull, R., and Vianu, V. (1995). *Foundations of databases*, volume 8. Addison-Wesley. 14

Abramsky, S., Gabbay, D., and Maibaum, T. (1992). *Handbook of Logic in Computer Science*. Oxford University Press, Oxford, England. 12

Adams, E. (1975). *The Logic of Conditionals*, volume 86 of *Synthese Library*. Springer. 101, 102, 108

Alchourrón, C. E., Gärdenfors, P., and Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *J. Symb. Log.*, 50(2):510–530. 92

Arlo-Costa, H. (1999). Qualitative and probabilistic models of full belief. In Buss, S., P.Hajek, and Pudlak, P., editors, *Logic Colloquium'98*, Lecture Notes on Logic 13. 78

Aubenque, P. (2012). *Encyclopædia Universalis [online]*, chapter Syllogisme. Encyclopædia Universalis S.A., `http://www.universalis-edu.com/encyclopedie/syllogisme/`. 7, 10

Aucher, G. (2010). An internal version of epistemic logic. *Studia Logica*, 1:1–22. 78

Aucher, G. (2012). DEL-sequents for regression and epistemic planning. *Journal of Applied Non-Classical Logics*, 22(4):337–367. 87

Aucher, G. (2015). Intricate axioms as interaction axioms. *Studia Logica*, pages 1–28. 78

Aucher, G. and Schwarzentruber, F. (2013). On the complexity of dynamic epistemic logic. *CoRR*, abs/1310.6406. 83

Aumann, R. (1976). Agreeing to disagree. *The annals of statistics*, 4(6):1236–1239. 12, 73

Baier, C. and Katoen, J. (2008). *Principles of model checking*. MIT press. 14

Baltag, A. and Moss, L. S. (2004). Logic for epistemic programs. *Synthese*, 139(2):165–224. 70

Baltag, A., Moss, L. S., and Solecki, S. (1998). The logic of public announcements and common knowledge and private suspicions. In Gilboa, I., editor, *TARK*, pages 43–56. Morgan Kaufmann. 46, 85

Baltag, A., Moss, L. S., and Solecki, S. (1999). The logic of public announcements, common knowledge and private suspicions. Technical report, Indiana University. 83

Battigalli, P. and Bonanno, G. (1999). Recent results on belief, knowledge and the epistemic foundations of game theory. *Research in Economics*, 53:149–225. 71, 72

Bellifemine, F. L., Caire, G., and Greenwood, D. (2007). *Developing multi-agent systems with JADE*, volume 7. John Wiley & Sons. 87

Ben-Ari, M. (2012). *Mathematical Logic for Computer Science*. Springer. 12, 28, 42

Bibel, W. and Eder, E. (1993). Methods and calculi for deduction. In Gabbay, D. M., Hogger, C. J., and Robinson, J. A., editors, *Handbook of Logic in Artificial Intelligence and Logic Programming (Vol. 1)*, pages 68–182. Oxford University Press, Inc., New York, NY, USA. 42

Blackburn, P., de Rijke, M., and Venema, Y. (2001). *Modal Logic*, volume 53 of *Cambridge Tracts in Computer Science*. Cambridge University Press. 26, 28, 40, 42

Boh, I. (1993). *Epistemic Logic in the later Middle Ages*. Routledge, London. 72

Börger, E., Grädel, E., and Gurevich, Y. (2001). *The classical decision problem*. Springer. 31

Conradie, W. and Goranko, V. (2015). *Logic and Discrete Mathematics - A Concise Introduction*. Wiley. 129

Cover, T. and Thomas, J. (1991). *Elements of Information Theory*. Wiley Series in Telecommunications. John Wiley & Sons Inc. 46

Dubois, D. and Prade, H. (1991). Possibilistic logic, preferential model and related issue. In *Proceedings of the 12th International Conference on Artificial Intelligence (IJCAI)*, pages 419–425. Morgan Kaufman. 108

Edgington, D. (2014). Indicative conditionals. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Winter 2014 edition. 4, 106

Enderton, H. B. (1972). *An Introduction to Mathematical Logic*. Academic Press. 28, 42

Fagin, R., Halpern, J., Moses, Y., and Vardi, M. (1995). *Reasoning about knowledge*. MIT Press. 45, 71, 72, 73, 87

Fermé, E. and Hansson, S. O. (2011). AGM 25 years. *Journal of Philosophical Logic*, 40(2):295–331. 121

Fitting, M. (1993). Basic modal logic. In Gabbay, D. M., Hogger, C. J., and Robinson, J. A., editors, *Handbook of Logic in Artificial Intelligence and Logic Programming (Vol. 1)*, pages 368–448. Oxford University Press, Inc., New York, NY, USA. 42

Frege, G. (1879). Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought. *From Frege to Gödel: A source book in mathematical logic*, 1931:1–82. 1

Friedman, N. and Halpern, J. Y. (2001). Plausibility measures and default reasoning. *Journal of the ACM*, 48(4):648–685. 47, 56, 108, 111

Gabbay, D. and Guenthner, F. (2001). *Handbook of philosophical logic*. Springer, second edition. 14, 26

Gabbay, D. and Robinson, J. (1998). *Handbook of Logic in Artificial Intelligence and Logic Programming: Volume 5: Logic Programming*, volume 5. Clarendon Press. 12, 14, 30

Gabbay, D. M., Hogger, C. J., Robinson, J. A., Siekmann, J., and Nute, D., editors (1998). *Handbook of logic in artificial intelligence and logic programming*, volume Nonmonotonic reasoning and uncertain reasoning (Volume 3). Clarendon Press. 2, 12, 91, 114

Gabbay, D. M. and Smets, P. (1998). *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, volume 1. Springer. 68

Gärdenfors, P. (1988). *Knowledge in Flux (Modeling the Dynamics of Epistemic States)*. Bradford/MIT Press, Cambridge, Massachusetts. 2, 12, 91, 92, 116, 118, 121

Gärdenfors, P. (1991). Belief revision and nonmonotonic logic: Two sides of the same coin? In *Logics in AI*, pages 52–54. Springer. 120

Gärdenfors, P. and Rott, H. (1995). *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume Volume 4, Epistemic and temporal reasoning, chapter Belief Revision, pages 35–132. Clarendon Press, Oxford. 92, 121

Gentzen, G. (1935). Untersuchungen über das logische schließen. i. *Mathematische zeitschrift*, 39(1):176–210. 1, 12

Goble, L. (2001). *The Blackwell guide to philosophical logic*. Wiley-Blackwell. 3, 26, 106, 114

Gochet, P. and Gribomont, P. (2006). Epistemic logic. In Gabbay, D. and Woods, J., editors, *Handbook of the History of Logic*, volume 7, Twentieth Century Modalities, pages 99–195. Elsevier, Amsterdam. 87

Grice, H. P. (1991). *Studies in the Way of Words*. Harvard University Press. 94

Grove, A. (1988). Two modellings for theory change. *Journal of Philosophical Logic*, 17:157–170. 119

Halpern, J. (2003). *Reasoning about Uncertainty*. MIT Press, Cambridge, Massachussetts. 3, 48, 59, 60, 61, 66, 68, 111, 112, 113, 114

Halpern, J., Harper, R., Immerman, N., Kolaitis, P., Vardi, M., and Vianu, V. (2001). On the unusual effectiveness of logic in computer science. *The Bulletin of Symbolic Logic*, 7(2):213–236. 12, 14

Halpern, J. and Moses, Y. (1992). A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54:311–379. 79

Halpern, J. Y. (1995). The effect of bounding the number of primitive propositions and the depth of nesting on the complexity of modal logic. *Artificial Intelligence*, 75(2):361 – 372. 79

Harel, D., Kozen, D., and Tiuryn, J. (2000). *Dynamic Logic*. MIT Press. 81

Hemp, D. (2006). The KK (knowing that one knows) principle. *The Internet Encyclopedia of Philosophy, http://www.iep.utm.edu/kk-princ/print/*. 78

Hintikka, J. (1962). *Knowledge and Belief, An Introduction to the Logic of the Two Notions*. Cornell University Press, Ithaca and London. 45, 72, 78, 87

Hintikka, J. and Sandu, G. (2007). What is logic ? In Jacquette, D., editor, *Philosophy of Logic*, pages 13 – 40. North Holland. 7

Hoare, C. (1969). An axiomatic basis for computer programming. *Communications of the ACM*, 12(10):567–580. 14

Hodges, W. (1997). *A Shorter Model Theory*. Cambridge University Press, New York, NY, USA. 40, 42

Huth, M. and Ryan, M. (2004). *Logic in Computer Science: Modelling and reasoning about systems*. Cambridge University Press. 28

Immerman, N. (1999). *Descriptive complexity*. Springer Verlag. 14

Jacquette, D. (2007). *Philosophy of logic*. North Holland. 28

Kant, E. (1800). *Logik*. Jäsche, G. B., Königsberg. 7

Katsuno, H. and Mendelzon, A. (1992). Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52(3):263–294. 118

Kleene, S. C., de Bruijn, N., de Groot, J., and Zaanen, A. C. (1971). *Introduction to metamathematics*. Wolters-Noordhoff Groningen. 28

Kraus, S., Lehmann, D. J., and Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44(1-2):167–207. 108, 111

Leeuwen, J. (1990). *Handbook of theoretical computer science*. Elsevier Science. 14

Lenzen, W. (1978). *Recent Work in Epistemic Logic*. Acta Philosophica 30. North Holland Publishing Company. 72, 77, 78

Levi, I. (1997). *The covenant of reason: rationality and the commitments of thought*, chapter The Logic of Full Belief, pages 40–69. Cambridge University Press. 78

Lewis, D. (1969). *Convention, a Philosophical Study*. Harvard University Press. 73

Lewis, D. K. (1973). *Counterfactuals*. Blackwell Publishers. 95, 102

Li, M. and Vitányi, P. (1993). *An introduction to Kolmogorov complexity and its applications*. Graduate texts in computer science. Springer-Verlag, Berlin, second edition. 46

Lutz, C. (2006). Complexity and succinctness of public announcement logic. In Nakashima, H., Wellman, M. P., Weiss, G., and Stone, P., editors, *AAMAS*, pages 137–143. ACM. 83

Makinson, D. (2005). *Bridges from classical to nonmonotonic logic*. King's College. 2, 12, 91

Makinson, D. and Gärdenfors, P. (1989). Relations between the logic of theory change and nonmonotonic logic. In Fuhrmann, A. and Morreau, M., editors, *The Logic of Theory Change*, volume 465 of *Lecture Notes in Computer Science*, pages 185–205. Springer. 120

Manna, Z. and Waldinger, R. (1985). *The logical basis for computer programming*. Addison-Wesley Professional. 12

Meyer, J.-J. C. and van der Hoek, W. (1995). *Epistemic Logic for AI and Computer Science*. Cambridge University Press, Cambridge. 45, 72, 87

Moore, R. C. (1983). Semantical considerations on nonmonotonic logic. In Bundy, A., editor, *Proceedings of the 8th International Joint Conference on Artificial Intelligence. Karlsruhe, FRG, August 1983*, pages 272–279. William Kaufmann. 108

Nagel, T. (1986). *The view from nowhere*. oxford university press. 71

Nute, D. (1994). *Handbook of logic in artificial intelligence and logic programming*. Oxford University Press, Inc., New York, NY. 106

Nute, D. and Cross, C. B. (2001). Conditional logic. In Gabbay, D. and Guenthner, F., editors, *Handbook of philosophical logic*, volume 4, pages 1–98. Kluwer Academic Pub. 2, 12, 91, 97, 106

Papadimitriou, C. (2003). *Computational complexity*. John Wiley and Sons Ltd. 14, 42

Pollock, J. L. (1976). *Subjunctive reasoning*. Reidel Dordrecht. 102

Pratt, V. (1976). Semantical considerations on floyd-hoare logic. In *Proceedings of the 17th IEEE Symposium on the Foundations of Computer Science*, pages 109–121. 12

Priest, G. (2011). *An introduction to non-classical logic*. Cambridge University Press. 3, 26, 106, 114

Ramsey, F. P. (1929). General propositions and causality. In Mellor, H., editor, *Philosophical Papers*. Cambridge University Press, Cambridge. 120

R.C.Moore (1984). Possible-world semantics for autoepistemic logic. In *Proceedings of the Non-Monotonic Reasoning Workshop*, pages 344–354, New Paltz NY. 78

R.C.Moore (1995). *Logic and Representation*. CSLI Lecture Notes. 78

Read, S. (1995). *Thinking about logic*. Oxford University Press Oxford. 28, 96

Restall, G. (2000). *An Introduction to Substructural Logics*. Routledge. 26, 27

Rosen, K. (2012). *Discrete Mathematics and its Applications*. McGraw-Hill. 129

Sanford, D. H. (2003). *If P, then Q: Conditionals and the Foundations of Reasoning*. Psychology Press. 93

Sipser, M. (2006). *Introduction to the Theory of Computation*, volume 2. Thomson Course Technology Boston. 42

Smullyan, R. M. (1968). *First-order logic*, volume 6. Springer. 28

Spohn, W. (1988a). A general non-probability theory of inductive reasoning. In Schachter, R., Levitt, T., Kanal, L., and Lemmer, J., editors, *Uncertainty in Artificial Intelligence 4*, pages 149–158. North-Holland. 108

Spohn, W. (1988b). Ordinal conditional functions: A dynamic theory of epistemic states. In Harper, W. L. and Skyrms, B., editors, *Causation in Decision, Belief Change, and Statistics*, volume 2, pages 105–134. reidel, Dordrecht. 108

Stalnaker, R. (1992). Notes on conditional semantics. In *Proceedings of the 4th Conference on Theoretical Aspects of Reasoning About Knowledge*, TARK '92, pages 316–327, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc. 106

Stalnaker, R. (2006). On logics of knowledge and belief. *Philosophical studies*, 128:169–199. 77

Stalnaker, R. C. (1968). A theory of conditionals. *Americal Philosophical Quarterly*, pages 98–112. 95, 98, 99, 102, 113

ter Meulen, A. and van Benthem, J. (2010). *Handbook of logic and language; 2nd ed.* Elsevier Science, Burlington. 12

van Benthem, J. (2001). Correspondence theory. In *Handbook of philosophical logic*, volume 3, pages 325–408. Kluwer Academic Publisher. 26

van Benthem, J. (2010). *Modal logic for open minds.* CSLI publications. 3, 26, 28, 34, 42

van Benthem, J. (2011). *Logical Dynamics of Information and Interaction.* Cambridge University Press. 2, 12, 46, 87

van Benthem, J. and Blackburn, P. (2007). Modal logic: a semantic perspective. In J. van Benthem, P. Blackburn, F. W., editor, *Handbook of Modal Logic*, pages 1–84. Elsevier. 18

van Benthem, J., Blackburn, P., and Wolter, F., editors (2007). *Handbook of Modal Logic.* Elsevier. 28

van Benthem, J., van Ditmarsch, H., van Eijck, J., and Jaspars, J. (2013). Logic in action. Manuscript available at `http://www.logicinaction.org/`. 28

van Ditmarsch, H., van der Hoek, W., and Kooi, B. (2007). *Dynamic Epistemic Logic*, volume 337 of *Synthese library.* Springer. 2, 12, 46, 87

Vardi, M. (2009). From philosophical to industrial logics. *Logic and Its Applications*, pages 89–115. 14

Index

Adams, 101
Additivity
    Finite, 48
    Sub-, 50
    Super-, 50
Algebra, 48
Algebra
    Popper, 61
Aristotle, 7, 10, 13
Artificial intelligence, 3, 14, 114
Assignment, 19
Assignment
    Faithful, 119
Atomic event, 83
Avicenna, 11
Axiom, 9, 11, 22
Axiom
    .2, 76
    .3.2, 76
    .3, 76
    .4, 76
    4, 76, 78
    5, 76, 78
    B, 76
    D, 78
    T, 76, 78

Bayes' rule, 62
Belief
    Base, 118
    Higher-order, 45
    Set, 116
    versus knowledge, 77
Belief function, 50, 58
Belief function
    Conditional, 65
Beth, 13
Bijective, 128
Bisimulation, 38
Bisimulation game, 38
BNF (Backus-Naur Form), 15
Boole, 1, 11, 13

Carroll, 13
Cartesian product, 129
Ceteris paribus, 101
Church, 13
Circumscription, 114
Codomain, 128
Compactness, 23, 34
Complementation, 127
Completeness, 9, 11, 23
Completeness
    Conditional Logic, 104, 105
    Conditional logic, Cond, 112
    Counterfactual logic, CFL, 113
    Default logic, P, 110
    Epistemic logic, 78
    First-order logic, 25
    Modal logic, 25
    Propositional logic, 25
    Public announcement logic, 82
    Strong, 23
Complexity
    Descriptive theory, 14
    Epistemic Logic, 79

Complexity theory, 14
Computer circuits, 15
Computer science, 3, 12
Conditional
    Indicative, 94
    Material, 1, 94
    Strict, 97
    Subjunctive, 94, 112
Confluence, 76
Constant, 19
Contraction operation, 115
Contraposition, 98, 106
Correspondence theory, 26
Counterfactual, 94, 112
Craig's interpolation theorem, 41
Curry-Howard isomorphism, 14

Database, 14
Decidability, 8, 29
Decision problem, 29, 41
Decision procedure, 29
Default logic, 114
Default reasoning, 91, 107
Defeasible inference, 107
Definability, 39
Difference, 127
Dignaga, 11
Domain, 20, 128
Dutch book, 55, 62

Element, 125
Elementary equivalence, 35
Epistemic model, 74, 79
Epistemic model
    Pointed, 74
Euclid, 11
Euclideanity, 26, 76
Event model, 79, 83
Event model
    Pointed, 83
Expansion operation, 116
Expressiveness, 8, 29, 40, 111
Expressiveness
    DEL versus PAL, 87

External approach, 69
External approach
    Imperfect, 48, 56, 69
    Perfect, 48, 56, 69

Filtration, 31
Finite model property, 31
Finite model property
    Effective, 34
FIPA, 87
Formula
    Atomic, 19
    Closed, 19
Frege, 1, 11, 13, 95, 96
Function, 128
Function
    Partial, 128
    Total, 128

Gabbay, 12
Game Semantics, 18, 20
Games
    p-Pebble, 39
    Bijection, 39
    Bisimulation, 38, 39
    Counting, 39
    Ehrenfeucht-Fraïssé, 36
Gangesa, 11
Gentzen, 12–14
Grove system of sphere, 119
Guarded fragment, 31
Gödel, 12, 13

Harper identity, 115
Herbrand, 13
Hilbert, 12, 13
Hintikka, 12, 13, 72
Hypothetical syllogism, 98, 106

Inference rule, 7, 11, 22
Information theory, 46
Injective, 128
Internal approach, 69
Interpretation, 15
Intersection, 127

Interval, 126
Isomorphism, 35
Isomorphism
    Partial, 35

JADE, 87
Jeffrey's rule, 67

Keisler's theorem, 40
Knowledge
    Common, 73
    Distributed, 73, 74
    General, 73
Kolmogorov complexity, 46
Kripke, 12, 13
Kripke frame, 16
Kripke model, 16
Kripke model
    pointed, 16

Lambek, 13
Language
    $\mathcal{L}_{\mathsf{Quant}}$, 58
    $\mathcal{L}_{\mathsf{Qual}}$, 59
    Conditional, 94
    Default, 108
    Dynamic epistemic logic, 87
    Event, 83
    First-order, 19
    Modal, 16
    Propositional, 15
    Public anouncement logic, 81
Law, 7, 10
Leibniz, 13
Levi identity, 115
Lewis, 13, 99
Lindström, 34
Logic
    Basic conditional, 99
    Conditional, 27, 102
    Deontic, 12
    Description, 27
    Dynamic, 27
    Dynamic epistemic (DEL), 86
    Epistemic, 12, 27
    First-order modal, 27
    Free, 27
    Fuzzy, 27
    Higher-order, 27
    Hoare, 28
    Independence-friendly, 27
    Intuitionistic, 26
    Linear, 26
    Many-valued, 27
    Modal, 12
    Non-monotonic, 27
    Paraconsistent, 27
    Possibilistic, 28
    Probabilistic, 28
    Provability, 27
    Public announcement (PAL), 81, 87
    Qualitative reasoning, 59
    Quantitative reasoning, 58
    Relevant, 26
    Separation, 28
    Spatial, 28
    Substructural, 27
    Temporal, 12, 27
Logic programming, 30
Logical consequence, 22
Logical language, 9, 15
Logics, 26, 27
Löwenheim-Skolem theorem, 34

Mapping, 128
Mass function, 51
Mass function
    Consonant, 53
Material implication, 1, 94
Measure
    Conditional inner, 63
    Conditional outer, 63
    Conditional probability, 60
    Inner, 49
    Outer, 49
    Plausibility, 57
    Possibility, 52, 58
    Probability, 48, 58

Qualitative plausibility, 57
Model checking, 14, 42
Monadic first-order logic, 31
Monotonicity, 98, 106
Montague, 12
Muddy children puzzle, 80

Nagel, 71
Natural deduction, 14, 23
Non-monotonic reasoning, 91
Non-standard model, 34
Nonmonotonic reasoning, 107

Only knowing, 86
Organon, 10

Plausibility function, 50, 52
Possibility function, 53
Possibility measure, 52, 58
Possibility measure
Conditional, 66
Possible event, 84
Possible world, 16
Postulates
Belief expansion, 116
Belief revision, 118
Potential isomorphism, 40
Precondition, 84
Predicate
Identity, 19
Symbol, 19
Prefixed formula, 30
Principle of indifference, 49, 54, 62
Probability
Conditional, 60
Conditional
Objective interpretation, 62
Subjective interpretation, 62
Conditional probability space, 61
Lower, 49
Measure, 48
Objective interpretation, 54
Space, 48
Subjective interpretation, 55
Upper, 49
Probability measure, 58
Product update, 79, 85
Program verification, 14
Programming, 14
Programming
Logic, 14
Proof, 11, 22
Proof system, 9, 11, 22
Proof system
Conditional logic, Cond, 112
Default logic, P, 110
Epistemic logic, 77
First-order logic, 24
Modal logic, 24
Propositional logic, 24
Public announcement logic, 82
Public announcement logic (PAL), 80

Qualitative reasoning, 59
Quantitative reasoning, 58

Ramsey, 13
Ramsey test, 120
Ranking function, 53, 58
Ranking function
Conditional, 66
Rational monotonicity, 120
Recursively enumerable, 29
Reflexivity, 26, 76
Relation, 129
Relation
Binary, 129
Representation theorem, 119
Resolution, 30, 42
Revision operation, 115, 118
Rule of inference, 8, 9
Russell, 13, 95

Sahlqvist formula, 26
Satisfaction relation
Conditional logic, 99, 111
Event language, 84
First-order logic, 20

Modal logic, 17
Propositional logic, 15
Satisfiability, 22, 42
Selection function, 98
Semantics, 15
Semantics
Rich, 111
Semi-Euclideanity, 76
Sentence, 19
Sequent calculus, 23
Seriality, 76
Set, 125
Sophists, 10
Soundness, 9, 11, 23
Soundness
Conditional logic, 104, 105
Conditional logic, Cond, 112
Counterfactual logic, CFL, 113
Default logic, P, 110
Epistemic logic, 78
First-order logic, 25
Modal logic, 25
Propositional logic, 25
Public announcement logic, 82
Speech act theory, 87
Stalnaker, 99
Standard translation, 40
Structure
First-order, 19
Lower probability, 57
Pointed measurable qualitative plausibility, 57
Pointed qualitative plausibility, 57
possibility, 57
preferential, 57
probability, 57
ranking, 57
Simple
Conditional probability, 108
Possibility, 108
probability, 108
Qualitative plausibility, 58
Ranking, 108
Subset, 127
Surjective, 128
Surprise, 53
Syllogism, 10
Symetry, 76
Syntax, 15
System of spheres, 99

Tableau, 30, 103
Tableau method, 30, 42, 103
Tarski, 13
Term, 19
Theorem, 9, 22
Three prisoners puzzle, 63
Transitivity, 25, 76
Truth, 8, 11, 22
Truth-functional, 95
Tuple, 127
Turing, 13, 14

Ultraproduct, 40
Undecidability, 29, 42
Union, 127

Validity, 8, 22
Valuation function, 16
van Benthem theorem, 41
Variable, 19
Variable
Bound, 19
Free, 19

Weak connectedness, 76
Well-formed formula, 9, 15
Wittgenstein, 13, 95

Logique et Raisonnement de Sens Commun

Guillaume Aucher

Pendant longtemps, la logique s'est intéressée au raisonnement mathématique ainsi qu'à la formalisation et aux fondements des mathématiques. C'est ce qui a motivé le développement de la logique dite mathématique dans la première moitié du $20^{ième}$ siècle. Dans la seconde moitié du $20^{ième}$ siècle, de par l'émergence de l'informatique et de l'intelligence artificielle, de nombreuses logiques et formalismes ont été développés avec pour objectif de modéliser des types de raisonnements plus proches de ceux que nous utilisons dans la vie courante, notamment en présence d'incertitude, et qui sont parfois assez éloignés du raisonnement de type mathématique.

Ces notes de cours sont une introduction aux logiques et formalismes qui étudient le raisonnement dit de sens commun, c'est à dire le raisonnement humain de "la vie courante". En particulier, elles abordent le raisonnement que nous effectuons dans des situations pour lesquelles nous avons une certaine incertitude quant à l'occurrence ou l'existence de certains événements ou faits. Ainsi, nous présentons les logiques non-monotones, les logiques pour conditionnels, la théorie de révision des croyances (pour le raisonnement de sens commun) ainsi que la logique épistémique (dynamique), les logiques probabilistes, possibilistes, les plausibilités, *etc.* (pour la représentation et le raisonnement en présence d'incertitude).

Ces notes de cours sont illustrées par de nombreux exemples intuitifs qui permettent de mieux appréhender et comprendre les différents concepts (formels) introduits. Elles ne nécessitent pas de recourir à d'autres ouvrages pour pouvoir être comprises et étudiées et incluent en particulier un rappel des rudiments de logique et de théorie des ensembles.

Bibliographie:

- J. Halpern. Reasoning about Uncertainty. MIT Press, Cambridge, Massachussetts, 2003.
- J. van Benthem. Modal Logic for Open Minds. CSLI, 2010.
- R. Fagin, J. Halpern, Y. Moses, M. Vardi. Reasoning about Knowledge. MIT Press, 1995.
- P. Gardenförs. Knowledge in Flux (Modelling the Dynamics of Epistemic States). Bradford/MIT Press, Cambridge, Massachussets, 1988.
- L. Goble. The Blackwell Guide to Philosophical Logic. Wiley-Blackwell, 2001.
- G. Priest. An Introduction to Non-classical Logic. Cambridge University Press, 2011.