FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Rätselhafte Ignoranz

Ein Rundgang durch epistemische Logik

MASTERARBEIT

zur Erlangung des akademischen Grades

Master of Science

im Rahmen des Studiums

European Master's Program in Computational Logic

eingereicht von

Ana A. Oliveira da Costa
Matrikelnummer 01624245

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Christian Fermüller

Wien, 4. April 2018

Ana A. Oliveira da Costa | Christian Fermüller

FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Puzzling Ignorance

A Modal Epistemic Tour

MASTER'S THESIS

submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

European Master's Program in Computational Logic

by

Ana A. Oliveira da Costa
Registration Number 01624245

to the Faculty of Informatics

at the TU Wien

Advisor: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Christian Fermüller

Vienna, 4th April, 2018

Ana A. Oliveira da Costa | Christian Fermüller

Erklärung zur Verfassung der Arbeit

Ana A. Oliveira da Costa
Döblinger Gürtel 8/18, Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 4. April 2018

Ana A. Oliveira da Costa

For Rosa Pereira,

Acknowledgements

I would like to thank my supervisor, Christian Fermüller, for challenging me with the topic of this Master thesis. His availability to discuss and review my ideas, kept me motivated and contributed for the success of this thesis.

I am thankful, as well, to all friends that I met during the past two years. With you by my side Dresden, Bolzano and Vienna felt a bit more like home. I am thankful to my friends in Portugal we were always there for me, even when I had so few time to be with them.

Finally, I am very grateful to my parents, for all the love they share with me. And, last but not least, I would like to thank Peter for being my support and to encourage me to study again.

Without you all this accomplishment would not have been possible.

Thank you. *Obrigado.*

Kurzfassung

Epistemische Modallogiken öffnen einen formalen Zugang zum Studium des Schließens über Wissen. Üblicherweise werden dabei *epistemische Szenarien* beschrieben um relevante Begriffe einzuführen und zu diskutieren. Unter diesen sind *Puzzles* und *Paradoxien* von besonderem Interesse, da sie unsere Intuitionen herausfordern und den Vergleich verschiedener Formalismen unterstützen. In dieser Arbeit verwenden wir einige bekannte Puzzles und Paradoxien um die Gundlagen der epistemischen Modallogik zu erläutern. Motiviert durch die Analyse epistemischer Szenarions, in denen die Akteure wiederholt ihre Nicht-Wissen deklarieren, diskutieren wir insbesondere neuere Entwicklungen zur Modallogik von “Ignoranz”.

Insgesamt versucht diese Masterarbeit eine einheitliche und in sich geschlossene Einführung in aussagenlogische Modallogik zu bieten. In Fokus sind dabei Modaloperatoren, die jeweils “wissen, dass”, “Gruppen-Wissen” und “Ignoranz” ausdrücken. Wir ergänzen die nötigen technischen Überlegungen mit vielen Beispielen und Einsichten zu verwandten Arbeiten der Erkenntnistheorie. Außerdem benutzen wir die “Logik der Ignoranz” um zu beweisen, dass die Anzahl unterscheidbarer epistemischer Zustände gleichmächtig mit dem Kontinuum der reellen Zahlen ist. Diese Einsicht wird in der Literatur zu epistemischer Logik nur selten berücksichtigt, obwohl aus diesem Resultat folgt, dass es epistemische Zustände gibt, die durch kein endliches Kripke-Modell adäquat representiert werden können.

Abstract

Epistemic modal logic is a formal approach to study reasoning about knowledge. It is common practice to describe *epistemic scenarios* to introduce and explore epistemic notions. Among those, the most interesting are *puzzles* and *paradoxes* because they challenge our intuitions and they can be used to compare formalisms. In this thesis we use some well-known puzzles and paradoxes to review the foundations of epistemic modal logic. In particular, motivated by the analysis of scenarios in which ignorance is repeatably announced, we review recent developments on a modal logic approach to capture the notion of *ignorance*.

Overall, this thesis provides an unified and self-contained introduction to propositional epistemic logic with a focus on modalities of *knowing that*, *group knowledge* and *ignorance*. We complement the standard technical considerations with plenty of examples and insights from related work in epistemology. In addition, we use the logic of ignorance to prove that the cardinality of the epistemic state space is the same as the cardinality of the *continuum*. This result implies, for instance, that it is not possible for an agent to name all epistemic states. This limitation is usually not included in the epistemic logic literature. In addition, the result implies that some epistemic states cannot be adequately represented by finite Kripke models.

Contents

CHAPTER 1

Introduction

A logical theory may be tested by its capacity for dealing with puzzles, and it is a wholesome plan, in thinking about logic, to stock the mind with as many puzzles as possible, since these serve much the same purpose as is served by experiments in physical science.

– Bertrand Russell, in *On Denoting*

Epistemic logic applies formal systems to study epistemic reasoning. This approach was motivated by the early idea held by philosophers like Rudolf Carnap or G.H. von Wright that some properties of knowledge could be captured by an axiomatic deductive system. The seminal book by Jaakko Hintikka, *Knowledge and Belief: An Introduction to the Logic of the Two Notions* [Hin62], was the first full-length book to discuss both a semantic view on epistemic reasoning, based on *possible world semantics*, and an axiomatic characterization of some epistemic notions. His approach turned out ot be very successful and it influences the work in this area up to these days. The contribution in late 90's from fields like computer science [FHVM95] and game theory [Aum99] boosted the development of epistemic logic and turned it into a multidisciplinary field driven by both philosophical and application related issues.

In [vDvdHK07, vDHvdHK15] the authors present an overview of epistemic logic recent technical progress, as well as, some of its applications. Those developments spawn from the analysis of different informational attitudes like an agent only knowing some statement or an agent having limited awareness [vDHvdHK15], to the study of the outcome of epistemic actions with dynamic epistemic logic [vDvdHK07,Pac13]. While epistemic logic initial motivation was to investigate questions in epistemology, its recent developments are mostly motivated by computational concerns and focused on propositional modal logic of *knowing that*.

This thesis uses *epistemic scenarios* to explore the foundations of epistemic logic. An *epistemic scenario* is a description of a situation involving epistemic reasoning. From those, the most interesting are *puzzles* and *paradoxes* because they challenge our intuitions and can be used as test cases for formalisms. Usually, in an epistemic puzzle, the goal is to find a consistent epistemic justification for the sequence of actions described. One such puzzle is presented below:

> *Example* 1.1 (*Epistemic puzzle [Pac13]*).
> Three logicians walk into a bar.
> The barman asks, 'Does everybody want a beer?'
> The first logician says, 'I don't know.'
> The second logician says, 'I don't know.'
> The third logician says, 'Yes.'

In this puzzle iterated announcements of ignorance about a statement leads to one of the agents to know whether that same statement holds. This pattern occurs in many epistemic puzzles, like the muddy children or the sum and product puzzles (for these and other epistemic puzzles see [vDK15]). This motivates us to have a closer look at ignorance.

We say that an agent *is ignorant about a statement* if and only if *neither the agent knows that the statement holds nor does he know that the statement does not hold*. In epistemic logic, *knowing that* is interpreted as a necessity modality. A proposition is *contingent* when it is neither necessarily true nor necessarily false. It is easy to see that ignorance is the contingent counterpart of knowledge. The dual of ignorance is *knowing whether*. Thus, an agent *knows whether a statement holds* if and only if the agent *either knows that the statements holds or he knows that it does not hold*.

The notions of ignorance and knowing whether are useful in epistemic applications. There are scenarios where it is only necessary to check whether an agent knows the truth-value of a sentence, independently of its actual value. For example, in [Rei01] the authors use a *know whether* operator to define succinct postconditions for *'knowledge-producing actions'*. In spite of their advantages and the fact they are recurrent in epistemic puzzles, these notions are barely mentioned in the epistemic logic literature. Exceptions are a logic for ignorance presented in [vdHL04] and a logic for knowing whether discussed in [FWvD15, Fan16].

The aim of this thesis is to fill a gap in the literature by presenting an uniform introduction to both the logic of knowing that and the logic of ignorance. To this end, we review the foundations of epistemic logic to later compare it with its contingent counterpart, ignorance.

This thesis is structured as follows. In chapter 2 we introduce our epistemic models and explain how to interpret the notions of *knowing that* and *knowledge within a group* in such models. We finish the chapter with an analysis of the Muddy Children puzzle.

The next chapter, is about deductive systems for knowledge and common knowledge. We start with a discussion about the relation between our modeling assumptions and some axioms with interesting epistemic interpretation. We, then, prove that the system $\mathcal{S}5$ is sound and complete for the class of our epistemic models $\mathbb{S}5$, for both the basic epistemic logic and its extension with common knowledge. At the end of the chapter we briefly discuss the Moorean paradox. The last chapter is dedicated to the logic of knowing whether and ignorance. In that chapter we prove, using the notion of *knowing whether*, that there continuum-many different epistemic states, when we consider at least two agents.

CHAPTER 2

Epistemic Models

In this chapter we introduce basic epistemic logic and its extension with notions referring to group of agents. We focus on their semantic interpretation over Kripke structures, due to its intuitive characterization. This chapter is meant to be used as a gentle introduction to the broad topic of epistemic modal logic.

We start the chapter by an historical overview of epistemic logic. The next section presents modal logic preliminaries. In section 2.3, we define the *basic epistemic logic* language and discuss the intended interpretation of its sentences over Kripke models. In section 2.4, we overview some group knowledge notions with a focus on the notions of *general knowledge* and *common knowledge*. We finalize this chapter by discussing a solution for the *Muddy Children puzzle* using all what was presented before. Along the chapter we use a simple epistemic scenario as a running example to help clarify some of the ideas discussed.

2.1 Historical Overview

Epistemology is a branch of philosophy that studies the notion of knowledge. This includes, for instance, the definition of knowledge, the analysis of its structural properties and limitations, and the study of mechanisms that enable an agent to acquire it [Ste17, Hol13]. Epistemic logic applies formal systems to study epistemic reasoning. Attempts to capture valid epistemic reasoning by means of a formal system can be traced back to the Middle Ages. Paul Gochet and Pascal Gribomont give a quick overview of the work done during this era in the introduction of [GG06]. In this section we present the contemporary developments in epistemic logic with emphasis on the modal logic approach.

In the late 1940's and early 1950's many philosophers and logicians believed that the way we reason about knowledge could be captured by means of an axiomatic system [HS15]. Von Wright [vW51] was the first to propose a syntactic characterization of formal epistemic reasoning based on modal notions. Later, the seminal book by Jaakko Hintikka, *Knowledge and Belief: An Introduction to the Logic of the Two Notions* [Hin62], was the first work to discuss both a semantic view on epistemic reasoning, based on *possible world semantics*, and an axiomatic characterization of some epistemic notions. Hintikka's goal was to establish formal criteria to determine the consistency of a set of statements expressing epistemic notions. In order to do so, he introduced and justified reasoning principles to characterize this notion. Later in the book he uses this formal system to investigate some of the problems discussed in epistemology at that time. He proves, for example, that it is inconsistent for an agent to believe in a Moorean sentence about his own current beliefs, i.e. an agent cannot consistently believe in a statement of the form 'it's raining, but I do not believe it is raining'.

The use of epistemic formal reasoning turned out to be relevant for other disciplines, as well. Since the late 70's there is an increasing interest in the study of knowledge notions in the context of multi-agents systems with contributions from fields like artificial intelligence, game theory and computer science. In artificial intelligence, for example, Robert Moore in [Moo77] proposed to use a first order version of epistemic modal logic to express preconditions on agents' actions. In [Aum76], Aumann uses the concept of common knowledge to prove his famous agreement theorem in game theory. This theorem states that if two rational agents have the same probability distribution before any evidence is presented (prior probabilities) and have common knowledge of each other's beliefs about the probability distribution after a relevant evidence is considered (posterior probabilities), then their posterior probabilities must be equal. In late 90's two text books from computer science [FHVM95, MH95] summarize some of the work done in this field at that time. All these contributions boosted the development of epistemic logic and turned it into a multidisciplinary field driven by both philosophical and application related issues.

2.2 Modal Logic Preliminaries

In this section we present modal logic concepts used throughout this thesis. We assume that the reader has basic knowledge about modal logic. Otherwise, the books [BdRV01, vB10] can be used to complement the material discussed here.

2.2.1 Language

We extend classical propositional logic with modal operators. We consider a countable (non-empty) set of atomic formulas At, which are used to express propositions that are considered to be logically elementary, for example *'It is sunny'*. Additionally, we have a finite (non-empty) set of agents Ag and a finite set of modal operators Op.

Definition 2.2.1 (Modal language $\mathcal{L}$)**.**
Let At be a countable set of atomic propositions, Op a finite set of modal operators, and Ag a finite set of agent symbols. All these sets are non-empty. The *multi-modal language*

$\mathcal{L}(At, Op, Ag)$ is defined by the following grammar in Backus normal form (BNF), where $p \in At$ and $\Box \in Op$:

$$\varphi \ ::= \ p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box\varphi.$$

It is common to interpret formulas of the form $\Box\varphi$, with $\Box \in Op$, as *'it is necessary that φ'*, while their dual $\Diamond\varphi$ is read as *'it is possible that φ'*. In table 2.1 we list standard abbreviations for other boolean operators and for the dual of each $\Box$ in our set Op.

Description/name	*definiendum*	*definiens*
false	$\bot$	$p \wedge \neg p$, for some $p \in At$
true	$\top$	$\neg\bot$
disjunction	$\varphi \vee \psi$	$\neg(\neg\varphi \wedge \neg\psi)$
implication	$\varphi \rightarrow \psi$	$\neg\varphi \vee \psi$
biconditional	$\varphi \leftrightarrow \psi$	$\varphi \rightarrow \psi \wedge \psi \rightarrow \varphi$
dual of $\Box$	$\Diamond\varphi$	$\neg\Box\neg\varphi$

Table 2.1: Abbreviations in the language $\mathcal{L}$ of other logical and modal operators.

Definition 2.2.2 (Iterated Application of $\Box$)**.**
Let $\Box$ be a modal operator, either one in Op or one defined as an abbreviation. The nth iterated application of $\Box$, written $\Box^n$, is defined as follows:

$$\Box^0\varphi = \varphi \quad \textit{and} \quad \Box^{n+1}\varphi = \Box\Box^n\varphi.$$

2.2.2 Semantics

Definition 2.2.3 (Kripke model and Kripke frame)**.**
Let At be a set of propositions and Op a set of modal operators. A *Kripke model* $\mathcal{M}$ is a structure $\mathcal{M} = \langle \mathcal{W}, \{\mathcal{R}_\Box \mid \Box \in Op\}, \mathcal{V}\rangle$ such that:

- $\mathcal{W} \neq \emptyset$ is a set of possible worlds;
- $\mathcal{R}_\Box \subseteq \mathcal{W} \times \mathcal{W}$ is an accessibility relation for each operator $\Box \in Op$;
- $\mathcal{V} : (\mathcal{W} \times At) \rightarrow \{\textit{true}, \textit{false}\}$ is a valuation assigning truth values to propositions at worlds. [1]

A pair $(\mathcal{M}, w)$ with $w \in \mathcal{W}$ is called a *pointed model*. We use either $(w, v) \in \mathcal{R}_\Box$ or $w\mathcal{R}_\Box v$, if worlds $w \in \mathcal{W}$ and $v \in \mathcal{W}$ are connected by the accessibility relation $\mathcal{R}_\Box$.

[1] It is often useful to consider an equivalent definition of valuation, $\mathcal{V}_{At} : At \rightarrow 2^{\mathcal{W}}$, where given a proposition $p \in At$ then $\mathcal{V}_{At}(p)$ yields the set of worlds in which p is true.

A *Kripke frame* $\mathcal{F}$ is a structure $\mathcal{F} = \langle \mathcal{W}, \{\mathcal{R}_\Box \mid \Box \in Op\}\rangle$ that abstracts from a specific valuation. Frames allow us to focus on models' structural properties. A *model is based on a frame* when they have the same set of worlds and accessibility relations. Given a class of frames $\mathbb{F}$ we define the class of models based on $\mathbb{F}$ as $\bigcup_{\mathcal{F}\in\mathbb{F}}\{\mathcal{M} \mid \mathcal{M}$ is based on $\mathcal{F}\}$.

It is important to note that there are two levels to be considered in our semantics. First, we define *satisfability* of a formula over a pointed model, i.e. we define the *truth value* of a given formula over a Kripke model from the point of view of one of its possible worlds. Second, we use frames to define *validity*, because Kripke frames abstracts from the evaluation given to proposition variables.

Definition 2.2.4 (Truth interpretation of modal formulas in Kripke structures)**.**
Let $\mathcal{M} = \langle \mathcal{W}, \{\mathcal{R}_\Box \mid \Box \in Op\}, \mathcal{V}\rangle$ be a Kripke model and $w \in \mathcal{W}$ a world. We define inductively that a modal formula $\varphi \in \mathcal{L}(At, Op, Ag)$ *holds in the pointed model* $(\mathcal{M}, w)$ as follows:

- $(\mathcal{M}, w) \models p$ iff $\mathcal{V}(w, p) = true$, for $p \in At$;
- $(\mathcal{M}, w) \models \neg\varphi$ iff it is not the case that $(\mathcal{M}, w) \models \varphi$, i.e. $(\mathcal{M}, w) \not\models \varphi$;
- $(\mathcal{M}, w) \models \varphi \wedge \psi$ iff $(\mathcal{M}, w) \models \varphi$ and $(\mathcal{M}, w) \models \psi$;
- $(\mathcal{M}, w) \models \Box\varphi$ iff for all $v \in \mathcal{W}$ if $w\mathcal{R}_\Box v$ then $(\mathcal{M}, v) \models \varphi$.

When $(\mathcal{M}, w) \vDash \varphi$ for all $w \in \mathcal{W}$, we write $\mathcal{M} \vDash \varphi$ and say that φ is true in $\mathcal{M}$. Moreover, if we consider a class of models $\mathbb{X} \subseteq \mathbb{K}$ we say that $\mathbb{X} \vDash \varphi$ holds when $\mathcal{M} \vDash \varphi$ holds for all $\mathcal{M} \in \mathbb{X}$.

Definition 2.2.5 (Satisfiability and validity)**.**
A modal formula φ is:

- *satisfiable* in a model $\mathcal{M}$ if there is a world w in $\mathcal{M}$ such that $(\mathcal{M}, w) \vDash \varphi$;
- *valid at a world w in a frame $\mathcal{F}$*, denoted $(\mathcal{F}, w) \vDash \varphi$, if φ is true at w in every model $\mathcal{M}$ based on $\mathcal{F}$;
- *valid in a frame $\mathcal{F}$*, denoted $\mathcal{F} \vDash \varphi$, if it is valid at every world in $\mathcal{F}$;
- *valid on a class of frames* $\mathbb{F}$, denoted $\vDash_{\mathbb{F}} \varphi$, if it is valid on every frame $\mathcal{F} \in \mathbb{F}$;
- *valid*, denoted $\vDash \varphi$, if it is valid on the class of all frames.

We present in table 2.2 a list of class of frames that are relevant for this thesis. Given that we have n agents, then $\mathbb{K}n$ is the class of all Kripke frames, at times we will refer to it as the class of all Kripke models.

Classe name	Agent's accessibility relation characterization
$\mathbb{T}n$	Reflexive
$\mathbb{D}n$	Serial
$\mathbb{B}n$	Symmetric
$4n$	Transitive
$5n$	Euclidean
$\mathbb{S}4n$	Reflexive and transitive
$\mathbb{S}5n$	Equivalence

Table 2.2: Class of frames used in this thesis and their characterization.

In this thesis we consider a variety of modal languages which can be interpreted over different class of models. Below we define that given a language, $\mathcal{L}$, and a class of frames, $\mathbb{F}$, the logic generated by them is the set of all $\mathcal{L}$ formulas that are valid in $\mathbb{F}$.

Definition 2.2.6 (Logic of $\mathcal{L}$ over $\mathbb{F}$)**.**
Let $\mathcal{L}$ be a modal language and $\mathbb{F}$ a class of frames. The set of all $\mathcal{L}$-formulas that are valid in $\mathbb{F}$ defines the *logic of $\mathcal{L}$ over* $\mathbb{F}$, $\mathbb{F}_{\mathcal{L}}$, as follows:

$$\mathbb{F}_{\mathcal{L}} = \{\varphi \mid \ \vDash_{\mathbb{F}} \varphi \text{ and } \varphi \in \mathcal{L}\}.$$

2.3 Basic Epistemic Modal Logic

In this section we introduce basic epistemic logic and epistemic models.

2.3.1 Epistemic Modal Language

Definition 2.3.1 (Basic epistemic modal language $\mathcal{L}_{\mathbf{K}}$)**.**
The *basic epistemic modal language* $\mathcal{L}_{\mathbf{K}}(At, Ag) = \mathcal{L}(At, \{\mathbf{K}_a \mid a \in Ag\}, Ag)$ is a multi-modal language with a knowledge operator $\mathbf{K}_a$ for each agent $a \in Ag$.

A formula $\mathbf{K}_a\varphi$ can be read as *'agent a knows φ'*. Another possible interpretation for $\mathbf{K}_a\varphi$ is *'agent a is informed that φ is true'*. The dual of $\mathbf{K}_a$ is denoted by $\hat{\mathbf{K}}_a$ and, as defined in table 2.1, $\hat{\mathbf{K}}_a\varphi = \neg\mathbf{K}_a\neg\varphi$. The formula $\hat{\mathbf{K}}_a\varphi$ is often read as *'φ is consistent with the knowledge of agent a'* or *'agent a considers φ possible'*.

Throughout this chapter we will use the Groningen-Liverpool-Otago scenario (GLO scenario), adapted from [vDvdHK07], as our running example of modal epistemic logic usage.

Example 2.1 (*Simple knowledge theory*).
Suppose we have one agent, say Bert (b), who lives in Groningen. For some reason, he builds a theory about the weather conditions in both Groningen and Liverpool: in

Groningen it is either sunny (denoted by the atom g), or not ($\neg g$); and likewise for Liverpool, it is either sunny (l) or not ($\neg l$).

We present below two $\mathcal{L}_{\mathbf{K}}$ formulas encoding Bert's knowledge about the weather in Groningen and Liverpool.

$$\begin{aligned} &\textit{Bert knows whether it is sunny in Groningen}: && \mathbf{K}_b g \vee \mathbf{K}_b \neg g \\ &\textit{Bert is ignorant about the weather in Liverpool}: && \neg \mathbf{K}_b l \wedge \neg \mathbf{K}_b \neg l \end{aligned}$$

The second formula encodes the fact that he is *ignorant* about the weather in Liverpool, i.e. he does not know that it is sunny there and he does not know the opposite either. If we consider the abbreviations defined in table 2.1 together with some classical logic equivalences, then we can encode his ignorance by any of the following equivalent formulas: $\neg \mathbf{K}_b l \wedge \neg \mathbf{K}_b \neg l \equiv \neg(\mathbf{K}_b l \vee \mathbf{K}_b \neg l) \equiv \hat{\mathbf{K}}_b \neg l \wedge \hat{\mathbf{K}}_b l$. The second formula can be read as *it not the case that Bert knows whether it is sunny in Liverpool*. We can read it using the interpretation of the dual of the knowledge operator, as well. Then the formula says *that each of the possibilities is consistent with Bert's current knowledge.*

Our language should allow us to encode statements concerning knowledge about knowledge, i.e. *high-order knowledge*. That includes statements about knowledge of an agent about other agents' knowledge, which are very useful to capture epistemic reasoning in multi-agent scenarios.

Example 2.2 (*Knowledge about knowledge*).
Bert is friend with Cat who lives in Liverpool. He knows that she knows whether it is sunny there.

We encode the knowledge of Bert about Cat's knowledge (Cat is the agent c), as follows:

$$\textit{Bert knows that Cat knows whether it is sunny in Liverpool}: \quad \mathbf{K}_b(\mathbf{K}_c l \vee \mathbf{K}_c \neg l).$$

If Bert gets to know that Cat knows that is is sunny in Liverpool, then we can encode his new knowledge with the formula:

$$\textit{Bert knows that Cat knows that it is sunny in Liverpool}: \quad \mathbf{K}_b \mathbf{K}_c l.$$

2.3.2 Epistemic Models

A Kripke model encoding agents' information is called an *epistemic model*. The design of our epistemic models is based on the following assumptions:

(a) *worlds* encode *possible configurations of the scenario* being modeled, which are consistent with agents' current knowledge;

(b) for each agent we define an *accessibility relation* connecting worlds that are *indistinguishable* for that agent, given his current knowledge.

The interpretation given to the accessibility relation may induce different theories of knowledge. In the interpretation adopted in this thesis each accessibility relation turns out to be an equivalence relation. It is easy to see that the following properties must hold in the agent's accessibility relation in our epistemic models:

Reflexivity An agent cannot distinguish a world from that same world;

Symmetry If an agent cannot distinguish world w from world v, then he cannot distinguish world v from world w;

Transitivity If an agent cannot distinguish world w from world v and world v from u, then he cannot distinguish world w from world u.

Therefore, we will work with the class of frames $\mathbb{S}5n$ (see table 2.2). From now on, to keep our models concise, we may not draw the reflexive arrows and arrows that can be obtained by transitive closure over one of the accessibility relations.

We could have adopted a different interpretation for each of the agents' accessibility relation. For example, we could have considered that for a given agent the world w is accessible to world v if and only if everything that the agent knows in w he knows it in v as well [Hol13, Wil00]. Under this interpretation epistemic models are not necessarily symmetric nor transitive [Hol13].

The *current world* represents the real configuration of the epistemic scenario being modeled. Epistemic formulas are usually interpreted at that world to narrow our analysis to only one of the possible epistemic configurations.

Example 2.3 (*Epistemic states and indistinguishable relations*).
Since Bert is situated in Groningen, we can assume that he is aware of the weather in Groningen, but not of that in Liverpool. Surprisingly, in our current scenario it is sunny in both cities.

Our goal is to build a model depicting what Bert knows about the weather in Groningen and Liverpool, i.e. an *epistemic model* of what was described so far.

We start by identifying four states matching the four possible combinations of the weather conditions considered by Bert (i.e. whether or not it is sunny) in both cities. We call those states *possible worlds*. Each pair inside of a world defines the valuation given to g (*it is sunny in Groningen*) and l (*it is sunny in Liverpool*) in that possible world. For example, in world w_1 both g and l are satisfied, which means that this world corresponds to the possibility that it is sunny in both cities; or the world w_2 that represents the situation in which it is not sunny in Groningen (g does not hold in this world) and it is sunny in Liverpool. As stated in the scenario description it is sunny in both cities, therefore the real world is represented by w_1.

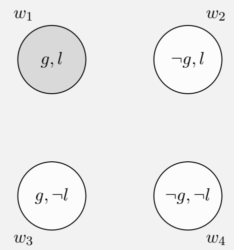

Figure 2.1: Possible worlds that we consider to model Bert's knowledge of weather conditions in Groningen and Liverpool.

Since Bert knows whether it is sunny in Groningen, he can distinguish between any two situations with different weather conditions in Groningen. This means that he can *distinguish*, for example, between states w_1 and w_2. However, Bert does not know about the weather in Liverpool and thus he cannot distinguish between states in which the weather is the same in Groningen but different in Liverpool. For example, the states w_1 and w_3.

Starting from the states identified in the previous picture we build our epistemic model of Bert's knowledge by connecting by an arrow, labeled with b, states that he cannot distinguish.

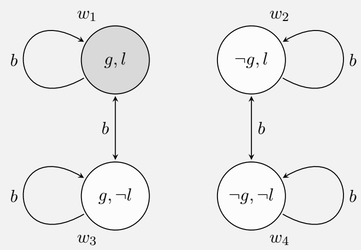

Figure 2.2: Epistemic model for Bert's current knowledge – $\mathcal{M}_{Bert}$.

We interpret $\mathbf{K}_a$ as a necessity operator. Then, given that we have one accessibility relation for each agent $a \in Ag$, we interpret $\mathcal{L}_{\mathbf{K}}$ modal operators as follows:

$$(\mathcal{M}, w) \models \mathbf{K}_a\varphi \text{ iff for all } v \in \mathcal{W}, \text{ if } w\mathcal{R}_a v \text{ then } (\mathcal{M}, v) \models \varphi$$

$$(\mathcal{M}, w) \models \hat{\mathbf{K}}_a\varphi \text{ iff there exists a } v \in \mathcal{W} \text{ such that } w\mathcal{R}_a v \text{ and } (\mathcal{M}, v) \models \varphi.$$

Given the above definition and the design assumptions about $\mathcal{R}_a$, we define that an agent *knows* or *is informed that* φ when the agent *has no uncertainty* about φ, i.e. φ is necessary given agent's current knowledge.

Example 2.4 (*Interpretation of epistemic formulas*).
Consider the previous example and the Kripke model, $\mathcal{M}_{Bert}$, represented by figure 2.2. We proceed now by checking if what we learned so far about Bert's knowledge holds in $\mathcal{M}_{Bert}$ at the current world (w_1).

Bert knows whether it is sunny in Groningen, which can be encoded by the formula $\mathbf{K}_b g \vee \mathbf{K}_b \neg g$ (from example 2.1). In fact, $\mathbf{K}_b g \vee \mathbf{K}_b \neg g$ holds at the world w_1 in $\mathcal{M}_{Bert}$ because $(\mathcal{M}_{Bert}, w_1) \vDash \mathbf{K}_b g$. Note that the worlds accessible from w_1 are w_1 and w_3, and in both the proposition g holds.

In addition, we know that Bert is ignorant about the weather in Liverpool. The formula $\neg\mathbf{K}_b l \vee \neg\mathbf{K}_b \neg l$ (from example 2.1) should hold at world w_1 in $\mathcal{M}_{Bert}$. As $(\mathcal{M}_{Bert}, w_1) \vDash l$ and $(\mathcal{M}_{Bert}, w_3) \vDash \neg l$, then $(\mathcal{M}_{Bert}, w_1) \vDash \neg\mathbf{K}_b l$. Therefore, $(\mathcal{M}_{Bert}, w_1) \vDash \neg\mathbf{K}_b \neg l$.

The most interesting scenarios to analyze using epistemic logic occur when we reason about the knowledge of multiple agents. Luckily, Kripke models provide a natural interpretation of arbitrary nested knowledge formulas. In scenarios with multiple agents we usually expect the knowledge of any two agents to be independent. For example, the formula $\mathbf{K}_a \varphi \rightarrow \mathbf{K}_b \varphi$ is not expected to be valid for arbitrary agents a, b and formula φ. This property makes it trivial to extend our single-agent logic to a multi-agent epistemic logic. In fact, given n agents we only need to consider n independent accessibility relations in our model and we have an epistemic model for n agents.

Example 2.5 (*High-order knowledge and epistemic models*).
Recall the example 2.2:

Bert is friend with Cat who lives in Liverpool. He knows that she knows whether it is sunny there.

As we were not informed otherwise we assume that Cat is ignorant about Groningen weather conditions. We will extend the model $\mathcal{M}_{Bert}$, as in figure 2.2, to include Cat's accessibility relations. Recall that, to keep the models concise we will not draw reflexive arrows and arrows that can be obtained by transitive closure of an accessibility relation.

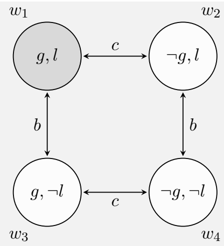

Figure 2.3: Kripke model for Bert and Cat's current knowledge.

The reader can check that the formulas $\mathbf{K}_c l$ and $\neg\mathbf{K}_c g \wedge \neg\mathbf{K}_c\neg g$ holds in this model at world w_1. Moreover, we can see that $\mathbf{K}_c l \vee \mathbf{K}_c\neg l$ holds everywhere in this model.

It is also useful to reason about *high-order knowledge*. For example, we can check that in this model at world w_1 Bert knows that Cat knows whether it is sunny in Liverpool, i.e. the following formula holds at w_1:

$$\mathbf{K}_b(\mathbf{K}_c l \vee \mathbf{K}_c\neg l).$$

In order to check that the formula holds at w_1, we will need to verify that $\mathbf{K}_c l \vee \mathbf{K}_c\neg l$ holds at the accessible worlds from w_1 using $\mathcal{R}_b$, i.e. at w_1 and w_3. We can then see that $(\mathcal{M}, w_1) \vDash \mathbf{K}_c l$ and $(\mathcal{M}, w_3) \vDash \mathbf{K}_c\neg l$.

Note that, even thought it is not explicit in our epistemic scenario description that it is the case that Cat knows that Bert knows whether it is sunny in Groningen, the formula $\mathbf{K}_c(\mathbf{K}_b g \vee \mathbf{K}_b\neg g)$ holds in this model at w_1, as well. In fact, it holds everywhere in this model. We may want to consider instead that Cat is ignorant about Bert's knowledge of Groningen weather conditions. We present below a Kripke model where the following formula for Cat's ignorance holds at w_1:

$$\neg\mathbf{K}_c(\mathbf{K}_b g \vee \mathbf{K}_b\neg g) \wedge \neg\mathbf{K}_c\neg(\mathbf{K}_b g \vee \mathbf{K}_b\neg g).$$

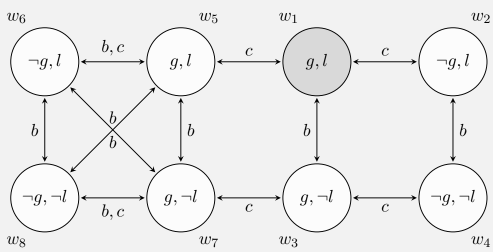

Figure 2.4: Kripke model for Bert and Cat's current knowledge assuming that Cat is ignorant about Bert's knowledge of Groningen weather conditions.

2.4 Group Knowledge

In this section we extend basic epistemic logic $\mathcal{L}_{\mathbf{K}}$ with operators for different types of group knowledge. The main idea is that given a group of agents we can combine their knowledge in different ways to capture such notions.

Example 2.6 (*Knowledge in a group*).
Consider the epistemic scenario discussed in example 2.5. The model below depicts the scenario in which Bert knows that Cat knows whether it is sunny in Liverpool and the other way around. Given a group of agents, $G \subseteq Ag$, we motivate below four group knowledge operators and use this model to illustrate their differences. Recall that to keep models concise we don't draw the reflexive and transitive arrows.

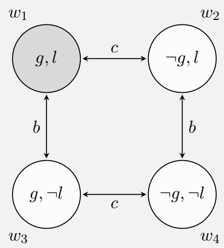

Figure 2.5: Kripke model for Bert and Cat's knowledge considering that each of them knows that the other knows whether it is sunny in their city – $\mathcal{M}_{Group}$.

Distributed knowledge ($\mathbf{D}_G$): *What would the agents in a group know if all of them shared their knowledge?*

If Bert and Cat would tell each other the weather in the cities they live in, then, for instance, Bert would not be uncertain about the weather in Liverpool any longer. If we consider the model above $\mathcal{M}_{Group}$, this means that Bert would be able to distinguish the worlds w_1 and w_3. Thus, his accessibility relation would not connect them anymore. It is easy to see that, if an agent in a group can distinguish two possible worlds, then it is distributed knowledge that these worlds are different. Therefore, we can interpret distributed knowledge as a $\Box$-operator defined over the intersection of all accessibility relations of the agents in the group.

In our example it is distributed knowledge among Bert and Cat that it is sunny in Groningen and in Liverpool (i.e. $g \wedge l$). Thus, the following formula holds at w_1:

$$\mathbf{D}_{\{b,c\}}(g \wedge l).$$

It is important to note that neither Bert nor Cat know this.

Dispersed knowledge ($\mathbf{S}_G$): *What does at least one of the agents in a group know?*

Dispersed knowledge is a collection of each agent's knowledge. It is usually defined as: $\mathbf{S}_G\varphi \overset{def}{\equiv} \bigvee_{a\in G}\mathbf{K}_a\varphi$.

We can see that it is not dispersed knowledge that it is sunny in Groningen and in Liverpool (i.e. $g \wedge l$), because none of the agents know it. Therefore, the following holds at w_1:

$$\neg\mathbf{S}_{\{b,c\}}(g \wedge l).$$

It is, however, dispersed knowledge that it is sunny in Groningen, because Bert knows it; and that it is sunny in Liverpool, because Cat knows it. Thus, the following two formulas hold at w_1:

$$\mathbf{S}_{\{b,c\}}g \quad \text{and} \quad \mathbf{S}_{\{b,c\}}l.$$

General knowledge ($\mathbf{E}_G$): *What do all agents in a group know?*

General knowledge is usually defined as: $\mathbf{E}_G\varphi \overset{def}{\equiv} \bigwedge_{a\in G}\mathbf{K}_a\varphi$.

In our current scenario it is not general knowledge that it is sunny in Groningen, because Cat does not know it. We can argue analogously to show that it is not general knowledge that it is sunny in Liverpool. Thus, the following formulas hold at w_1:

$$\neg\mathbf{E}_{\{b,c\}}g \quad \text{and} \quad \neg\mathbf{E}_{\{b,c\}}l.$$

However, each of the agents know that is sunny in one of the two cities, then it is general knowledge that it is sunny either in Groningen or in Liverpool (i.e. $g \vee l$). Thus, the formula below holds at w_1:

$$\mathbf{E}_{\{b,c\}}(g \vee l).$$

Common knowledge ($\mathbf{C}_G$): *What do all the agents in a group know about what all of them know, and so forth up to infinite depth?*

A statement is common knowledge if it holds in all possible iterations of general knowledge. Therefore, common knowledge of φ is usually defined as the following infinite conjunction: $\mathbf{C}_G\varphi \overset{def}{\equiv} \bigwedge_{k=1}^{\infty}\mathbf{E}_G^k\varphi$.

In our example, $(\mathcal{M}_{Group}, w_1) \not\models \mathbf{E}_{\{b,c\}}\mathbf{E}_{\{b,c\}}(g \vee l)$. This follows from the fact that, the world w_2 is accessible from w_1 by Cat's accessibility relation and $(\mathcal{M}_{Group}, w_2) \not\models \mathbf{K}_b(g \vee l)$, i.e. $(\mathcal{M}_{Group}, w_1) \not\models \mathbf{K}_c\mathbf{K}_b(g \vee l)$. Therefore, it is not common knowledge that it is either sunny in Groningen or in Liverpool ($g \vee l$), i.e. the formula below holds at w_1:

$$\neg \mathbf{C}_{\{b,c\}}(g \vee l).$$

It is, however, common knowledge that Bert knows whether it is sunny in Groningen and that Cat knows whether it is sunny in Liverpool. The following two formulas hold at w_1:

$$\mathbf{C}_{\{b,c\}}(\mathbf{K}_b g \vee \mathbf{K}_b \neg g) \quad \text{and} \quad \mathbf{C}_{\{b,c\}}(\mathbf{K}_c l \vee \mathbf{K}_c \neg l).$$

We can see that for all worlds that we reach from w_1 in $\mathcal{M}_{Group}$, the formulas ($\mathbf{K}_b g \vee \mathbf{K}_b \neg g$) and ($\mathbf{K}_c l \vee \mathbf{K}_c \neg l$) hold there. We will see later, in section 2.4.2, that this characterization captures the notion of common knowledge.

2.4.1 General Knowledge

In this section we explore the notion of *general knowledge* using epistemic modal logic. A statement is general knowledge in a group if everybody in that group knows it.

Definition 2.4.1 (General knowledge in a group $\mathbf{E}_G$).
Let $\mathcal{L}_{\mathbf{KE}}(At, Ag) = \mathcal{L}(At, \{\mathbf{K}_a,\ \mathbf{E}_G \mid a \in Ag \text{ and } G \subseteq Ag\}, Ag)$ be the basic epistemic language extended with a general knowledge operator, $\mathbf{E}_G$, for each possible group of agents G. Let Ag be a set of agents, $G \subseteq Ag$ a group and $\varphi \in \mathcal{L}_{\mathbf{KE}}(At, Ag)$. The modal operator $\mathbf{E}_G$ is defined as follows:

$$\mathbf{E}_G \varphi \overset{def}{\equiv} \bigwedge_{a \in G} \mathbf{K}_a \varphi.$$

As mentioned in section 2.3.1, in $\mathcal{L}_{\mathbf{K}}$ we only consider a finite number of agents, thus our groups are of finite size, as well. This means, using the definition above, that we can translate any occurrence of the general knowledge into an equivalent well-formed basic epistemic modal formula. Therefore, adding general knowledge as a modal operator to our basic language does not make the language more expressive. However, having this operator in our language makes it exponentially more succinct than $\mathcal{L}_{\mathbf{K}}$ (more details in [vDHvdHK15]). In addition, later we use this notion to define common knowledge.

Example 2.7 (*General knowledge*).
Bert is informed by a reliable source, but not by Cat, that it is sunny in Liverpool.

After Bert being informed about the weather conditions in Liverpool it follows that both Bert and Cat know whether it is sunny in Liverpool, and particularly they know that it is sunny there. Therefore, this is general knowledge among both agents.

We can encode this new information with the formula:

Bert and Cat know that it is sunny in Liverpool: $\mathbf{K}_b l \wedge \mathbf{K}_c l \overset{def}{\equiv} \mathbf{E}_{\{b,c\}} l.$

The new Kripke model, updated from the example 2.5 with this new information, is presented below.

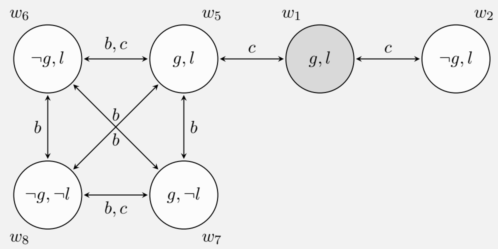

Figure 2.6: Kripke model for Bert and Cat's knowledge after Bert is informed about the weather in Liverpool.

It is important to note that Cat does not know that Bert knows that it is sunny in Liverpool. Therefore, it follows, from the definition of the operator $\mathbf{E}_G$ and other modal logic equivalences, that it is not general knowledge that the fact that is sunny in Liverpool is general knowledge:

It is not general knowledge that both Bert and Cat know that it is sunny in Liverpool: $\neg \mathbf{E}_{\{b,c\}} \mathbf{E}_{\{b,c\}} l$, *because* $\neg \mathbf{K}_c \mathbf{E}_{\{b,c\}} l.$

With this example we can see that the modal operator $\mathbf{E}_G$ does not satisfy the positive introspection property. It is easy to come up with an example to illustrate that $\mathbf{E}_G$ does not satisfy negative introspection, i.e. an epistemic scenario in which not everybody in a group knows p but all elements in the same group know that not everybody knows p.

An agent knows a statement iff it holds in all worlds that to him are epistemically distinguishable from the current world. The interpretation of general knowledge is similar. It considers, instead, all worlds that are indistinguishable to at least one of the agents.

Definition 2.4.2 (Accessibility relation for $\mathbf{E}_G$ [vDvdHK07])**.**
Let $\mathcal{F} = \langle \mathcal{W}, \{\mathcal{R}_a \mid a \in Ag\} \rangle$ be a frame and $G \subseteq Ag$ a group of agents, we define the *general knowledge accessibility relation* $\mathcal{R}_{\mathbf{E}_G}$ as:

$$\mathcal{R}_{\mathbf{E}_G} = \bigcup_{a \in G} \mathcal{R}_a.$$

In the lemma below we prove that the relation defined above characterizes correctly general knowledge as presented in the definition 2.4.1.

Lemma 2.4.1.
Let $\mathcal{M} = \langle \mathcal{W}, \{\mathcal{R}_a \mid a \in Ag\}, \mathcal{V}\rangle$ *be a Kripke model,* $G \subseteq Ag$ *a group and* $\varphi \in \mathcal{L}_{\boldsymbol{KE}}$*:*

$$(\mathcal{M}, w) \models \boldsymbol{E}_G\varphi \text{ iff for all } v \in \mathcal{W}, \text{ if } w\mathcal{R}_{\boldsymbol{E}_G}v \text{ then } (\mathcal{M}, v) \models \varphi.$$

Proof. We consider an arbitrary pointed model $(\mathcal{M} = \langle \mathcal{W}, \{\mathcal{R}_a \mid a \in Ag\}, \mathcal{V}\rangle, w)$ with $w \in \mathcal{W}$, and an arbitrary group $G \subseteq Ag$.

$$(\mathcal{M}, w) \models \mathbf{E}_G\varphi \quad \overset{\text{Def. 2.4.1 and 2.2.4}}{\Longleftrightarrow}$$

$$\text{for all } a \in G, (\mathcal{M}, w) \models \mathbf{K}_a\varphi \quad \overset{\text{Def. 2.2.4}}{\Longleftrightarrow}$$

$$\text{for all } a \in G, \text{ for all } v \in \mathcal{W} \text{ if there exists } (w, v) \in \mathcal{R}_a \text{ then } (\mathcal{M}, v) \models \varphi \quad \overset{\text{Gen. union}}{\Longleftrightarrow}$$

$$\text{for all } v \in \mathcal{W}, \text{ if } (w, v) \in \bigcup_{a \in G} \mathcal{R}_a \text{ then } (\mathcal{M}, v) \models \varphi \quad \overset{\text{Def. 2.4.2}}{\Longleftrightarrow}$$

$$\text{for all } v \in \mathcal{W}, \text{ if } (w, v) \in \mathcal{R}_{\mathbf{E}_G} \text{ then } (\mathcal{M}, v) \models \varphi$$

□

We interpret *iterated general knowledge* (see definition 2.2.2) over our epistemic models by means of *group reachability.* A statement is general knowledge up to iteration k iff it holds in all worlds reachable in k steps through the accessibility relations of the agents in the group. We will prove this correspondence in the visualization lemma below (lemma 2.4.2).

Definition 2.4.3 (Path and group reachability in a frame)**.**
A sequence $(s_0, s_1, \ldots, s_n)$ is a *path* π from s_0 to s_n (with size n) in the relation $\mathcal{R}$ iff $(s_i, s_{i+1}) \in \mathcal{R}$ for all $0 \leq i < n$. Each pair is called a *step.*

Let $\mathcal{F} = \langle \mathcal{W}, \{\mathcal{R}_a \mid a \in Ag\}\rangle$ be a frame and $G \subseteq Ag$ a group of agents. A state $v \in \mathcal{W}$ is *G-reachable (in* $\mathcal{F}$*)* from a state $w \in \mathcal{W}$ iff there exists a path from w to v in the relation $\bigcup_{a \in G} \mathcal{R}_a$.

Lemma 2.4.2 (Visualization for $\mathbf{E}^k_G$ [FHVM95])**.**
Let $\mathcal{M} = \langle \mathcal{W}, \{\mathcal{R}_a \mid a \in Ag\}, \mathcal{V}\rangle$ *be a Kripke model,* $w \in \mathcal{W}$ *a world,* $G \subseteq Ag$ *a group and* $\varphi \in \mathcal{L}_{\boldsymbol{KE}}(At, Ag)$*:*

$$(\mathcal{M}, w) \vDash \boldsymbol{E}^k_G\varphi \text{ iff for all } v \in \mathcal{W},$$
$$\text{if } v \text{ is } G\text{-reachable from } w \text{ in } k \text{ steps, then } (\mathcal{M}, v) \vDash \varphi.$$

Proof. We prove the statement by induction in k. We consider an arbitrary model $\mathcal{M} = \langle \mathcal{W}, \{\mathcal{R}_a \mid a \in Ag\}, \mathcal{V}\rangle$, $w \in \mathcal{W}$, group $G \subseteq Ag$ and formula $\varphi \in \mathcal{L}_{\mathbf{KE}}(At, Ag)$.

The base case, $k = 0$, follows by definition of iterated operator (2.2.2) and the fact that the world w is the only world reachable from itself in 0 steps: $(\mathcal{M}, w) \vDash \mathbf{E}^0_G\varphi$ iff $(\mathcal{M}, w) \vDash \varphi$.

We proceed to the induction step and we assume as induction hypothesis that:

$$(\mathcal{M}, w) \vDash \mathbf{E}_G^k \varphi' \text{ iff for all } v \in \mathcal{W},$$
$$\text{if } v \text{ is } G\text{-reachable from } w \text{ in } k \text{ steps, then } (\mathcal{M}, v) \vDash \varphi'. \qquad (\mathbf{IH})$$

By the definition of iterated operator (2.2.2) and the induction hypothesis (**IH**):

$$(\mathcal{M}, w) \vDash \mathbf{E}_G^{k+1} \varphi \text{ iff } (\mathcal{M}, w) \vDash \mathbf{E}_G^k \mathbf{E}_G \varphi \text{ iff for all } v \in \mathcal{W},$$
$$\text{if } v \text{ is } G\text{-reachable from } w \text{ in } k \text{ steps, then } (\mathcal{M}, v) \vDash \mathbf{E}_G \varphi. \qquad (*)$$

Consider an arbitrary $v \in \mathcal{W}$. Assume that v is G-reachable from w in k steps, i.e. there exists a path $\pi = (w, s_1, \ldots, s_{k-1}, v)$ in $\bigcup_{a \in G} \mathcal{R}_a$. By (*), $(\mathcal{M}, v) \vDash \mathbf{E}_G \varphi$, which is equivalent to $(\mathcal{M}, v) \vDash \mathbf{K}_a \varphi$ for all $a \in G$, by definition 2.4.1. Thus, by the truth interpretation of $\mathbf{K}_a$ (def. 2.2.4), by the definition of generalized union and by G-reachability (def. 2.4.3), this is equivalent to:

$$\text{for all } v' \in \mathcal{W}, \text{ if } v' \text{ is } G\text{-reachable from } v \text{ in 1 step, then } (\mathcal{M}, v') \vDash \varphi. \qquad (**)$$

Now, consider a $v' \in \mathcal{W}$ that is G-reachable from v in 1 step. Then, v' is G-reachable from w in $k+1$ steps, because there exists a path $\pi^{+1} = (w, s_1, \ldots, s_{k-1}, v, v')$ in $\bigcup_{a \in G} \mathcal{R}_a$ of length $k+1$ from w to v'. Finally, by (**), $(\mathcal{M}, v') \vDash \varphi$. Therefore, we proved that:

$$\text{for all } v' \in \mathcal{W}, \text{ if } v' \text{ is } G\text{-reachable from } w \text{ in } k+1 \text{ steps, then } (\mathcal{M}, v') \vDash \varphi. \qquad \square$$

Example 2.8 (*General knowledge distributes over conjunction*).
We we want to prove that: $\vDash \mathbf{E}_G(A \wedge B) \leftrightarrow (\mathbf{E}_G A \wedge \mathbf{E}_G B)$. Our goal with this proof is twofold: to show that the visualization lemma give us an intuitive interpretation of (iterated) general knowledge; and to prove a general knowledge property that will be useful in the next section on common knowledge.

If we consider the abbreviations in table 2.1 together with the classical interpretation of logical operators, then we need to prove for all Kripke models $\mathcal{M}$ that:

$$\mathcal{M} \vDash \mathbf{E}_G(A \wedge B) \text{ iff } \mathcal{M} \vDash \mathbf{E}_G A \wedge \mathbf{E}_G B.$$

Given the visualization lemma (lemma 2.4.2) and the interpretation of conjunction over Kripke models, it follows that:

$$\mathcal{M} \vDash \mathbf{E}_G(A \wedge B) \text{ iff for all } v, w \text{ worlds in } \mathcal{M},$$
$$\text{if } v \text{ is } G\text{-reachable from } w \text{ in 1 step, then } (\mathcal{M}, v) \vDash A \text{ and } (\mathcal{M}, v) \vDash B.$$

Using classical logic equivalences together with the visualization lemma, we can conclude that the previous is equivalent to $\mathcal{M} \vDash \mathbf{E}_G A \wedge \mathbf{E}_G B$.

2.4.2 Common Knowledge

General knowledge does not satisfy introspection properties (see example 2.7). This means that it can be the case that $\mathbf{E}^m_G\varphi$ holds but $\mathbf{E}^n_G\varphi$ doesn't. *Common knowledge* can be seen as a limiting notion for those iterations. In other words, all agents in a group know that all agents in the group know that ... the statement holds, *ad infinitum.*

The language $\mathcal{L}_{\mathbf{KC}} = \mathcal{L}(At, \{\mathbf{K}_a, \mathbf{C}_G \mid a \in Ag \text{ and } G \subseteq Ag\}, Ag)$ extends basic epistemic language with modal operators for common knowledge for each possible group of agents.

Definition 2.4.4 (Group common knowledge $\mathbf{C}_G$)**.**
Let At be a set of agents and $G \subseteq Ag$ a group of agents. We define *common knowledge* of $\varphi \in \mathcal{L}_{\mathbf{KC}}$ among the agents in group G by the following infinitary conjunction:

$$\mathbf{C}_G\varphi \overset{def}{\equiv} \bigwedge_{n=1}^{\infty} \mathbf{E}^n_G\varphi.$$

We do not include the iteration $\mathbf{E}^0_G\varphi$ in this conjunction, because this would force common knowledge to be veridical (only true statements could be common knowledge). This would be independent of whether the theory of knowledge we decide to work with is veridical, as well. We prefer to not impose such a restriction on common knowledge alone. In addition, it is easy to see that, if we assume knowledge to be veridical, then both general knowledge and common knowledge are veridical, as well.

In the basic epistemic modal language $\mathcal{L}_{\mathbf{K}}$ we can only have finite formulas. Thus, we cannot use the definition above to translate a formula of the type $\mathbf{C}_G\varphi$ in to an equivalent well-formed basic epistemic formula. It turns out that adding this operator to basic epistemic logic makes it more expressive (see [vDHvdHK15]).

We generalize below the visualization lemma for general knowledge (2.4.2) to common knowledge. Common knowledge of a formula φ must hold for any iteration of general knowledge about that sentence, thus the visualization lemma for common knowledge must hold for all path sizes. This semantic characterization will be useful later to prove the validity of some common knowledge properties.

Lemma 2.4.3 (Visualization for $\mathbf{C}_G$ [FHVM95])**.**
Let $\mathcal{M} = \langle \mathcal{W}, \{\mathcal{R}_a \mid a \in Ag\}, \mathcal{V}\rangle$ be a Kripke model, $w \in \mathcal{W}$ a world, $G \subseteq Ag$ a group and $\varphi \in \mathcal{L}_{\boldsymbol{KC}}(At, Ag)$ a formula:

$$(\mathcal{M}, w) \vDash \boldsymbol{C}_G\varphi \text{ iff for all } k \geq 1, \text{ for all } v \in \mathcal{W}, \text{ if } v \text{ is } G\text{-reachable from } w \text{ in } k \text{ steps, then } (\mathcal{M}, v) \vDash \varphi.$$

Example 2.9 (*Common knowledge*).
Bert has a call with Cat and she tells him that it is sunny in Liverpool.

After Cat shares the weather conditions of Liverpool with Bert, it becomes *common* *knowledge* between them that it is sunny in Liverpool. They do not consider it possible that one of them does not know that all of them know that ... it is sunny in Liverpool, independently of the number of iterations. The new Kripke model is presented below.

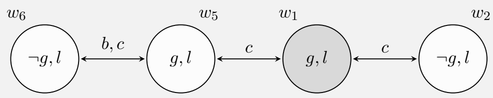

Figure 2.7: Kripke model for Bert and Cat's knowledge after it becomes common knowledge that it is sunny in Liverpool.

Using the visualization lemma (lemma 2.4.3) we can prove that the following formula holds at w_1:

It is common knowledge that it is sunny in Liverpool: $\mathbf{C}_{\{b,c\}} l$.

Note that all worlds are $\{b, c\}$-accessible in any number of steps from w_1, and l holds in all worlds in the model in figure 2.7.

The visualization lemma for this notion (lemma 2.4.3) tell us that we need to consider all G-reachable states to check if a given statement is commonly known between the agents in G. Therefore, we can interpret common knowledge as a $\Box$-operator over an accessibility relation that connects two worlds iff they are G-reachable in any number of steps. Given the definition of G-reachability (def. 2.4.3), it is easy to see that this relation must be the transitive closure of $\bigcup_{a \in G} \mathcal{R}_a$.

Definition 2.4.5 (Accessibility relation for $\mathbf{C}_G$ [vDvdHK07])**.**
Let $\mathcal{M} = \langle \mathcal{W}, \{\mathcal{R}_a \mid a \in Ag\}, \mathcal{V} \rangle$ be a Kripke model and $G \subseteq Ag$ a group of agents. The *common knowledge accessibility relation* $\mathcal{R}_{\mathbf{C}_G}$ is defined as the transitive closure of $\bigcup_{a \in G} \mathcal{R}_a$, as shown below:

$$\mathcal{R}_{\mathbf{C}_G} \stackrel{def}{\equiv} (\bigcup_{a \in G} \mathcal{R}_a)^+.$$

Lemma 2.4.4.
Let $\mathcal{M} = \langle \mathcal{W}, \{\mathcal{R}_a \mid a \in Ag\}, \mathcal{V} \rangle$ *be a Kripke model,* $w \in \mathcal{W}$ *a world and* $G \subseteq Ag$ *a group:*

$$(\mathcal{M}, w) \models \boldsymbol{C}_G \varphi \text{ iff for all } v \in \mathcal{W}, \text{ if } w \mathcal{R}_{\boldsymbol{C}_G} v, \text{ then } (\mathcal{M}, v) \models \varphi.$$

Example 2.10 (*Common knowledge as a limit to iterated general knowledge*).
In this example we prove two properties of common knowledge that clarify its definition as a limit to iterated general knowledge. The first property establishes that common knowledge can be interpreted as the fixed point of the boolean function

$f(\varphi) = \mathbf{E}_G(\varphi \wedge f(\varphi))$. In the second property, we have an inductive definition of common knowledge based on general knowledge.

(1) $\mathbf{C}_G\varphi \leftrightarrow \mathbf{E}_G(\varphi \wedge \mathbf{C}_G\varphi)$

Proof. Given the abbreviations in table 2.1 and the classical interpretation of logical operators, then we want to prove, for all Kripke models $\mathcal{M}$, that:

$$\mathcal{M} \vDash \mathbf{C}_G\varphi \text{ iff } \mathcal{M} \vDash \mathbf{E}_G(\varphi \wedge \mathbf{C}_G\varphi).$$

Let $\mathcal{M} = \langle \mathcal{W}, \{\mathcal{R}_a \mid a \in Ag\}, \mathcal{V}\rangle$ be a Kripke model and $w \in \mathcal{W}$:

$(\mathcal{M}, w) \vDash \mathbf{C}_G\varphi \overset{\text{lemma 2.4.3}}{\Leftrightarrow}$

for all $v \in \mathcal{W}$, for all $k \geq 1$,
if v is G-reachable from w in k steps, then $(\mathcal{M}, v) \vDash \varphi \;\Leftrightarrow$

for all $v \in \mathcal{W}$, if v is G-reachable from w in 1 steps, then $(\mathcal{M}, v) \vDash \varphi$ and
for all $k \geq 1$, if v is G-reachable from w in $k+1$ steps, then $(\mathcal{M}, v) \vDash \varphi \overset{\text{def. 2.4.2}}{\underset{\text{def. 2.4.3}}{\Leftrightarrow}}$

for all $v \in \mathcal{W}$, $(\mathcal{M}, w) \vDash \mathbf{E}_G\varphi$ and
for all $k \geq 1$, for all $v' \in \mathcal{W}$, if there exists a path $(w, s_1, s_2 \ldots, s_k, v)$
in $\bigcup_{a \in G} \mathcal{R}_a$, then $(\mathcal{M}, v) \vDash \varphi \overset{\text{path composition}}{\underset{\text{lemma 2.4.2}}{\Leftrightarrow}}$

$(\mathcal{M}, w) \vDash \mathbf{E}_G\varphi$ and $(\mathcal{M}, v) \vDash \mathbf{E}_G\mathbf{C}_G\varphi \overset{\text{classical logic}}{\Leftrightarrow}$

$(\mathcal{M}, w) \vDash \mathbf{E}_G\varphi \wedge \mathbf{E}_G\mathbf{C}_G\varphi \overset{\text{example 2.8}}{\Leftrightarrow} (\mathcal{M}, w) \vDash \mathbf{E}_G(\varphi \wedge \mathbf{C}_G\varphi)$ □

(2) $(\varphi \to \mathbf{E}_G(\varphi \wedge \psi)) \to (\varphi \to \mathbf{C}_G\psi)$

Proof. Let $\mathcal{M} = \langle \mathcal{W}, \{\mathcal{R}_a \mid a \in Ag\}, \mathcal{V}\rangle$ be a Kripke model, $\varphi, \psi \in \mathcal{L}_{\mathbf{KC}}$ and assume that **(a1)** $\mathcal{M} \vDash (\varphi \to \mathbf{E}_G(\varphi \wedge \psi))$. We want to prove that:

$$\mathcal{M} \vDash \varphi \to \mathbf{C}_G\psi.$$

We assume **(a2)** $\mathcal{M} \vDash \varphi$. By **(a1)**, it follows that **(*)** $\mathcal{M} \vDash \mathbf{E}_G(\varphi \wedge \psi)$. We will now assume, towards a contradiction, that $\mathcal{M} \nvDash \mathbf{C}_G\psi$. Then, by the visualization lemma (lemma 2.4.3), there exists a $k \geq 1$ and $w, v \in \mathcal{W}$ such that v is G-reachable from w in k steps and $(\mathcal{M}, v) \nvDash \psi$. Given the definition of G-reachability (def. 2.4.3), there exists a path $\pi = (w, s_1, \ldots, s_{k-1}, v)$ in $\bigcup_{a \in G} \mathcal{R}_a$, and in particular there exists a path π' of size 1 from s_{k-1} to v. Note that by **(*)**, $(\mathcal{M}, s_{k-1}) \vDash \mathbf{E}_G(\varphi \wedge \psi)$. Therefore, by the lemma 2.4.2 and the existence of π', $(\mathcal{M}, v) \vDash \varphi \wedge \psi$. This contradicts **(a2)**. □

2.5 Puzzle: Muddy Children

The Muddy children is a well-known epistemic puzzle, with many solutions presented in epistemic logic literature [vB04, vDvdHK07, GS11]. We work with the simple version of this puzzle presented by van Benthem in [vB04]:

> *After playing outside, two of three children have mud on their foreheads. They all see the others, but not themselves, so they do not know their own status. Now their father comes and says: 'At least one of you is dirty'. He then asks: 'Does anyone know if he is dirty?' The children answer truthfully. As this question–answer episode repeats, what will happen?*

An epistemic puzzle is solvable when there is a sequence of actions, which are consistent with the agents' knowledge, after which no agent is *uncertain* about the epistemic scenario described. In this puzzle, we use epistemic logic to discuss how each agent's knowledge justify the sequence of actions described and which actions are needed to solve it.

2.5.1 Epistemic Model For Muddy Children

We start with a Kripke model describing agents' initial knowledge about the epistemic scenario described. We consider three agents, $Ag = \{1, 2, 3\}$, which correspond to the children in the puzzle. The set of atomic propositions is $At = \{m_a \mid a \in Ag\}$, where m_a means that the child a has mud in the forehead. In our first model we consider 8 possible worlds, because each of the three kids can either have or not have mud in the forehead. All agents have common knowledge about each other epistemic state. This means that an agent knows whether other agent knows whether m_a holds, for all $a \in Ag$. Thus, we don't need to consider worlds representing alternative states of mind.

Note that in a first instance we consider the case that none of the children have mud in the forehead. We proceed now by translating the information in the puzzle into $\mathcal{L}_{\mathbf{KC}}$ formulas.

'They all see the others' − They know whether the others' forehead is dirty, and this is common knowledge between them:

$$\mathbf{C}_{Ag}\big(\bigwedge_{a\in Ag}\bigwedge_{\substack{b\in Ag\\ b\neq a}}(\mathbf{K}_a m_b \vee \mathbf{K}_a\neg m_b)\big) \tag{2.1}$$

'but not themselves' − They are ignorant about their own forehead, and this is common knowledge between them:

$$\mathbf{C}_{Ag}\big(\bigwedge_{a\in Ag}(\neg\mathbf{K}_a m_a \wedge \neg\mathbf{K}_a\neg m_a)\big) \tag{2.2}$$

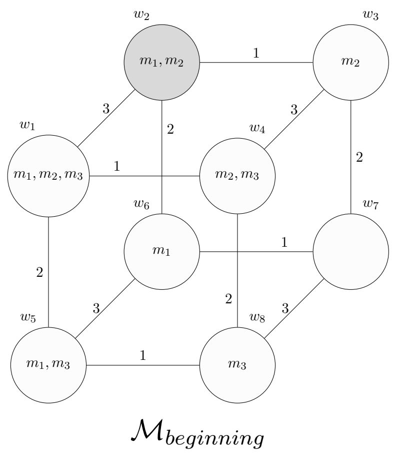

Figure 2.8: Muddy children: Kripke model for the beginning of the puzzle – $\mathcal{M}_{beginning}$.

In addition, it is easy to see that it is *general knowledge* that each child see at least one child with mud in the forehead. This follows from the number of children with a dirty forehead being greater than 1. Therefore, each of them know that at least one child is dirty. However, this is not common knowledge among the children.

$$\mathbf{E}_{Ag}(\bigvee_{b \in Ag} m_b) \tag{2.3}$$

$$\neg\mathbf{C}_{Ag}(\bigvee_{b \in Ag} m_b) \tag{2.4}$$

The Kripke model in Figure 2.8 shows an epistemic model for the children knowledge before their father's announcement. The epistemic formulas presented so far are satisfied by this model at world w_2, which we consider to be our current world. Recall that accessibility relations in our epistemic models are equivalence relations and to improve readability we do not draw reflexive and transitive arrows.

Announcement of 'At least one of you is dirty': With the father announcement what previously was just general knowledge becomes common knowledge:

$$\mathbf{C}_{Ag}(\bigvee_{b \in Ag} m_b) \tag{2.5}$$

This announcement will remove the world w_8 as a possible word for all our agents, because at the world w_8 we have that $(\mathcal{M}_{beginning}, w_8) \not\models (m_1 \vee m_2 \vee m_3)$ and as a consequence $(\mathcal{M}_{beginning}, w_2) \not\models \mathbf{C}_{Ag}(\bigvee_{b \in Ag} m_b)$. Therefore, we eliminate this world from the model $\mathcal{M}_{init}$. The new Kripke model is presented in Figure 2.9.

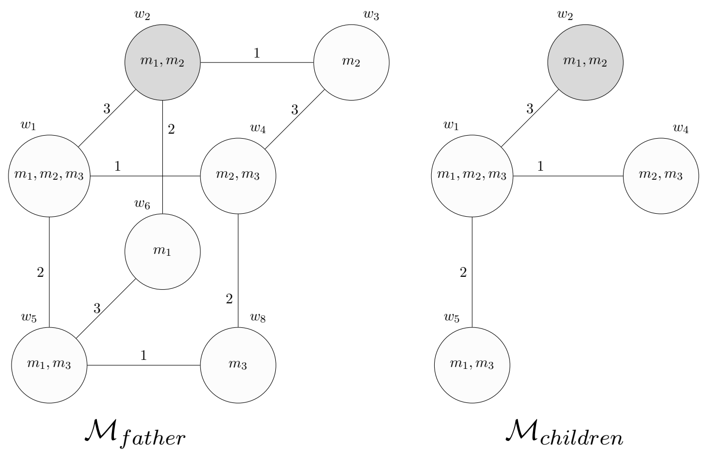

Figure 2.9: Kripke models after the father's announcement, $\mathcal{M}_{father}$, and after children's announcement of ignorance, $\mathcal{M}_{children}$, respectively.

First question *'Does anyone know if he is dirty?'*: No children answers affirmatively, because no child knows whether he is dirty. This means that they are ignorant about that. For all agents $a \in Ag$ the formula below is entailed on the model $\mathcal{M}_{father}$ at world w_2:

$$\neg\mathbf{K}_a m_a \wedge \neg\mathbf{K}_a \neg m_a \tag{2.6}$$

No children announces that he knows that he has mud in the forehead: Therefore, all children announce their ignorance about the state of their forehead. After this action it becomes common knowledge that all kids are ignorant about the state of their forehead:

$$\mathbf{C}_{Ag}(\bigwedge_{a \in Ag} (\neg\mathbf{K}_a m_a \wedge \neg\mathbf{K}_a \neg m_a)) \tag{2.7}$$

The worlds w_6, w_3 and w_8 are not consistent with this new information, because at those worlds the formulas $\mathbf{K}_1 m_1$, $\mathbf{K}_2 m_2$ and $\mathbf{K}_3 m_3$ hold, respectively. Therefore, the formula 2.7 does not hold on model $\mathcal{M}_{first}$ at world w_2. We eliminate those worlds from our model and obtain the model in Figure 2.9.

Second question *'Does anyone know if he is dirty?'*: Child 1 and child 2 answer affirmatively because $\mathbf{K}_1 m_1$ and $\mathbf{K}_2 m_2$ hold on model $\mathcal{M}_{children}$ at world w_2.

At this point the puzzle is solved. Two rounds of *'question-answer'* were necessary to reach this state. Note that all the formulas presented here (from 2.1 to 2.7) are independent of the number of the children. This epistemic modal specification can be applied to any instance of this puzzle. Moreover, it can be used to prove that the number of steps we need to solve the puzzle is same as the number of muddy children [vDvdHK07].

2.5.2 Discussion and Further Topics

This puzzle captures some interesting features that become clear after its analysis with epistemic logic. For example, after the first father's announcement a statement that was general knowledge becomes common knowledge. This breaks the Kripke model symmetry and makes the puzzle solvable. Or the fact that after the children announce their ignorance for the second time, they are not ignorant anymore.

Additionally, this puzzle is a good introduction to dynamic epistemic logic (see [vD-vdHK07]). While in basic epistemic logic we are concerned with static properties of knowledge, in dynamic epistemic logic we focus on *change* of knowledge. We model this change by transforming Kripke models (as we did in the puzzle's solution). It is important to note that such transformations do not change the content of the epistemic states. In fact, they only affect the agents view over these. The first dynamic epistemic logic, called *logic of public communication*, was presented in [Pla89]. The author considered communications done in discrete time between a given set of agents, and defined that if an agent a announces φ at time $t+1$ then at time t it becomes common knowledge among all the agents that a knew φ. In [GG97] the authors developed, independently, a logic for information change equivalent to *logic of public communication*. In *public announcement logic* finding out that φ is the case by a reliable source amounts to eliminate all the worlds that are not consistent with φ.

CHAPTER 3

Knowledge Axioms

In the previous chapter we introduced epistemic models, which are an helpful tool to analyze epistemic scenarios. In this chapter we present a *syntactic characterization of formal epistemic reasoning* to derive all valid formulas in such models. This means that we can derive all formulas that describe invariant knowledge properties with respect to the epistemic theory encoded in our models, and only these.

The purpose of this chapter is twofold: to make clear advantages and limitations of using deductive systems to capture formal epistemic reasoning; and, to get acquainted with the notions of *normal modal logic*, *canonical model*, *maximal consistent sets* and *completeness-via-canonicity*. These concepts are used, in the next chapter, in the context of a non-normal logic and applied to prove a new result. The material discussed in here is based on [BdRV01, FHVM95, vDvdHK07, vDHvdHK15]. In addition, we complement their material with some useful insights and carry out proofs that are left as exercise. For instance, proofs related to maximally consistent sets compositional properties are usually left to the reader. We decide to include them, because they are essential to understand the role of each system's component.

We start the chapter by looking at some epistemic principles that follow from the theory of knowledge implicit in our epistemic models. Afterwards, in section 3.2, we introduce a deductive system for basic epistemic logic, that we prove to be sound and complete for this class of models. In the next section we present a deductive system for basic epistemic logic extended with common knowledge, and prove that it is sound and complete for the class of all frames. The last section, section 3.2, is about Moorean sentences (i.e. sentences of the form *'p but I do not know p'*). Using the deductive system for basic epistemic logic we prove that is is inconsistent for an agent to announce such a sentence while referring to himself. This means that, within our theory of knowledge it is not consistent for an agent to know that *'p but I do not know p'*.

3.1 Epistemic Principles and Frame Properties

We identify epistemic principles that are valid in epistemic models. In order to do so, we use results from *modal correspondence theory*, as it establishes a *connection between modal formulas and frame properties*. This is not meant to be an in depth discussion of it, as this is not one of the main topics of this work. More information about frame correspondence can be found in the third chapter of [BdRV01] and in [vB01].

The formalism presented in this work is based on the realization that propositional knowledge can be interpreted as a necessity modality. We assume, in addition, that agents are *perfect reasoners* and that *they do not have computational limitations*. Given these two considerations, we adopt propositional modal logic as our base language and interpret its sentences over Kripke models. This idealized view on the agents reasoning abilities lead us to consider *logical omniscient* agents, i.e. they know all the logical consequences of their knowledge. This principle can be schematized by the following rule of inference[1]:

$$\frac{(\varphi_1 \wedge \ldots \wedge \varphi_n) \rightarrow \varphi}{(\mathbf{K}_a\varphi_1 \wedge \ldots \wedge \mathbf{K}_a\varphi_n) \rightarrow \mathbf{K}_a\varphi}\ R\mathcal{K} \quad \text{, for } n \geq 0.$$

This rule together with all propositional tautologies and modus ponens define a deductive system for the minimal normal modal logic $\mathcal{K}$ [Che80]. In the next section we specify a different system for $\mathcal{K}$. Nevertheless, both systems define the same logic [Che80] and they are both sound and complete for the class of all Kripke models.

We say that a formula or a set of formulas defines a property, like reflexivity, if it defines the class of frames satisfying that property in the following sense.

Definition 3.1.1 (Definability).
Let φ be a modal formula and $\mathbb{F}$ a class of frames. We say that φ defines (or characterizes) $\mathbb{F}$ if for all frames $\mathcal{F}$, then $\mathcal{F} \in \mathbb{F}$ iff $\mathcal{F} \vDash \varphi$.

In table 3.1 we list some formulas that have an interesting epistemic interpretation along with the frame property they define. These formulas can be combined in a modular way with the minimal modal logic $\mathcal{K}$ to specify the deductive systems listed in the table 3.2.

The axiom **T** captures the idea that only true things can be known. It is often used to distinguish knowledge from belief. The axiom **B** tells us that true sentences should always be known to be epistemically possible. Axioms **4** and **5** describe positive and negative introspective agents, respectively. This means that they describe agents that are able to know about their own knowledge state. It is interesting to note that, it is enough to add axioms **T** and **5** to $\mathcal{K}$ to derive all the other axioms in table 3.1. This shows the power of negative introspection. In fact, axiom **5** is rejected by most epistemologists [Hol13], while

[1] A rule depicts that, if all formulas on top of the line are theorems in the system considered, then the formula below the line is a theorem, as well.

Name	Formula	Epistemic interpretation	Property
Truth axiom	$\mathbf{K}_a\varphi \to \varphi$	Veridical	Reflexive
D axiom	$\neg\mathbf{K}_a\bot$	Consistent	Serial
4 axiom	$\mathbf{K}_a\varphi \to \mathbf{K}_a\mathbf{K}_a\varphi$	Positive introspection	Transitive
5 axiom	$\neg\mathbf{K}_a\varphi \to \mathbf{K}_a\neg\mathbf{K}_a\varphi$	Negative introspection	Euclidean
Brower axiom	$\varphi \to \mathbf{K}_a\hat{\mathbf{K}}_a\varphi$	Truth is epistemically possible	Symmetric

Table 3.1: Axioms with epistemic interpretation and the frame properties they define.

Name	Basic axioms	Other valid axioms	Constraints
$\mathcal{D}$	**D**	-	Serial
$\mathcal{K}45$	**4**,**5**	-	Transitive, euclidean
$\mathcal{KD}45$	**D**,**4**,**5**	-	Serial, transitive, euclidean
$\mathcal{T}$	**T**	**D**	Reflexive
$\mathcal{B}$	**T**,**B**	**D**	Reflexive, symmetric
$\mathcal{S}4$	**T**,**4**	**D**	Reflexive, transitive
$\mathcal{S}5$	**T**,**5**	**D**,**B**,**4**	Reflexive, symmetric, transitive

Table 3.2: Propositional systems with axioms from table 3.1 and the respective constraints they impose on frames' accessibility relations [GG06].

positive introspection, axiom **4**, is accepted by some of them. For instance, in [Hin62] axiom **4** is included in the system described, but not axiom **5**. However, the inclusion of **4** is not based on agent's introspective capabilities, but on the idea that an agent knowing a sentence ($\mathbf{K}_a\varphi$) differs only in words from an agent knowing that he knows that sentence ($\mathbf{K}_a\mathbf{K}_a\varphi$).

In this work, we define knowledge as a primitive modality in our language. We do not aim at analyzing what knowledge is in terms of something else. Instead, we want to evaluate what an agent or group of agents know, given a knowledge theory that can be encoded in terms of Kripke structures. Therefore, we do not advocate for any specific theory of knowledge. It is nevertheless relevant to understand the implications of our modeling assumptions over Kripke models, which we list as follows:

(M1) worlds encode possible configurations of the epistemic scenario being modeled, which are compatible with the agents current knowledge;

(M2) there is one accessibility relation for each agent;

(M3) the accessibility relations for the various agents are pairwise independent;

(M4) an agent's accessibility relation connects worlds that the agent is unable to distinguish, given its current knowledge.

Given these assumptions, accessibility relations in our epistemic models turn out to be equivalence relations. This corresponds to the system $\mathcal{S}5$ in table 3.2. We chose to work mostly within $\mathcal{S}5$, because it proved to be an adequate and successful approach for epistemic reasoning within computer science applications [FHVM95].

Finally, it is important to note that giving our modeling assumptions, we are working with *perfect introspective agents*. This is a reasoning ability that is not included in our initial assumptions. However, it was made explicit by the correspondence of axioms **4** and **5** to the transitivity and euclideaness of agents' accessibility relations. This is an example of how sound and complete axiomatizations can help us to detect hidden assumptions.

3.2 Basic Epistemic Logic

The main goal of this section is to present sound and complete axiomatizations for basic epistemic logic with respect to the class of all Kripke models ($\mathbb{K}_n$) and the class of our epistemic models ($\mathbb{S}5_n$).

3.2.1 Minimal Normal Modal Logic

A *normal modal logic* is a set of formulas satisfying some closure properties.

Definition 3.2.1 (Normal modal logic [BdRV01])**.**
Let $\mathcal{L}$ be a modal language. A *modal logic* Λ is a set of modal formulas, $\Lambda \subseteq \mathcal{L}$, that contains all propositional tautologies; and it is closed under *modus ponens* (i.e. if $\varphi \in \Lambda$ and $\varphi \to \psi \in \Lambda$, then $\psi \in \Lambda$) and *uniform substitution* (i.e. if $\varphi \in \Lambda$ then all of its substitution instances are in Λ as well).

A modal logic Λ is *normal* when it contains the formula $\mathcal{K}$ (i.e. $\Box(p \to q) \to (\Box p \to \Box q)$), and it is closed under *necessitation* (i.e. if $\Box\varphi \in \Lambda$, then $\varphi \in \Lambda$).

Example 3.1 (*Frames, models and normal modal logics*).
Validity is defined with respect to Kripke frames (def 2.2.5), because they abstract from any specific valuation. It is easy to see that given a modal language $\mathcal{L}$ and a class of frames $\mathbb{F}$, then $\mathbb{F}_{\mathcal{L}}$ (def 2.2.6) is a normal modal logic. This establishes a link between semantics, as defined previously, and normal modal logics.

It is worth to note that the same does not apply to Kripke models, i.e. if $\mathcal{M}$ is a Kripke model, then $\Lambda_{\mathcal{M}} = \{\varphi \mid \mathcal{M} \vDash \varphi\}$ is not a (normal modal) logic.

If we consider an arbitrary set of modal formulas Γ, then there is a smallest normal modal logic containing Γ. We call it the *logic generated (or axiomatised) by* Γ. The empty set generates the *minimal normal modal logic* $\mathcal{K}$, i.e. for any normal modal logic Λ we have $\mathcal{K} \subseteq \Lambda$.

3.2.2 System $\mathcal{K}_n$

Definition 3.2.2 (Minimal normal modal logic system $\mathcal{K}_n$)**.**
We denote by $\mathcal{K}_n$ the *Hilbert system for the minimal normal modal logic over* $\mathcal{L}_{\boldsymbol{K}}$. This system is defined by the following axiom and inference rules schemes[2], for each $a \in Ag$:

Taut. all instances of classic propositional logic tautologies;

$\mathcal{K}$**.** $\mathbf{K}_a(A \to B) \to (\mathbf{K}_a A \to \mathbf{K}_a B)$;

Modus Ponens. $\dfrac{A \qquad A \to B}{B}\ {\scriptstyle MP}$;

Necessitation. $\dfrac{A}{\mathbf{K}_a A}\ {\scriptstyle Nec}$.

The propositional fragment is denoted by $\mathcal{P}$ and defined by *Taut* and *Modus Ponens*.

We consider the usual definitions of proof and provable for a deductive system $\mathcal{X}$. A $\mathcal{X}$*-proof* or $\mathcal{X}$*-derivation* of a formula φ is a finite ordered list containing φ, such that each element is either an axiom of $\mathcal{X}$ or the application of one of the inference rules to formulas that appear earlier in the list. We sometimes refer to φ as a *theorem of* $\mathcal{X}$ or we say that φ is $\mathcal{X}$*-provable*, denoted by $\vdash_{\mathcal{X}} \varphi$. By $\nvdash_{\mathcal{X}} \varphi$ we denote that φ is not $\mathcal{X}$-provable . We refer to *propositional reasoning* in the context of a proof whenever we reach a conclusion by using only the system $\mathcal{P}$, i.e. classic propositional tautologies and modus ponens.

Example 3.2 ($\mathcal{K}$*-derivations*).
In this example we show common derivation patterns and present $\mathcal{K}$-proofs for some of them. We use these results later as lemmas in proofs.

N$\mathcal{K}$ **lemma:** We explore the use of $\mathcal{K}$ axiom together with the necessitation rule. The following lemma allow us to distribute $\mathbf{K}_a$ over an implication.

$$\frac{\vdash_{\mathcal{K}} A \to B}{\vdash_{\mathcal{K}} \mathbf{K}_a A \to \mathbf{K}_a B}\ {\scriptstyle N\mathcal{K}}$$

[2]Axioms and rules are presented as schemes because they can be instantiated by any element of $\mathcal{L}$. For example, the axiom $\mathcal{K}$ should be read as $\{\mathbf{K}_a(A \to B) \to (\mathbf{K}_a A \to \mathbf{K}_a B) \mid A, B \in \mathcal{L}\}$.

$\mathcal{K}$**-proof:**

1. $A \to B$

2. $\mathbf{K}_a(A \to B)$ — $Nec(1)$

3. $\mathbf{K}_a(A \to B) \to (\mathbf{K}_a A \to \mathbf{K}_a B)$ — $\mathcal{K}$

4. $\mathbf{K}_a A \to \mathbf{K}_a B$ — $MP(2,3)$

Propositional reasoning lemmas: We state below some lemmas that can be derived using only propositional reasoning.

$$\frac{\vdash_{\mathcal{K}} A \to B}{\vdash_{\mathcal{K}} \neg B \to \neg A}\ PL_1 \qquad \frac{\vdash_{\mathcal{K}} A \to B \quad \vdash_{\mathcal{K}} B \to C}{\vdash_{\mathcal{K}} A \to C}\ PL_2 \qquad \frac{\vdash_{\mathcal{K}} A \to (B \to C)}{\vdash_{\mathcal{K}} (A \wedge B) \to C}\ PL_3$$

$$\frac{\vdash_{\mathcal{K}} A \to B \quad \vdash_{\mathcal{K}} A \to C}{\vdash_{\mathcal{K}} A \to (B \wedge C)}\ PL_4 \qquad \frac{\vdash_{\mathcal{K}} A \to B \quad \vdash_{\mathcal{K}} C \to (D \wedge A)}{\vdash_{\mathcal{K}} C \to (D \wedge B)}\ PL_5$$

Distribution of $\Box$ over $\wedge$: We prove below a distribution law of $\Box$ over $\wedge$. It is interesting to note that, given our interpretation of knowledge as a necessity operator, this law states that an agent knowing the conjunction of two statements is equivalent to the agent knowing each of the statements in isolation.

$$\frac{}{\vdash_{\mathcal{K}} \mathbf{K}_a(A \wedge B) \leftrightarrow (\mathbf{K}_a A \wedge \mathbf{K}_a B)}\ Dist$$

$\mathcal{K}$**-proof** $\to$**:**

1. $(A \wedge B) \to A$ — $Taut$

2. $\mathbf{K}_a(A \wedge B) \to \mathbf{K}_a A$ — $N\mathcal{K}(1)$

3. $(A \wedge B) \to B$ — $Taut$

4. $\mathbf{K}_a(A \wedge B) \to \mathbf{K}_a B$ — $N\mathcal{K}(3)$

5. $\mathbf{K}_a(A \wedge B) \to (\mathbf{K}_a A \wedge \mathbf{K}_a B)$ — $PL_4(2,4)$

$\mathcal{K}$**-proof** $\leftarrow$**:**

1. $A \to (B \to (A \wedge B))$ — $Taut$

2. $\mathbf{K}_a A \to \mathbf{K}_a(B \to (A \wedge B))$ — $N\mathcal{K}(1)$

3. $\mathbf{K}_a(B \to (A \wedge B)) \to (\mathbf{K}_a B \to \mathbf{K}_a(A \wedge B))$ — $\mathcal{K}$

4. $\mathbf{K}_a A \to (\mathbf{K}_a B \to \mathbf{K}_a(A \wedge B))$ — $PL_2(2,3)$

5. $(\mathbf{K}_a A \wedge \mathbf{K}_a B) \to \mathbf{K}_a(A \wedge B)$ — $PL_3(4)$

***R*$\mathcal{K}$ lemma:** We present below the $\mathcal{K}$-proof for the following lemma:

$$\frac{\vdash_{\mathcal{K}} (A \wedge B) \rightarrow C}{\vdash_{\mathcal{K}} (\mathbf{K}_a A \wedge \mathbf{K}_a B) \rightarrow \mathbf{K}_a C} \; R\mathcal{K}$$

$\mathcal{K}$-proof:

1. $(A \wedge B) \rightarrow C$

2. $\mathbf{K}_a(A \wedge B) \rightarrow \mathbf{K}_a C$ — $N\mathcal{K}(1)$

3. $(\mathbf{K}_a A \wedge \mathbf{K}_a B) \rightarrow \mathbf{K}_a(A \wedge B)$ — *Dist*

4. $(\mathbf{K}_a A \wedge \mathbf{K}_a B) \rightarrow \mathbf{K}_a C$ — $PL_2(3, 2)$

A system is a set of formulas. We define a system by the axioms and inference rules that can be used to derive its elements. A system $\mathcal{X}$ is an *extension of a system* $\mathcal{X}'$ if $\mathcal{X}$ includes $\mathcal{X}'$ (i.e. $\mathcal{X}' \subseteq \mathcal{X}$). For example, the system $\mathcal{K}$ is an extension of $\mathcal{P}$. In the next section, we will look at all the possible extensions of $\mathcal{K}$ that are still consistent and complete, and use them to build a canonical model. Hence, we will review these notions below, to define *maximal consistent sets* in the end.

Definition 3.2.3 (Consistent and complete systems)**.**
Let $\mathcal{X}$ be a system. $\mathcal{X}$ is *consistent* iff $\nvdash_{\mathcal{X}} \bot$. $\mathcal{X}$ is *complete* for a language $\mathcal{L}$ iff for all $\varphi \in \mathcal{L}$, $\vdash_{\mathcal{X}} \varphi$ or $\vdash_{\mathcal{X}} \neg\varphi$.

Below we define consistency of a set of formulas Γ with respect to a system $\mathcal{X}$. We define this relative consistency by the consistency of the extension of $\mathcal{X}$ with Γ. If we consider a system $\mathcal{X}$ that includes classic propositional theorems (i.e. extensions of $\mathcal{P}$), then $\mathcal{X}$ is consistent iff for all formulas φ it is not the case that $\vdash_{\mathcal{X}} \varphi$ and $\vdash_{\mathcal{X}} \neg\varphi$. Given that derivations are finite, then $\mathcal{X}$ is consistent iff we cannot find two finite subsets of $\mathcal{X}$ that can be used to prove φ and its negation.

Definition 3.2.4 (Consistency with respect to $\mathcal{X}$)**.**
Let $\mathcal{X}$ be a system. A formula φ is *consistent with the system* $\mathcal{X}$ if we cannot derive $\neg\varphi$ in $\mathcal{X}$ ($\nvdash_{\mathcal{X}} \neg\varphi$). Likewise, a finite set of formulas $\{\varphi_1, \ldots, \varphi_n\}$ is consistent with the axiom system $\mathcal{X}$ if the conjunction $\varphi_1 \wedge \ldots \wedge \varphi_n$ is consistent with $\mathcal{X}$. An infinite set of formulas Γ is consistent with $\mathcal{X}$ if each finite subset of Γ is consistent with $\mathcal{X}$.

Example 3.3 (*Consistent set of formulas and Kripke models*).
Let $\mathcal{X}$ be an axiom system that is sound and complete for a class of frames $\mathbb{F}$. Then, for each Kripke model based on a frame $\mathcal{F} \in \mathbb{F}$ each of its states defines a $\mathcal{X}$ consistent set, i.e. $\{\varphi \mid (\mathcal{M}, w) \vDash \varphi\}$ is a $\mathcal{X}$-consistent.

Proof. Let $\Gamma = \{\varphi \mid (\mathcal{M}, w) \vDash \varphi\}$, then **(i)** $(\mathcal{M}, w) \vDash \varphi$ for all $\varphi \in \Gamma$. Assume towards a contradiction that there exists a finite subset $\{\varphi_1, \ldots, \varphi_n\}$ of Γ that is not $\mathcal{X}$-consistent. Then, $\vdash_{\mathcal{X}} \neg(\varphi_1 \wedge \ldots \wedge \varphi_n)$ and, by $\mathcal{X}$ being sound, $\vDash_{\mathbb{F}} \neg(\varphi_1 \wedge \ldots \wedge \varphi_n)$. This is a contradiction. By **(i)**, $(\mathcal{M}, w) \vDash \varphi_i$ for all $i \in \{1, \ldots, n\}$. Thus, $\mathcal{M}$ can be used as a counter-example for the validity of $\neg(\varphi_1 \wedge \ldots \wedge \varphi_n)$. □

The set defined above is, in fact, maximally consistent. This means that every proper superset is a $\mathcal{X}$ inconsistent set.

Definition 3.2.5 (Maximal consistent set of $\mathcal{X}$)**.**
Let $\mathcal{L}$ be a language, $\Phi \subseteq \mathcal{L}$ and $\mathcal{X}$ a system. A set of formulas $\Gamma \subseteq \Phi$, is a *maximal consistent set of* $\mathcal{X}$ *(in* Φ*)*, denoted $\mathcal{X}$-MCS, if:

- Γ is consistent with respect to $\mathcal{X}$; and
- for every formula $\varphi \in \Phi$, if $\varphi \notin \Gamma$ then $\Gamma \cup \{\varphi\}$ is not $\mathcal{X}$-consistent.

If not mentioned otherwise we assume that $\Phi = \mathcal{L}$.

Proposition 3.2.1 ($\mathcal{K}$-MCS is complete and deductively closed)**.**
Let Γ *be a* $\mathcal{K}$*-MCS and* φ *a modal formula.*

(1) $\varphi \in \Gamma$ *iff* $\neg\varphi \notin \Gamma$.

(2) *Let* $\{\varphi_1, \ldots, \varphi_n\} \subseteq \Gamma$. *If* $\vdash_{\mathcal{K}} (\varphi_1 \wedge \ldots \wedge \varphi_n) \to \varphi$, *then* $\varphi \in \Gamma$.

Proof. Assume that Γ is a $\mathcal{K}$-MCS.

(1) $\Rightarrow$: Consider an arbitrary $\varphi \in \Gamma$ and assume towards a contradiction that $\neg\varphi \in \Gamma$. We can prove $\vdash_{\mathcal{K}} \neg(\varphi \wedge \neg\varphi)$, because $\neg(\varphi \wedge \neg\varphi)$ is a propositional tautology. Therefore, $\{\varphi, \neg\varphi\} \subseteq \Gamma$ is not $\mathcal{K}$-consistent, which contradicts Γ being a $\mathcal{K}$-MCS.

$\Leftarrow$ Analogously by proving the contraposition: if $\neg\varphi \in \Gamma$ then $\varphi \notin \Gamma$

(2) Assume $\vdash_{\mathcal{K}} (\varphi_1 \wedge \ldots \wedge \varphi_n) \to \varphi$, for $\{\varphi_1, \ldots, \varphi_n\} \subseteq \Gamma$. Wsing propositional reasoning, we prove that $\vdash_{\mathcal{K}} \neg(\varphi_1 \wedge \ldots \wedge \varphi_n \wedge \neg\varphi)$. Therefore, $\neg\varphi \notin \Gamma$ because it would contradict the initial assumption that Γ is a $\mathcal{K}$-MCS. Then, by $\mathcal{K}$-MCS being complete (1), $\varphi \in \Gamma$. □

3.2.3 Soundness and Completeness of $\mathcal{K}$

In what follows we work with only one modal operator, i.e. with the system $\mathcal{K}$. The results in this section can straightforwardly be generalized to the multi-modal case, because none of the axioms nor rules use two or more distinct modal operators. This generalization turns out to be sufficient to capture *knowing that* in the multi-agent setting, due to our assumption that propositional knowledge of any pair of agents is mutually independent.

We will now check whether the system $\mathcal{K}$ fully characterizes $\mathbb{K}_{\mathcal{L}_{\mathbf{K}}}$. In this thesis we only look at weak soundness and completeness, as this is enough to establish a perfect match between the syntactic notion of logic (def. 3.2.1) and its corresponding Kripke-style semantics (def. 2.2.6).

Definition 3.2.6 (Soundness and completeness [vDHvdHK15]).
Let $\mathcal{L}$ be a language, $\mathcal{X}$ an axiom system and $\mathbb{S}$ a class of structures. We use $\vDash_{\mathbb{S}} \varphi$ to denote that for all structures $\mathcal{S} \in \mathbb{S}$, then $\mathcal{S} \vDash \varphi$.

$\mathcal{X}$ is *sound* with respect to $\mathbb{S}_{\mathcal{L}}$ if for all formulas $\varphi \in \mathcal{L}$,

$$\vdash_{\mathcal{X}} \varphi \quad \text{implies} \quad \vDash_{\mathbb{S}} \varphi.$$

$\mathcal{X}$ is *complete* with respect to $\mathbb{S}_{\mathcal{L}}$ if for all formulas $\varphi \in \mathcal{L}$,

$$\vDash_{\mathbb{S}} \varphi \quad \text{implies} \quad \vdash_{\mathcal{X}} \varphi.$$

If $\mathcal{L}$ is clear from the context, then we omit it and say that $\mathcal{X}$ is sound and complete with respect to $\mathbb{S}$.

The soundness of $\mathcal{K}$ for the class of all Kripke models[3] tells us that all theorems of $\mathcal{K}$ will, in fact, be valid formulas with respect to Kripke semantics. Soundness is easy to prove. In order to do so, we need to prove that all axioms are valid formulas and that all rules preserve validity, i.e. if we assume that the premises are valid then the conclusion given by the application of that rule is a valid formula as well.

The completeness of $\mathcal{K}$ for the class of all Kripke models establishes that we can derive all valid formulas. We will prove this result by contraposition. That is, given the class of all Kripke models $\mathbb{K}$ and the system $\mathcal{K}$, our goal is to prove that for all modal formulas φ:

$$\nvdash_{\mathcal{K}} \varphi \quad \text{implies} \quad \mathcal{M} \nvDash \varphi \text{ for some } \mathcal{M} \in \mathbb{K}.$$

We start by assuming that $\nvdash_{\mathcal{K}} \varphi$. Thus, by the definition of a formula being consistent with a system (def. 3.2.4), $\neg\varphi$ is consistent with $\mathcal{K}$. We can reformulate once again our statement, using the definition of $\mathcal{K}$-consistent formula together with the interpretation of formulas over Kripke models (def. 2.2.4), as follows:

$$\neg\varphi \text{ is consistent with } \mathcal{K} \quad \text{implies} \quad \mathcal{M} \vDash \neg\varphi \text{ for some } \mathcal{M} \in \mathbb{K}.$$

[3] We refer, interchangeably, to elements of a class of frames $\mathbb{F}$ as frames or as models (see def. 2.2.3).

The proposition below clarifies the relation between consistency and satisfiability for complete systems. Given this result we will proceed in our completeness proof by constructing a suitable model, called *canonical model*. In this model every consistent formula will be satisfiable in some of its states.

Proposition 3.2.2 (Consistency and satisfiability [BdRV01]).
An axiomatic system $\mathcal{X}$ is complete with respect to a class of frames $\mathbb{F}$ iff all $\mathcal{X}$-consistent formulas are satisfiable in some $\mathcal{F} \in \mathbb{F}$.

Proof. $\Rightarrow$ Assume that $\mathcal{X}$ is complete with respect to a class of frames $\mathbb{F}$, then (i) $\nvdash_{\mathcal{X}} \varphi$ implies $\nvDash \varphi$. Consider an arbitrary formula φ that is $\mathcal{X}$-consistent. This means that $\nvdash_{\mathcal{X}} \neg\varphi$ (by definition 3.2.4) and by our assumption (i) it follows that $\nvDash \neg\varphi$. Therefore, there exists $\mathcal{F} \in \mathbb{F}$ such that $\mathcal{F} \nvDash \neg\varphi$ and, by our interpretation of formulas over Kripke frames (def. 2.2.4), φ is satisfiable in $\mathcal{F}$ ($\mathcal{F} \vDash \varphi$).

$\Leftarrow$ Assume that all $\mathcal{X}$-consistent formulas are satisfiable in some $\mathcal{F} \in \mathbb{F}$. Now, consider an arbitrary φ that is $\mathcal{X}$-consistent ($\nvdash_{\mathcal{X}} \neg\varphi$). By our assumption, it follows that φ is satisfiable in some $\mathcal{F} \in \mathbb{F}$ ($\mathcal{F} \vDash \varphi$). By our interpretation of formulas over Kripke frames (def. 2.2.4) we have that $\mathcal{F} \nvDash \neg\varphi$. Thus, $\neg\varphi$ is not valid ($\mathcal{F} \nvDash \neg\varphi$ then $\nvDash \neg\varphi$). In conclusion, we just proved that given an arbitrary φ that is $\mathcal{X}$-consistent ($\nvdash_{\mathcal{X}} \neg\varphi$), then $\nvDash \neg\varphi$, i.e. $\mathcal{X}$ is complete with respect to $\mathbb{F}$. □

Recall that in the example 3.3 we mentioned that each state in a Kripke model induces a $\mathcal{K}$-MCS. In the canonical model each state will be identified by the $\mathcal{K}$-MCS it defines. We will use the canonical model to prove the truth lemma (lemma 3.2.4). This lemma states that if a formula belongs to a $\mathcal{K}$-MCS, then it is satisfiable in the canonical model at the state of that $\mathcal{K}$-MCS. Therefore, we need to guarantee that the states are coherently connected (lemma 3.2.3). This means that, if a $\mathcal{K}$-MCS includes the formula $\mathbf{K}_a\varphi$, then all the states connected to the state representing that $\mathcal{K}$-MCS must include the formula φ. We present below the canonical model definition.

Definition 3.2.7 (Canonical model for $\mathcal{X}$ (in Φ) [vDHvdHK15]).
Let $\mathcal{L}$ be a modal language and $\Phi \subseteq \mathcal{L}$. The canonical model for the axiom system $\mathcal{X}$ (in Φ) is the Kripke model $\mathcal{M}^{\mathcal{X}}_{\Phi} = \langle \mathcal{W}^{\mathcal{X}}_{\Phi}, \mathcal{R}^{\mathcal{X}}_a, \mathcal{V}^{\mathcal{X}} \rangle$ defined as follows:

- $\mathcal{W}^{\mathcal{X}}_{\Phi}$ is the set of all maximal consistent sets for $\mathcal{X}$ (in Φ);
- $\mathcal{R}^{\mathcal{X}}_a$ is a binary relation over $\mathcal{W}^{\mathcal{X}}_{\Phi}$, such that:

$$(\Gamma, \Delta) \in \mathcal{R}^{\mathcal{X}}_a \quad \text{iff} \quad \Gamma|_{\mathbf{K}_a} \subseteq \Delta, \quad \text{with } \Gamma|_{\mathbf{K}_a} = \{\varphi \mid \mathbf{K}_a\varphi \in \Gamma\};$$

- $\mathcal{V}^{\mathcal{X}}(\Gamma)(p) = \mathit{true}$ iff $p \in \Gamma$.

If not mentioned otherwise, we assume that $\Phi = \mathcal{L}$ and denote the canonical model as $\mathcal{M}^{\mathcal{X}} = \langle \mathcal{W}^{\mathcal{X}}, \mathcal{R}^{\mathcal{X}}_a, \mathcal{V}^{\mathcal{X}} \rangle$.

The following lemma tells us that every $\mathcal{K}$-consistent set is contained in a $\mathcal{K}$-MCS. This result guarantees that, in particular, every $\mathcal{K}$-consistent formula is in a $\mathcal{K}$-MCS.

Lemma 3.2.1 (Lindenbaum's lemma for $\mathcal{K}$ [BdRV01])**.**
If Σ is a $\mathcal{K}$-consistent set of formulas then there is an $\mathcal{K}$-MCS Σ^+ such that $\Sigma \subseteq \Sigma^+$.

Proof sketch. Let Σ be a $\mathcal{K}$-consistent set, and let the following be an enumeration of all the formulas in language $\mathcal{L}$: φ_1, φ_2,... . We define $\Sigma^+ = \bigcup_{n\in\mathbb{N}} \Sigma_n$, with each of the stages Σ_n defined as below:

$$\Sigma_0 = \Sigma$$
$$\Sigma_{n+1} = \begin{cases} \Sigma_n \cup \{\varphi_{n+1}\} & \text{if it is } \mathcal{K}\text{-consistent,} \\ \Sigma_n \cup \{\neg\varphi_{n+1}\} & \text{otherwise.} \end{cases}$$

It is easy to see that Σ^+ is $\mathcal{K}$-consistent, because our construction guarantees that all of its finite subsets are $\mathcal{K}$-consistent. In addition, note that we iterated through all formulas of $\mathcal{L}$ to build Σ^+. Therefore, no proper extension of Σ^+ is $\mathcal{K}$-consistent, as it would make some stage $\mathcal{K}$-inconsistent. Therefore, Σ^+ is a $\mathcal{K}$-MCS. ☐

Below we prove $\mathcal{K}$-MCS decomposition properties for boolean operators, using the fact that $\mathcal{K}$ contains all classic propositional tautologies and admits modus ponens.

Lemma 3.2.2 ($\mathcal{K}$-MCS and boolean reasoning)**.**
Let Γ be a $\mathcal{K}$-MCS and φ, φ' be modal formulas. Then, $\varphi \wedge \varphi' \in \Gamma$ iff $\varphi \in \Gamma$ and $\varphi' \in \Gamma$.

Proof. Assume that Γ is a $\mathcal{K}$-MCS.

$\Rightarrow$: Consider an arbitrary $\varphi \wedge \varphi' \in \Gamma$ and assume towards a contradiction that $\varphi \notin \Gamma$ or $\varphi' \notin \Gamma$. We will proceed by cases.

If $\varphi \notin \Gamma$, then $\neg\varphi \in \Gamma$, by $\mathcal{K}$-MCS being complete (prop. 3.2.1). As $\neg((\varphi \wedge \varphi') \wedge \neg\varphi)$ is a propositional tautology, it follows $\vdash_{\mathcal{K}} \neg((\varphi \wedge \varphi') \wedge \neg\varphi)$. Therefore, $\{\varphi \wedge \varphi', \neg\varphi\}$ is not $\mathcal{K}$-consistent, which is a contradiction.

We can prove, analogously, that $\varphi' \notin \Gamma$ cannot be the case.

$\Leftarrow$: Similar proof, using the tautology $\neg(\varphi \wedge \varphi' \wedge \neg(\varphi \wedge \varphi'))$. ☐

The following lemma proves that worlds in the canonical model are coherently connected. The interesting part of the proof is to prove that if a $\mathcal{K}$-MCS includes the formula $\mathbf{K}_a\varphi$, then all the states connected to the state representing that $\mathcal{K}$-MCS must include the formula φ. We follow the same strategy as in the previous lemma. We prove it by contraposition and use $\mathcal{K}$ axiom and *N$\mathcal{K}$* rule.

Lemma 3.2.3 (Coherently connected states)**.**
Let Γ be $\mathcal{K}$-MCS and φ a modal formula. Then, $\boldsymbol{K}_a\varphi \in \Gamma$ iff for all Δ that are $\mathcal{K}$-MCS if $(\Gamma, \Delta) \in \mathcal{R}_a^{\mathcal{K}}$ then $\varphi \in \Delta$.

Proof. Let Γ be a $\mathcal{K}$-MCS.

$\Rightarrow$ We assume that $\mathbf{K}_a\varphi \in \Gamma$ and we consider an arbitrary Δ such that $(\Gamma, \Delta) \in \mathcal{R}_a^{\mathcal{K}}$. Then, by definition of $\Gamma|_{\mathbf{K}_a}$ and $\mathcal{R}_a^{\mathcal{K}}$ in 3.2.7, it follows that $\varphi \in \Delta$.

$\Leftarrow$ We will prove this direction by contraposition. We start by assuming that $\mathbf{K}_a\varphi \notin \Gamma$, and by lemma 3.2.1 it follows that:

$$\neg\mathbf{K}_a\varphi \in \Gamma. \qquad (\mathbf{a1})$$

Our goal is to prove that:

$$\text{there exists a } \mathcal{K}\text{-MCS } \Delta \text{ such that } (\Gamma, \Delta) \in \mathcal{R}_a^{\mathcal{K}} \text{ and } \varphi \notin \Delta. \qquad (*)$$

By proposition 3.2.1 and the definition of $\mathcal{R}_a^{\mathcal{K}}$, our goal (*) is equivalent to:

there exists a $\mathcal{K}$-MCS Δ such that $\{\psi \mid \mathbf{K}_a\psi \in \Gamma\} \subseteq \Delta$ and $\neg\varphi \in \Delta$ $\overset{\text{Def. set inclusion}}{\Leftrightarrow}$

there exists a $\mathcal{K}$-MCS Δ such that $\{\psi \mid \mathbf{K}_a\psi \in \Gamma\} \cup \{\neg\varphi\} \subseteq \Delta$ $\underset{\text{lemma 3.2.1}}{\overset{\text{Def. } \mathcal{K}\text{-MCS (3.2.5) and}}{\Leftrightarrow}}$

$\{\psi \mid \mathbf{K}_a\psi \in \Gamma\} \cup \{\neg\varphi\}$ is $\mathcal{K}$-consistent

We will now proceed by proving the following equivalent statement of (*):

$$\{\psi \mid \mathbf{K}_a\psi \in \Gamma\} \cup \{\neg\varphi\} \text{ is } \mathcal{K}\text{-consistent.} \qquad (**)$$

Assume towards a contradiction that $\{\psi \mid \mathbf{K}_a\psi \in \Gamma\} \cup \{\neg\varphi\}$ is not $\mathcal{K}$-consistent, then:

$$\text{there exists } \{\mathbf{K}_a\psi_1, \ldots, \mathbf{K}_a\psi_n\} \in \Gamma \text{ and } \vdash_{\mathcal{K}} \neg(\psi_1 \wedge \ldots \wedge \psi_n \wedge \neg\varphi). \qquad (\mathbf{a2})$$

By propositional reasoning, it follows that:

$$\vdash_{\mathcal{K}} (\psi_1 \wedge \ldots \wedge \psi_n) \to \varphi.$$

In example 3.2 we proved the lemma $R\mathcal{K}$. This lemma can be generalized to:

$$\frac{\vdash_{\mathcal{K}} (\psi_1 \wedge \ldots \wedge \psi_n) \to \varphi}{\vdash_{\mathcal{K}} (\mathbf{K}_a\psi_1 \wedge \ldots \wedge \mathbf{K}_a\psi_n) \to \mathbf{K}_a\varphi}\ .$$

Using propositional reasoning:

$$\frac{\vdash_{\mathcal{K}} (\psi_1 \wedge \ldots \wedge \psi_n) \to \varphi}{\vdash_{\mathcal{K}} \neg(\mathbf{K}_a\psi_1 \wedge \ldots \wedge \mathbf{K}_a\psi_n \wedge \neg\mathbf{K}_a\varphi)}\ .$$

So, $\{\mathbf{K}_a\psi_1, \ldots, \mathbf{K}_a\psi_n, \neg\mathbf{K}_a\varphi\}$ is not $\mathcal{K}$-consistent. This contradicts the assumption that Γ is $\mathcal{K}$-MCS, because from (**a1**) and (**a2**) it follows $\{\mathbf{K}_a\psi_1, \ldots, \mathbf{K}_a\psi_n, \neg\mathbf{K}_a\varphi\} \subseteq \Gamma$. $\square$

We have all the ingredients to prove the connection between a formula φ being satisfiable in the canonical model at some $\mathcal{K}$-MCS Γ and φ being an element of Γ.

Lemma 3.2.4 (Truth Lemma)**.**
Let φ be a modal formula and Γ a $\mathcal{K}$-MCS. $(\mathcal{M}^{\mathcal{K}}, \Gamma) \vDash \varphi$ iff $\varphi \in \Gamma$.

Proof. We prove this lemma by induction on the structure of an arbitrary modal formula φ. The base case, $(\mathcal{M}^{\mathcal{K}}, \Gamma) \vDash p$ iff $p \in \Gamma$ for a propositional variable p, follows from the valuation function defined in 3.2.7 and the interpretation of propositional variables using Kripke semantics (see def. 2.2.4).

Consider arbitrary modal formula φ. Then, for all $\mathcal{K}$-MCS Γ we assume as induction hypothesis:

$$(\mathcal{M}^{\mathcal{K}}, \Gamma) \vDash \varphi \text{ iff } \varphi \in \Gamma. \tag{IH}$$

The inductive cases with boolean operators as the topmost operator follow from lemma 3.2.2 and their interpretation under Kripke semantics (see def. 2.2.4).

The inductive case $\mathbf{K}_a\varphi$ is proved as follows. Assume for arbitrary Δ that:

$$\begin{aligned}
&(\mathcal{M}^{\mathcal{K}}, \Delta) \vDash \mathbf{K}_a\varphi \quad \overset{\text{Kripke semantics (def. 2.2.4)}}{\Leftrightarrow} \\
&(\mathcal{M}^{\mathcal{K}}, \Delta') \vDash \varphi \text{ for all } \Delta' \text{ such that } (\Delta, \Delta') \in \mathcal{R}_a^{\mathcal{K}} \quad \overset{IH}{\Leftrightarrow} \\
&\varphi \in \Delta' \text{ for all } \Delta' \text{ such that } (\Delta, \Delta') \in \mathcal{R}_a^{\mathcal{K}} \quad \overset{\text{lemma 3.2.3}}{\Leftrightarrow} \\
&\mathbf{K}_a\varphi \in \Delta
\end{aligned}$$

□

Theorem 3.2.5 ($\mathcal{K}$ soundness and completeness w.r.t. $\mathbb{K}$)**.**
$\mathcal{K}$ is complete with respect to the class of all Kripke models $\mathbb{K}$.

Proof. Soundness is straightforward. The completeness proof is presented below.

By Lindenbaum's lemma (lemma 3.2.1) we know that every $\mathcal{K}$-consistent formula will be part of a $\mathcal{K}$-MCS. It then follows, by the truth lemma 3.2.4, that every consistent formula will be satisfiable in the canonical model at the state representing a $\mathcal{K}$-MCS it belongs to. We summarize this below:

$$\mathcal{K}\text{-consistent } \varphi \xrightarrow{\text{lemma 3.2.1}} \text{exists a } \mathcal{K}\text{-MCS } \Gamma \text{ and } \varphi \in \Gamma \xrightarrow{\text{lemma 3.2.4}} (\mathcal{M}^{\mathcal{K}}, \Gamma) \vDash \varphi$$

Therefore, we can use the canonical model to prove that every $\mathcal{K}$-consistent formula is satisfiable in some Kripke model and, by proposition 3.2.2, conclude that $\mathcal{K}$ is complete with respect to the class of all Kripke models $\mathbb{K}$. □

3.2.4 Soundness and Completeness of $\mathcal{S}5$

The system $\mathcal{S}5$ (table 3.2) extends $\mathcal{K}$ with the **T** axiom ($\mathbf{K}_a\varphi \to \varphi$) and the **5** axiom ($\neg\mathbf{K}_a\varphi \to \mathbf{K}_a\neg\mathbf{K}_a\varphi$). All previous results for the minimal normal modal logic $\mathcal{K}$ are applicable to $\mathcal{S}5$, because this system contains all of its rules and axioms. The theorem 3.2.5 refers only to the class of all frames, so we still need to prove that $\mathcal{S}5$ is complete with respect to the class $\mathbb{S}5$. One way to do it is by proving that the canonical model is a member of $\mathbb{S}5$.

A theorem φ of $\mathcal{X}$ is *canonical for a frame property* P, when φ is both valid in all frames that satisfy P and valid in the canonical model of φ. This means that φ, besides defining the class of frames characterized by P, forces the canonical model of $\mathcal{X}$ to satisfy the structural property P. Recall that, in section 3.1, we discussed *frame definability* and listed some formulas with the frame property they define (table 3.1). Those formulas are, as well, canonical for the frame property they define [BdRV01].

Definition 3.2.8 (Canonicity for a property [BdRV01])**.**
Let φ be a modal formula and P be a frame property. Let $\mathcal{X}$ be a system such that $\varphi \in \mathcal{X}$. If the canonical model for $\mathcal{X}$ is based on a frame that satisfies property P and φ is valid in any classes of frames with property P, then φ is canonical for P.

Example 3.4 (*Axiom **T** forces reflexivity in canonical model*).
We want to prove that the canonical model for normal modal logic extended with the axiom **T** is reflexive.

Consider an arbitrary normal model logic $\mathcal{K}X$, such that:

$$\mathbf{K}_a\varphi \to \varphi \in \mathcal{K}X. \tag{*}$$

Let $\mathcal{M}^{\mathcal{K}X} = \langle \mathcal{W}^{\mathcal{K}X}, \mathcal{R}_a^{\mathcal{K}X}, \mathcal{V}^{\mathcal{K}X} \rangle$ be the canonical model of $\mathcal{K}X$ and $\Gamma \in \mathcal{W}^{\mathcal{K}X}$ a world in this model. Assume that $\mathbf{K}_a\varphi \in \Gamma$. Γ is a $\mathcal{K}X$-MCS, so it is easy to see that (*), implies that $\mathbf{K}_a\varphi \to \varphi \in \Gamma$. Then, by proposition 3.2.1 and $\varphi \in \Gamma$, it follows $\varphi \in \Gamma$. Finally, by canonical model's definition (def. 3.2.7), $(\Gamma, \Gamma) \in \mathcal{R}_a^{\mathcal{K}X}$. This means that $\mathcal{R}_a^{\mathcal{K}X}$ is reflexive.

The completeness of $\mathcal{S}5$ follows from the canonicity of the axioms **T** and **5** for reflexive and euclidean frames, respectively. This proof technique is called *completeness-via-canonicity*.

Theorem 3.2.6 ($\mathcal{S}5$ soundness and completeness w.r.t. $\mathbb{S}5$)**.**
$\mathcal{S}5$ is complete with respect to the class of frames $\mathbb{S}5$.

Proof. Soundness follows from both axiom **T** and **5** being valid in the class $\mathbb{S}5$.

To prove completeness, by proposition 3.2.2, we need to prove that all $\mathcal{S}5$-consistent formulas are satisfiable in some model based on a frame of the class $\mathbb{S}5$.

Let φ be an arbitrary $\mathcal{S}5$-consistent formula. By lemma 3.2.1 we know that there exists a $\mathcal{S}5$-MCS Γ such that $\varphi \in \Gamma$. We need to prove that: (i) φ is satisfied in the canonical model $\mathcal{M}^{\mathcal{S}5} = \langle \mathcal{W}^{\mathcal{S}5}, \mathcal{R}_a^{\mathcal{S}5}, \mathcal{V}^{\mathcal{S}5} \rangle$ at the state Γ; (ii) the accessibility relation $\mathcal{R}_a^{\mathcal{S}5}$ is an equivalence relation. The first part (i) follows from the truth lemma. The second (ii) follows from the fact that the **T** and **5** axioms are canonical for reflexive and euclidean frames, respectively. Therefore, the accessibility relation in the canonical model, $\mathcal{R}_a^{\mathcal{S}5}$, is an equivalence relation. □

3.3 Group Knowledge

In this section we present a finite axiomatization for basic epistemic logic extended with common knowledge. This contrasts with the infinite nature of its semantic interpretation presented in the previous chapter (def. 2.4.4). The system we introduce is based on the view of common knowledge as the greatest fixed-point of what everybody in a group knows.

3.3.1 Normal modal logic with common knowledge $\mathcal{KC}$

In section 2.4.2 (see example 2.10) we proved that the formulas below are valid:

$$\mathbf{C}_G\varphi \leftrightarrow (\varphi \wedge \mathbf{E}_G\mathbf{C}_G\varphi) \qquad \text{(Fixed-point)}$$
$$(\varphi \to \mathbf{E}_G(\varphi \wedge \psi)) \to (\varphi \to \mathbf{C}_G\psi) \qquad \text{(Induction)}$$

It turns out that these properties are all we need to extend the syntactic characterization of basic epistemic logic to accommodate common knowledge.

Definition 3.3.1 (Minimal normal modal logic with common knowledge $\mathcal{KC}$**).**
Let $\mathcal{L}_{\mathbf{KC}}$ be the basic epistemic language with common knowledge and $n = |Ag|$. The system $\mathcal{KC}$ is the minimal modal logic $\mathcal{K}_n$ extended with the following axioms and rule schemes:

General Knowledge (AE). $\mathbf{E}_G A \leftrightarrow \bigwedge_{a\in G} \mathbf{K}_a A$;

Fixed-Point Axiom (FP). $\mathbf{C}_G A \leftrightarrow \mathbf{E}_G(A \wedge \mathbf{C}_G A)$;

Induction Rule (IC). $$\frac{A \to \mathbf{E}_G(C \wedge B)}{A \to \mathbf{C}_G B}\ IC.$$

We prove, in theorem 3.3.6, that $\mathcal{KC}$ is sound and complete with respect to the class of all Kripke models, for the epistemic language $\mathcal{L}_{\mathbf{KC}}$. This proof follows the same structure as the proof presented earlier for basic epistemic logic (theorem 3.2.5). The main difference is that the maximally consistent sets used in the canonical model need to be of finite size, because $\mathbb{K}_{\mathcal{L}_{\mathbf{KC}}}$ is not compact. We can see this with the set below:

$$\mathbf{S} = \{\mathbf{E}_G^n\varphi \mid n \in \mathbb{N}\} \cup \{\neg\mathbf{C}_G\varphi\}.$$

If the group G has two or more agents, then all $\mathbf{S}$ finite subsets are satisfiable while $\mathbf{S}$ is not satisfiable.

Example 3.5 (*Compactness and Canonical Model*).
Why do we need a compact logic to use the canonical model, as in definition 3.2.7, to prove a completeness result?

We proved that the minimal modal logic $\mathcal{K}$ is complete with respect to the class of all Kripke models (theorem 3.2.5), by means of a truth lemma (lemma 3.2.4). This lemma tell us that the canonical model, which consists of coherently connected $\mathcal{K}$ maximally consistent sets ($\mathcal{K}$-MCS), satisfies a given formula iff the formula belongs to some $\mathcal{K}$-MCS. Now, recall that an infinite set of formulas is consistent with respect to an axiomatic system (def. 3.2.4) iff each of its finite subsets is consistent, as well.

Let $\mathcal{X}$ be a system that we want to prove to be complete with respect to a semantic characterization that is not compact. We assume that we follow the same structure as in the proof of the theorem 3.2.5. In particular, we assume that we proved already (i) soundness, (ii) the Lindenbaum's lemma for $\mathcal{X}$ and (iii) a truth lemma. Let S be an infinite set of its sentences that is not satisfiable while its finite subsets are. By soundness (i), all of S finite sets are $\mathcal{X}$-consistent and thus, S is $\mathcal{X}$-consistent as well. Therefore, given the Lindenbaum's lemma for $\mathcal{X}$ (ii), S is a set Γ that is a $\mathcal{X}$-MCS. We could, then, use the truth lemma (iii) to show that S is satisfied in the canonical model at Γ. This is a contradiction, because there should be no model for that set.

Instead of looking at all well-formed formulas in the language to build a large canonical model, which satisfies all consistent formulas, we restrict our attention to a finite set of formulas defined with respect to the formula we want to prove satisfiable. We call this set the *closure* of that formula (def. 3.3.2). The maximally consistent sets, used as the building blocks of the canonical model, are subsets of it. Therefore, we must ensure that it has all formulas we need in the maximally consistent sets to prove a truth lemma later. This means that, the closure of a formula φ should include, for instance, all subformulas of φ and all formulas that can be obtained from these by the application of inference rules of the system we want to prove to be complete.

Definition 3.3.2 (Closure of φ for $\mathcal{KC}$ [FHVM95])**.**
Let $\varphi \in \mathcal{L}_{\mathbf{KC}}$ and $Sub(\varphi)$ the set of all subformulas of φ. We denote by $Cl_C(\varphi)$ the closure of φ for $\mathcal{L}_{\mathbf{KC}}$, as the smallest set defined inductively as follows:

- if $\psi \in Sub(\varphi)$, then $\psi \in Cl_C(\varphi)$;
- if $\mathbf{C}_G\psi \in Cl_C(\varphi)$, then $\{\mathbf{E}_G(\psi \wedge \mathbf{C}_G\psi),\ \psi \wedge \mathbf{C}_G\psi,\ \mathbf{K}_a(\psi \wedge \mathbf{C}_G\psi),\ \mathbf{K}_a\psi\} \in Cl_C(\varphi)$, for all $a \in G$;
- if $\mathbf{E}_G\psi \in Cl_C(\varphi)$, then $\mathbf{K}_a\psi \in Cl_C(\varphi)$, for all $a \in G$;

- if $\psi \in Cl_C(\varphi)$ and ψ is not negated, then $\neg\psi \in Cl_C(\varphi)$.

It is easy to prove that the closure as defined above is a finite set. It follows, then, that all $\mathcal{KC}$-MCS in $Cl_C(\varphi)$ (def. 3.2.5), for any formula $\varphi \in \mathcal{L}_{\mathbf{KC}}$, are finite sets and that there are only a finite number of such sets. Thus, the $\mathcal{KC}$'s canonical model for φ in $Cl_C(\varphi)$ (def. 3.2.7) is of finite size, too.

We define below the encoding of $\mathcal{KC}$-MCS into a formula, which can be used to define an encoding with all the canonical model worlds that include a given formula φ. We call the later *characteristic of* φ .

Definition 3.3.3 ($\mathcal{X}$-MCS encoding into a formula [FHVM95])**.**
Let Γ be a $\mathcal{X}$-MCS in a finite set Φ. We denote by $\underline{\Gamma}$ the encoding of Γ into a formula, defined as $\underline{\Gamma} \overset{def}{\equiv} \bigwedge_{\varphi\in\Gamma} \varphi$.

Definition 3.3.4 (Characteristic of formula [FHVM95])**.**
Let φ be a modal formula, $\mathcal{M}$ a canonical model in a finite set Φ and $\mathcal{W}$ the set of its worlds. We denote by $[\varphi]^{\mathcal{M}}$ the *set of all worlds in* $\mathcal{M}$ *that satisfies* φ, which is defined as follows:

$$[\varphi]^{\mathcal{M}} \overset{def}{=} \{\Delta \in \mathcal{W} \mid (\mathcal{M}, \Delta) \vDash \varphi\}.$$

We define the *formula that characterizes* φ in that model $\mathcal{M}$ as follows:

$$\chi_{\varphi}^{\mathcal{M}} \overset{def}{\equiv} \bigvee_{\Delta\in[\varphi]^{\mathcal{M}}} \underline{\Delta}.$$

If $\mathcal{M}$ is clear from the context, then we denote them by $[\varphi]$ and χ_{φ}, respectively.

Given that our canonical model is finite, then the characteristic of formula will be a well-formed $\mathcal{L}_{\mathbf{Kw}}$ formula. We prove below that the this characterization is deductively closed and complete.

Proposition 3.3.1.
Let φ be a modal formula and $\mathcal{M}_{\Phi}^{\mathcal{KC}} = \langle \mathcal{W}_{\Phi}^{\mathcal{KC}}, \{\mathcal{R}_a^{\mathcal{KC}} \mid a \in Ag\}, \mathcal{V}^{\mathcal{KC}}\rangle$ the canonical model in $\Phi = Cl_C(\varphi)$.

(1) *if $\varphi \in \Phi$ and $\Gamma \in \mathcal{W}_{\Phi}^{\mathcal{KC}}$, then $\varphi \in \Gamma$ iff $\vdash_{\mathcal{KC}} \underline{\Gamma} \to \varphi$;*

(2) $\vdash_{\mathcal{KC}} \bigvee_{\Delta\in\mathcal{W}_{\Phi}^{\mathcal{KC}}} \underline{\Delta}$.

(3) *if $\varphi \in \Phi$, then $\vdash_{\mathcal{KC}} \chi_{\varphi} \leftrightarrow \neg\chi_{\neg\varphi}$.*

Proof. **(1)** We assume that $\varphi \in \Phi$ and $\Gamma \in \mathcal{W}_{\Phi}^{\mathcal{KC}}$.

$\Rightarrow$: Assume $\varphi \in \Gamma$, then $\underline{\Gamma} = \ldots \wedge \varphi \wedge \ldots$. By propositional reasoning, it follows that $\vdash_{\mathcal{KC}} \underline{\Gamma} \to \varphi$.

$\Leftarrow$: Assume $\vdash_{\mathcal{KC}} \underline{\Gamma} \to \varphi$. By propositional reasoning, $\vdash_{\mathcal{KC}} \neg(\underline{\Gamma} \wedge \neg\varphi)$. Therefore, $\{\Gamma, \neg\varphi\}$ is not $\mathcal{KC}$-consistent. From the assumption that $\Gamma \in \mathcal{W}_{\Phi}^{\mathcal{KC}}$, then Γ is a $\mathcal{KC}$-MCS. Thus, $\neg\varphi \notin \Gamma$. This together with $\varphi \in \Phi$ entails that $\varphi \in \Gamma$.

(2) Assume towards a contradiction that $\nvdash_{\mathcal{KC}} \bigvee_{\Delta \in \mathcal{W}_{\Phi}^{\mathcal{KC}}} \underline{\Delta}$. By propositional reasoning, this is equivalent to $\nvdash_{\mathcal{KC}} \neg(\bigwedge_{\Delta \in \mathcal{W}_{\Phi}^{\mathcal{KC}}} \neg\underline{\Delta})$. Thus, $\Theta = \{\neg\underline{\Delta} \mid \Delta \in \mathcal{W}_{\Phi}^{\mathcal{KC}}\}$ is $\mathcal{KC}$-consistent. This is equivalent to:

$$\Theta = \{ \bigvee_{\varphi \in \Delta} \neg\varphi \mid \Delta \text{ is a } \mathcal{KC}\text{-consistent in } \Phi\} \text{ is } \mathcal{KC}\text{-consistent.}$$

Therefore, for all Δ that is a $\mathcal{KC}$-consistent in Φ there exists a $\varphi_{\Delta} \in \Delta$ such that the following set is $\mathcal{KC}$-consistent, as well:

$$\Theta' = \{\neg\varphi_{\Delta} \mid \Delta \text{ is a } \mathcal{KC}\text{-consistent in } \Phi\}.$$

Note that $\Theta' \subseteq \Phi$ and that we can prove a Lindenbaum's lemma for $\mathcal{KC}$ consistent sets in Φ similar to lemma 3.2.1. Given these, it follows that there exists a $\mathcal{KC}$-MCS in Φ that contains Θ'. However, this is a contradiction, because Θ', by its construction, is inconsistent with all $\mathcal{KC}$-MCS in Φ.

(3) Consequence of **(2)**. □

3.3.2 Soundness and Completeness of $\mathcal{KC}$ and $\mathcal{S}5\mathcal{C}$

In the results that follow, given a formula $\varphi \in \mathcal{L}_{\mathbf{KC}}$, we consider the structures and abbreviations below:

- $Sub(\varphi)$ as the set with all φ subformulas;
- $\Phi = Cl_C(\varphi)$ as its closure;
- $\mathcal{M}_{\Phi}^{\mathcal{KC}} = \langle \mathcal{W}_{\Phi}^{\mathcal{KC}}, \{\mathcal{R}_a^{\mathcal{KC}} \mid a \in Ag\}, \mathcal{V}^{\mathcal{KC}} \rangle$ as the canonical model of $\mathcal{KC}$ in Φ.

Below we prove that, although we restrict the domain of $\mathcal{KC}$-MCS to the closure of a formula, they still have the compositional properties we need to prove the truth lemma (lemma 3.3.5) later.

Lemma 3.3.1.
Let $\varphi \in \mathcal{L}_{\boldsymbol{KC}}$ and Γ a $\mathcal{KC}$-MCS in $Cl_C(\varphi)$.

(1) *if $\psi \in Cl_C(\varphi)$, then $\psi \in \Gamma$ iff $\neg\psi \notin \Gamma$.*

(2) *if* $\psi \in Cl_C(\varphi)$*,* $\{\psi_1, \ldots, \psi_n\} \subseteq \Gamma$ *and* $\vdash_{\mathcal{KC}} (\psi_1 \wedge \ldots \wedge \psi_n) \rightarrow \psi$ *then* $\psi \in \Gamma$*.*

(3) *if* $\psi \wedge \psi' \in Cl_C(\varphi)$*, then* $\psi \wedge \psi' \in \Gamma$ *iff* $\psi \in \Gamma$ *and* $\psi' \in \Gamma$*.*

Proof. Analogous to proof of proposition 3.2.1 and lemma 3.2.2. □

In addition we prove that $\mathcal{KC}$-MCS, restricted to the closure of a formula, are coherently connected w.r.t. modal operators $\mathbf{K}_a$ and $\mathbf{E}_G$. To prove the $\mathbf{E}_G$ case we use the general knowledge axiom **(AE)**.

Lemma 3.3.2 ($\mathcal{KC}$-MCS are coherently connected)**.**
Let $\varphi \in \mathcal{L}_{\boldsymbol{KC}}$ *be a formula and* Γ *be a* $\mathcal{KC}$*-MCS in* Φ*.*

(1) *if* $\boldsymbol{K}_a\psi \in Cl_C(\varphi)$*, then* $\boldsymbol{K}_a\psi \in \Gamma$ *iff for all* $\Delta \in \mathcal{W}_\Phi^{\mathcal{KC}}$ *if* $(\Gamma, \Delta) \in \mathcal{R}_a^{\mathcal{KC}}$ *then* $\psi \in \Delta$*;*

(2) *if* $\boldsymbol{E}_G\psi \in Cl_C(\varphi)$*, then* $\boldsymbol{E}_G\psi \in \Gamma$ *iff for all* $a \in G$*,* $\boldsymbol{K}_a\psi \in \Gamma$*.*

Proof. Consider an arbitrary $\varphi \in \mathcal{L}_{\mathbf{KC}}$ and its closure $\Phi = Cl_C(\varphi)$. Let $G \subseteq Ag$ be a group of agents. We assume, in all proofs that follow, that Γ is a $\mathcal{KC}$-MCS in Φ.

(1) Note that, if $\mathbf{K}_a\psi \in \Phi$ then $\{\mathbf{K}_a\psi, \psi, \neg\mathbf{K}_a\psi, \neg\psi\} \subseteq \Phi$. Given this, the proof is analogous to the proof of lemma 3.2.3.

(2) Assume that $\mathbf{E}_G\psi \in \Phi$. It is easy to prove that, for all $a \in G$, $\{\mathbf{K}_a\psi, \neg\mathbf{K}_a\psi\} \subseteq \Phi$.

$\Rightarrow$: Assume that $\mathbf{E}_G\psi \in \Gamma$. Assume towards a contradiction that there exists $a' \in G$ such that:

$$\mathbf{K}_{a'}\psi \notin \Gamma.$$

Then, by (1) and (2):

$$\neg \bigwedge_{a \in G} \mathbf{K}_a\psi \in \Gamma.$$

Using the axiom for general knowledge and propositional reasoning:

$$\vdash_{\mathcal{KC}} \neg(\mathbf{E}_G\psi \wedge \neg \bigwedge_{a \in G} \mathbf{K}_a\psi).$$

Therefore, the subset $\{\mathbf{K}_a\psi \mid a \in G \setminus \{a'\}\} \cup \{\neg\mathbf{K}_{a'}\psi\} \cup \{\mathbf{E}_G\psi\}$ of Γ is not $\mathcal{KC}$-consistent, which contradicts our initial assumption that Γ is a $\mathcal{KC}$-MCS.

$\Leftarrow$: Assume that $\mathbf{K}_a\psi \in \Gamma$, for all $a \in G$. Then, by lemma 3.3.1:

$$\bigwedge_{a \in G} \mathbf{K}_a\psi \in \Gamma.$$

If we consider, in addition, the axiom for general knowledge (AE) and lemma 3.3.1, it follows that $\mathbf{E}_G\psi \in \Gamma$. □

To prove that canonical model worlds are coherently connected w.r.t. the $\mathbf{C}_G$ is more complicated. Recall that common knowledge axiomatization is based on its view as the fixed-point of $f(\varphi) = \mathbf{E}_G(\varphi \wedge f(\varphi))$. Our strategy is to use the characteristic formula of $\mathbf{C}_G\varphi$ to prove this result.

Lemma 3.3.3 ($\chi_{\mathbf{C}_G\varphi}$)**.**
Let $\varphi \in \mathcal{L}_{\boldsymbol{KC}}$ be a formula, $\Phi = Cl_C(\varphi)$ be its closure and $\mathcal{M} = \mathcal{M}^{\mathcal{KC}}_{\Phi}$ be its canonical model for $\mathcal{KC}$ in Φ. If $\boldsymbol{C}_G\psi \in Cl_C(\varphi)$, then for all $\Delta \in [\boldsymbol{C}_G\psi]^{\mathcal{M}}$ and $\Theta \in [\neg\boldsymbol{C}_G\psi]^{\mathcal{M}}$:

- ***(1)*** *$\nvdash_{\mathcal{KC}} \underline{\Delta} \rightarrow \boldsymbol{K}_a\psi$ iff there exists $\Delta' \in \mathcal{W}^{\mathcal{KC}}_{\Phi}$ such that $\Delta\mathcal{R}_a\Delta'$ and $\psi \notin \Delta'$;*
- ***(2)*** *$\vdash_{\mathcal{KC}} \underline{\Delta} \rightarrow \boldsymbol{K}_a\neg\underline{\Theta}$.*

Proof. Consider an $\varphi \in \mathcal{L}_{\mathbf{KC}}$ and $\mathbf{C}_G\psi \in Cl_C(\varphi)$. Then, by definition of $Cl_C(\varphi)$:

$$\{\mathbf{C}_G\psi,\ \mathbf{E}_G(\psi \wedge \mathbf{C}_G\psi),\ \psi \wedge \mathbf{C}_G\psi\} \cup \{\mathbf{K}_a(\psi \wedge \mathbf{C}_G\psi),\ \mathbf{K}_a\psi \mid a \in G\} \subseteq Cl_C(\varphi). \qquad (*)$$

(1) Consider a $\Delta \in [\mathbf{C}_G\psi]^{\mathcal{M}}$, then Δ is a $\mathcal{KC}$-MCS in $Cl_C(\varphi)$.

$$\begin{aligned}
&\nvdash_{\mathcal{KC}} \underline{\Delta} \rightarrow \mathbf{K}_a\psi && \overset{\text{Propositional reasoning}}{\Leftrightarrow}\\
&\nvdash_{\mathcal{KC}} \neg(\underline{\Delta} \wedge \neg\mathbf{K}_a\psi) && \underset{(*)}{\overset{\text{def. 3.2.5}}{\Leftrightarrow}}\\
&\neg\mathbf{K}_a\psi \in \Delta && \overset{\text{lemma 3.3.1}}{\Leftrightarrow}\\
&\mathbf{K}_a\psi \notin \Delta && \underset{(*)}{\overset{\text{lemma 3.3.2}}{\Leftrightarrow}}\\
&\text{there exists } \Delta' \in \mathcal{W}^{\mathcal{KC}}_{\Phi},\ \Delta\mathcal{R}\Delta' \text{and } \psi \notin \Delta'.
\end{aligned}$$

(2) Consider $\Theta \in [\neg\mathbf{C}_G\psi]$. Given the equivalences below, Θ is not G-reachable from any $\Delta \in [\mathbf{C}_G\psi]$.

$$\begin{aligned}
[\neg\mathbf{C}_G\psi] \overset{def}{\equiv} &\{\Theta \mid (\mathcal{M}^{\mathcal{KC}}_{\Phi},\ \Theta) \vDash \neg\mathbf{C}_G\psi\} \overset{\text{lemma 2.4.3}}{\Leftrightarrow}\\
&\{\Theta \mid \text{there exists } \Theta' \ G\text{-reachable from } \Theta \text{ and } (\mathcal{M}^{\mathcal{KC}}_{\Phi},\ \Theta') \nvDash \psi\} \underset{\text{def. 3.3.4}}{\overset{\text{def. 2.4.3}}{\Leftrightarrow}}\\
&\{\Theta \mid \Theta \ \text{ is not } G\text{-reachable from } \Delta \in [\mathbf{C}_G\psi]\}
\end{aligned}$$

In particular, $(\Delta, \Theta) \notin \mathcal{R}^{\mathcal{KC}}_a$. By definition of canonical model (def. 3.2.7), this means that, there exists a formula ψ' such that $\mathbf{K}_a\psi' \in \Delta$ but $\psi' \notin \Theta$. Therefore, as $\mathbf{K}_a\psi' \in \Delta$ then $\psi' \in Cl_C(\varphi)$ and, by lemma 3.3.1, $\neg\psi' \in \Theta$. It follows then:

$$\begin{aligned}
&\vdash_{\mathcal{KC}} \underline{\Theta} \rightarrow \neg\psi' && \overset{\text{Prop. reasoning}}{\Leftrightarrow}\\
&\vdash_{\mathcal{KC}} \psi' \rightarrow \neg\underline{\Theta} && \overset{N\mathcal{K}\text{ lemma}}{\Leftrightarrow}\\
&\vdash_{\mathcal{KC}} \mathbf{K}_a\psi' \rightarrow \mathbf{K}_a\neg\underline{\Theta} && \underset{\text{Prop. reasoning}}{\Leftrightarrow}\\
&\vdash_{\mathcal{KC}} \underline{\Delta} \rightarrow \mathbf{K}_a\neg\underline{\Theta}
\end{aligned}$$

□

Lemma 3.3.4 ($\mathcal{KC}$-MCS and common knowledge)**.**
Let $\varphi \in \mathcal{L}_{\boldsymbol{KC}}$ be a formula, $\mathcal{M}_\Phi^{\mathcal{KC}} = \langle \mathcal{W}_\Phi^{\mathcal{KC}}, \{\mathcal{R}_a^{\mathcal{KC}} \mid a \in Ag\}, \mathcal{V}^{\mathcal{KC}}\rangle$ be the canonical model of φ in Φ, and $\Gamma \in \mathcal{W}_\Phi^{\mathcal{KC}}$ be a $\mathcal{KC}$-MCS in Φ. If $\boldsymbol{C}_G\psi \in Sub(\varphi)$, then:

$$\boldsymbol{C}_G\psi \in \Gamma \text{ iff for all } \Delta \in \mathcal{W}_\Phi^{\mathcal{KC}}, \text{ if } \Delta \text{ is } G\text{-reachable in } k \geq 1 \text{ steps in } \mathcal{M}_\Phi^{\mathcal{KC}} \text{ from } \Gamma, \text{ then } \{\psi, \boldsymbol{C}_G\psi\} \subseteq \Delta.$$

Proof. Consider an arbitrary $\varphi \in \mathcal{L}_{\mathbf{KC}}$ and its closure $\Phi = Cl_C(\varphi)$. Let $G \subseteq Ag$ be a group of agents. We assume, in all proofs that follow, that Γ is a $\mathcal{KC}$-MCS in Φ and that $\mathbf{C}_G\psi \in Sub(\varphi)$. Then, by definition of $Cl_C(\varphi)$ (def. 3.3.2):

$$\{\mathbf{C}_G\psi,\ \mathbf{E}_G(\psi \wedge \mathbf{C}_G\psi),\ \psi \wedge \mathbf{C}_G\psi\} \cup \{\mathbf{K}_a(\psi \wedge \mathbf{C}_G\psi),\ \mathbf{K}_a\psi \mid a \in G\} \subseteq \Phi. \qquad (*)$$

$\Rightarrow$: Assume that (**r1**) $\mathbf{C}_G\psi \in \Gamma$. We will prove by induction in k that, for all $\Delta \in \mathcal{W}_\Phi^{\mathcal{KC}}$:

$$\text{if } \Delta \text{ is } G\text{-reachable in } k \geq 1 \text{ steps in } \mathcal{M}_\Phi^{\mathcal{KC}} \text{ from } \Gamma, \text{ then } \{\psi, \mathbf{C}_G\psi\} \subseteq \Delta. \qquad (\dagger)$$

We use the fixed-point axiom (*FP*) to prove the base case, $k = 1$. Given (*), (**r1**) and $\vdash_{\mathcal{KC}} \mathbf{C}_G\psi \to \mathbf{E}_G(\psi \wedge \mathbf{C}_G\psi)$ (*FP*), then, by lemma 3.3.1, $\mathbf{E}_G(\psi \wedge \mathbf{C}_G\psi) \in \Gamma$. We can use the results proved in the previous lemma, to show that:

$$\text{for all } a \in G, \text{ for all } \Delta \in \mathcal{W}_\Phi^{\mathcal{KC}}, \text{ if } (\Gamma, \Delta) \in \mathcal{R}_a^{\mathcal{KC}}, \text{ then } \{\psi, \mathbf{C}_G\psi\} \subseteq \Delta.$$

This is equivalent to (†) for $k = 1$, by the definition of G-reachability (def. 2.4.3) and generalized union.

We assume as induction hypothesis that (†) holds for k. Additionally, we consider an arbitrary $\Delta \in \mathcal{W}_\Phi^{\mathcal{KC}}$ and assume that:

$$\Delta \text{ is } G\text{-reachable in } k+1 \text{ steps in } \mathcal{M}_\Phi^{\mathcal{KC}} \text{ from } \Gamma.$$

By the definition of G-reachability, the previous is equivalent to:

$$\text{there exists a path } (\Gamma, S_1, \ldots, S_k, \Delta) \text{ in } \mathcal{M}_\Phi^{\mathcal{KC}}. \qquad (\mathbf{r2})$$

Therefore, we can apply the induction hypothesis to the state S_k, as it is G-reachable from Γ in k steps, and conclude that $\{\psi, \mathbf{C}_G\psi\} \subseteq S_k$. As $\mathbf{C}_G\psi \in S_k$, we can apply the same reasoning as in the base case, and conclude that, for all $\Delta' \in \mathcal{W}_\Phi^{\mathcal{KC}}$:

$$\text{if } \Delta' \text{ is } G\text{-reachable in 1 step in } \mathcal{M}_\Phi^{\mathcal{KC}} \text{ from } S_k, \text{ then } \{\psi, \mathbf{C}_G\psi\} \subseteq \Delta'.$$

Δ is G-reachable from S_k in one step, due to **(r2)**. Thus, $\{\psi, \mathbf{C}_G\psi\} \subseteq \Delta$.

$\Leftarrow$: Assume that, for all $\Delta \in \mathcal{W}_\Phi^{\mathcal{KC}}$:

$$\text{if } \Delta \text{ is } G\text{-reachable in } k \geq 1 \text{ steps in } \mathcal{M}_\Phi^{\mathcal{KC}} \text{ from } \Gamma, \text{ then } \{\psi, \mathbf{C}_G\psi\} \subseteq \Delta. \qquad (\mathbf{l1})$$

We want to prove that $\mathbf{C}_G\psi \in \Gamma$.

We use the characteristic formula of $\mathbf{C}_G\varphi$ (def. 3.3.4) to prove the following:

$$\vdash_{\mathcal{KC}} \chi_{\mathbf{C}_G\psi} \to \mathbf{E}_G(\psi \wedge \chi_{\mathbf{C}_G\psi}).$$

Given the definition of $\chi_{\mathbf{C}_G\psi}$ (def. 3.3.4) and the axiom for general knowledge in $\mathcal{KC}$ (AE), then the previous is equivalent to:

$$\text{for all } \Delta \in \chi_{\mathbf{C}_G\psi} \text{ and } a \in G,\ \vdash_{\mathcal{KC}} \underline{\Delta} \to \mathbf{K}_a(\psi \wedge \chi_{\mathbf{C}_G\psi}).$$

We prove earlier, in example 3.2, that $\mathbf{K}_a$ distributes over conjunction. Thus, the previous is equivalent to:

$$\text{for all } \Delta \in \chi_{\mathbf{C}_G\psi} \text{ and } a \in G,\ \vdash_{\mathcal{KC}} \underline{\Delta} \to \mathbf{K}_a\psi \wedge \mathbf{K}_a\chi_{\mathbf{C}_G\psi}.$$

By propositional reasoning, this means that we need to prove the following two statements, for all $\Delta \in \chi_{\mathbf{C}_G\psi}$ and $a \in G$:

$$\textbf{(i)} \vdash_{\mathcal{KC}} \underline{\Delta} \to \mathbf{K}_a\psi \quad \text{and} \quad \textbf{(ii)} \vdash_{\mathcal{KC}} \underline{\Delta} \to \mathbf{K}_a\chi_{\mathbf{C}_G\psi}.$$

In what follows we consider an arbitrary $\Delta \in \chi_{\mathbf{C}_G\psi}$ and $a \in G$.

We assume, towards a contradiction that **(i)** does not hold. By lemma 3.3.3, there exists $\Delta' \in \mathcal{W}_\Phi^{\mathcal{KC}}$ such that $\Delta\mathcal{R}_a\Delta'$ and $\psi \notin \Delta'$. Given our assumption **l1** and the fact that Δ' is G-reachable from Δ, then $\{\psi, \mathbf{C}_G\psi\} \subseteq \Delta'$. This is a contradiction, because before we said that $\psi \notin \Delta'$. Therefore, $\vdash_{\mathcal{KC}} \underline{\Delta} \to \mathbf{K}_a\psi$ holds.

By 3.3.1, we can prove $\vdash_{\mathcal{KC}} \underline{\Delta} \to \mathbf{K}_a\chi_{\mathbf{C}_G\psi}$ by proving instead $\vdash_{\mathcal{KC}} \underline{\Delta} \to \mathbf{K}_a\neg\chi_{\neg\mathbf{C}_G\psi}$. This is equivalent to the statement proved in 3.3.3:

$$\text{for all } \Theta \in \chi_{\neg\mathbf{C}_G\psi} \text{ and } a \in G,\ \vdash_{\mathcal{KC}} \underline{\Delta} \to \mathbf{K}_a\neg\Theta.$$

Thus, $\vdash_{\mathcal{KC}} \chi_{\mathbf{C}_G\psi} \to \mathbf{E}_G(\psi \wedge \chi_{\mathbf{C}_G\psi})$. and, by application of the **IC** rule, it follows $\vdash_{\mathcal{KC}} \chi_{\mathbf{C}_G\psi} \to \mathbf{C}_G\psi$.

We want to prove that $\Gamma \in [\mathbf{C}_G\psi]$, i.e. that $(\mathcal{M}_\Phi^{\mathcal{KC}}, \Gamma) \vDash \mathbf{C}_G\psi$. Assume towards a contradiction that is not the case, then there exists a G-path $(\Gamma, s_1, \dots, s_{n-1}, \Delta)$, with $n \geq 1$, such that $(\mathcal{M}_\Phi^{\mathcal{KC}}, \Delta) \vDash \neg\psi$. In particular, we have that $(\mathcal{M}_\Phi^{\mathcal{KC}}, s_{n-1}) \vDash \neg\mathbf{K}_a\psi$, for some $a \in G$. By 3.3.2, this means that, for some $a \in G$, there exists a Δ' such that $s_{n-1}\mathcal{R}_a\Delta'$ and $\psi \notin \Delta'$. Note that such Δ' is G-reachable from Γ and this contradicts our initial assumption (**l1**).

By $\Gamma \in [\mathbf{C}_G\psi]$ and (*****), then $\vdash_{\mathcal{KC}} \underline{\Gamma} \to \chi_{\mathbf{C}_G\psi}$. Together with $\vdash_{\mathcal{KC}} \chi_{\mathbf{C}_G\psi} \to \mathbf{C}_G\psi$ it follows that $\vdash_{\mathcal{KC}} \underline{\Gamma} \to \mathbf{C}_G\psi$. By proposition 3.3.1, then $\mathbf{C}_G\psi \in \Gamma$. □

We can use all the previous lemmas to prove the following truth lemma.

Lemma 3.3.5 (Truth Lemma)**.**
Let $\psi \in \mathcal{L}_{\boldsymbol{KC}}$*,* $\Phi = Cl_C(\psi)$ *the closure of* ψ *and* $\mathcal{M}^{\mathcal{KC}}_{\Phi}$ *be the canonical model of* $\mathcal{KC}$ *in* Φ*. For all* Γ *that are* $\mathcal{KC}$*-MCS in* Φ *and* $\varphi \in Sub(\psi)$*,* $(\mathcal{M}^{\mathcal{KC}}, \Gamma) \vDash \varphi$ *iff* $\varphi \in \Gamma$*.*

Proof. We prove it by induction on $\varphi \in \mathcal{L}_{\mathbf{KC}}$. The base case follows from definition of canonical model (def. 3.2.7). The inductive cases for the boolean operators and the knowing that operator ($\mathbf{K}_a$), can be proved as in the previous section using lemma 3.3.1 and lemma 3.3.2, respectively. For the general knowledge and common knowledge we use lemma 3.3.2 and lemma 3.3.4, respectively. □

Theorem 3.3.6 ($\mathcal{KC}$ soundness and completeness w.r.t. $\mathbb{K}$)**.**
$\mathcal{KC}$ is sound and complete with respect to the class of all Kripke models $\mathbb{K}$*.*

Proof. The soundness follows trivially from the proofs in the example 2.10.

We can use the canonical model, as we did in theorem 3.2.5, to prove that every $\mathcal{KC}$-consistent formula is satisfiable in some Kripke model. By proposition 3.2.2, it follows that $\mathcal{KC}$ is complete with respect to the class of all Kripke models $\mathbb{K}$. We can prove that, by proving a Lindenbaum's lemma for $\mathcal{KC}$, similar to lemma 3.2.1. Then, by the truth lemma 3.3.5, every consistent formula is satisfiable in the canonical model at the state representing a $\mathcal{K}$-MCS it belongs to. □

For completeness result for $\mathcal{S}5\mathcal{C}$ we need to modify the accessibility relation in the canonical model to be an equivalence relation. This is nicely proved in [vDvdHK07].

Theorem 3.3.7 ($\mathcal{S}5\mathcal{C}$ soundness and completeness w.r.t. $\mathbb{S}5$ [vDvdHK07])**.**
$\mathcal{S}5\mathcal{C}$ is sound and complete with respect to the class of all Kripke models $\mathbb{S}5$*.*

3.4 Paradox: Agents' Knowledge and Moore's Sentences

Moore's sentences are of the form *'It is raining, but I don't believe it is raining.'*. If we assume that someone asserting a statement implies that he believes in that same statement, then asserting a Moore's sentence is paradoxical. In [Hin62], Hintikka provides an analysis of such sentences involving both the concepts of belief and knowledge. He uses his system to prove that it is inconsistent for an agent to believe in a Moore's sentence if it involves their own beliefs.

We consider an analogous to Moore's sentences involving only knowledge. We present below a $\mathcal{T}$-proof, based on [Hin62], for the fact that Moore's sentences cannot be known. Therefore, we prove that $\vdash_{\mathcal{T}} \neg\mathbf{K}_a(C \wedge \neg\mathbf{K}_a C)$. In order to improve the proof's readability we refer to lemmas presented in example 3.2.

$\mathcal{T}$-proof:

1.	$\mathbf{K}_a(\neg\mathbf{K}_a C) \rightarrow \neg\mathbf{K}_a C$	*Axiom T*
2.	$\mathbf{K}_a(C \wedge \neg\mathbf{K}_a C) \rightarrow (\mathbf{K}_a C \wedge \mathbf{K}_a \neg\mathbf{K}_a C)$	*Dist*
3.	$\mathbf{K}_a(C \wedge \neg\mathbf{K}_a C) \rightarrow (\mathbf{K}_a C \wedge \neg\mathbf{K}_a C)$	$PL_5(1,2)$
4.	$\neg(\mathbf{K}_a C \wedge \neg\mathbf{K}_a C) \rightarrow \neg\mathbf{K}_a(C \wedge \neg\mathbf{K}_a C)$	$PL_1(3)$
5.	$\neg(\mathbf{K}_a C \wedge \neg\mathbf{K}_a C)$	*Taut*
6.	$\neg\mathbf{K}_a(C \wedge \neg\mathbf{K}_a C)$	$MP(4,5)$

In [HII10] the authors establish a connection between Moore's sentences an unsuccessful announcements. An announcement is unsuccessful, if the sentence announced becomes false after its announcement.

CHAPTER 4

Knowing Whether and Ignorance

In this chapter we present a modal logic that captures epistemic notions of *ignorance* and its dual *to know whether*. In some epistemic puzzles, like the muddy children presented in chapter 2, ignorance is repeatably announced by the agents until some of them knows whether a given proposition holds. This motivated us to have a close look at these epistemic notions and investigate how are they related with the notion of *knowing that*.

We start the chapter by discussing our definition of ignorance. Afterwards, in section 4.2, we present the state of the art on formal characterization of knowing whether and ignorance using modal logic. In the next section we introduce a *logic for knowing whether*, as in [FWvD15]. The main goal of that section is to develop intuitions about this non-normal logic. Therefore, we complement the technical introduction with examples. Finally, in the last section we use this logic to prove that, given two agents and a fixed proposition, there are uncountably many distinct states of knowledge. This result was proved before in [HHS96] using information structures. Here we prove it using Kripke structures.

4.1 What is Ignorance?

The definition of ignorance adopted in this work is motivated by epistemic scenarios in which an agent *does not know the answer to a 'know whether question'*. Therefore, an agent being *ignorant* about some statement, $\mathbf{I}_a\varphi$, is the opposite of that agent *knowing whether* that statement holds, $\mathbf{Kw}_a\varphi$.

Definition 4.1.1.
We define below *knowing whether* in terms of *knowing that* and *ignorance* in terms of

knowing whether, as follows:

$$\mathbf{Kw}_a\varphi \stackrel{def}{\equiv} \mathbf{K}_a\varphi \vee \mathbf{K}_a\neg\varphi$$
$$\mathbf{I}_a\varphi \stackrel{def}{\equiv} \neg\mathbf{Kw}_a\varphi.$$

Example 4.1 (*Ignorance vs. Not knowing that*).
Cat does not know whether Bert is currently visiting Lisbon. She was not paying attention when Bert told her about his trip and she does not want him to know that. Thus, instead of asking him directly about his location, she decides to ask about the weather conditions in Lisbon. She knows that the only reason for him to not be ignorant about the weather there is if he is there.

If Cat decides to ask Bert if he *knows that* it is raining in Lisbon, then he can be there and still answer truthfully that he does not know that. As he may know that it is sunny, instead. Therefore, this question fails to test his ignorance on this matter. Cat should ask *whether he knows* if it is raining there. In this case, Bert's answer will allow Cat to infer whether he is currently in Lisbon.

It is not the concern of this thesis to elaborate on either linguistics or philosophical justifications for the definitions just presented. It is, nevertheless, relevant to have an overview of the current discussions in epistemology on this matter.

Ignorance is defined by Oxford dictionary as *'Lack of knowledge or information.'* This view is widely accepted by epistemologists [LM13] and it is referred as the *'standard view'*. In [Kyl15] the duality between ignorance and knowledge, with respect to this view, is summarized as follows:

> On the standard view, knowledge and ignorance are mutually exclusive and jointly exhaustive; they are contraries and contradictories. (...) The theory can be stated as necessary and sufficient conditions for ignorance:
>
> Standard View: For any truth P, S is ignorant of the fact that P if and only if S does not know that P.

Recently Peels challenged the standard view by defining ignorance as the *absence of true belief* [Pee10][1]. While arguing in favor of the standard view Le Morvan identifies two notions of ignorance [LM13]: propositional and factative. Later, in [LM15], he defends that ignorance is a failure of *knowing of*, instead of a failure of *knowing that*. The following quote clarifies his view:

[1] In epistemic logic, knowledge is defined as true belief. This is not the case in epistemology. Gettier-cases are counterexamples for this definition (see [Ste17]).

> Accordingly, on the intuitive idea that knowledge and ignorance are complements, the complement of ignorance of a proposition p is not knowledge **that** p but rather knowledge **of** p—an acquaintance with or knowledge of an entity, where the entity in question is a proposition. Such acquaintance or knowledge may be occurrent (as when one is conscious of it) or dispositional (as when one retains it in memory). It requires the deployment of concepts in the grasping or comprehension (whether occurrent or dispositional) of a proposition. Knowledge of p is required to have—and is therefore entailed by, and a precondition of—any propositional attitude in relation to p such as believing that p, considering that p, doubting that p, hoping that p, or knowing that p.

It is not clear to us if this can still be considered the standard view, because the standard view definition presented in [Kyl15] defines ignorance as the failure of knowing that.

The stance we adopt in this work is classified in [Fan16] as the *'logical view'*, in which ignorance is defined as the contingent counterpart of knowing that. A proposition is *contigent* when it is neither necessarily true nor necessarily false, likewise a proposition is *non-contingent* when it is either necessarily true or necessarily false. Given that we interpret knowing that as a necessity operator, then it is easy to understand how the concepts *knowing whether* and *ignorance* are related with *non-contingency* and *contingency*, respectively.

4.2 State of the Art

Ignorance is barely mentioned in the current literature on epistemic logic, and only recently the relation between ignorance and contingency was established [FWvD15, vDF16, CL16].

Contingency modal logic, first introduced in [MR66], investigates the logic with the sole primitive operator Δ with the intended reading of *'it is non-contingent whether...'*. Its dual is denoted by ∇. In [H$^+$95] an infinite axiomatization for the class of all frames and for the class of serial frames is presented. Later, a finite axiomatization is proposed in [Kuh95], which includes a system for transitive frames. In [Zol99], the authors introduce a finite axiomatization that looks alike the system for minimal normal modal logic as in definition 3.2.1. We present the finite axiomatizations on table 4.1.

A propositional modal logic for ignorance is introduced in [vdHL04]. Its defined over a language that considers only one primitive modal operator $\mathbf{I}_a$, for each agent a, read as *'the agent a is ignorant about ...'*. Ignorance is interpreted as the contingent counterpart of knowledge. However, the authors seem to not be aware of the work done previously in contingent modal logic. They prove that their deductive system (in table 4.2) is sound and complete with respect to the class of all Kripke models, by defining an appropriate canonical model. Additionally, they a sound and complete axiomatization is shown for the class of transitive frames, by extending the system for ignorance with a new axiom scheme. However, in [FWvD15] they present a counter example for the validity of this

	[Kuh95]	[Zol99]
Axioms:	All instances of propositional tautologies	
	$\Delta\neg A \to \Delta A$	$\Delta A \leftrightarrow \Delta\neg A$
	$\Delta A \wedge \nabla(A \wedge B) \to \nabla B$	$\Delta(A \leftrightarrow B) \to (\Delta A \leftrightarrow \Delta B)$
	$\Delta A \wedge \nabla(A \vee B) \to \Delta(\neg A \wedge C)$	$\Delta A \to (\Delta(A \to B) \vee \Delta(A \to C))$
Rules:	$\dfrac{\vdash A}{\vdash \Delta A}\ Nec\Delta$	$\dfrac{\vdash A \leftrightarrow B}{\vdash \Delta A \leftrightarrow \Delta B}$ $\qquad \dfrac{A \quad A \to B}{B}\ MP$

Table 4.1: Minimal contingent modal logic systems by [Kuh95] and by [Zol99].

Axioms:	All instances of propositional tautologies
	$\mathbf{I}\varphi \leftrightarrow \mathbf{I}\neg\varphi$
	$\mathbf{I}(\varphi \wedge \psi) \to (\mathbf{I}\varphi \vee \mathbf{I}\psi)$
	$(\neg\mathbf{I}\varphi \wedge \mathbf{I}(\alpha_1 \wedge \varphi)) \wedge \neg\mathbf{I}(\varphi \to \psi) \wedge \mathbf{I}(\alpha_2 \wedge (\varphi \to \psi)) \to (\neg\mathbf{I}\psi \wedge \mathbf{I}(\alpha_1 \wedge \psi))$
	$(\neg\mathbf{I}\psi \wedge \mathbf{I}\alpha) \to (\mathbf{I}(\alpha \wedge \psi) \vee \mathbf{I}(\alpha \wedge \neg\psi))$
Rules:	Substitution of equivalents
	$\dfrac{\vdash \varphi}{\vdash \neg\mathbf{I}\varphi \wedge (\mathbf{I}\alpha \to \mathbf{I}(\alpha \wedge \varphi))}\ R\mathbf{I} \qquad \dfrac{\varphi \quad \varphi \to \psi}{\psi}\ MP$

Table 4.2: Axiomatization of ignorance by [vdHL04].

axiom scheme over the class of transitive frames. The main problem with this approach is that it mostly an *ad hoc* solution with axioms that are difficult to interpret, specially in epistemic terms.

In [vDF16] they present a contingency logic with a non-contingent operator as the only primitive operator. They axiomatize contingency logic over different classes of frames based on an *almost definable* schema for necessity, define in a previous work [FWvD14]. In addition they introduce dynamic operators in the context of contingency logics by extending the basic contingency logic with public announcements and later introducing action models in this context. A contingency logic with arbitrary announcements (**ACLA**) is proposed in [vDF16].

4.3 A Logic for Knowing Whether and Ignorance

4.3.1 Semantics

We define a propositional modal language with one non-contingent modal operator $\mathbf{Kw}_a$, for each $a \in Ag$.

Definition 4.3.1 (Knowing whether language $\mathcal{L}_{\mathbf{Kw}}$)**.**
Let At be a countable set of atomic propositions and Ag a finite set of agent symbols. The language for knowing whether, $\mathcal{L}_{\mathbf{Kw}}(At, Ag)$, is defined by the following BNF grammar, where $p \in At$ and $a \in Ag$:

$$\varphi \ ::= \ p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \mathbf{Kw}_a\varphi.$$

We read $\mathbf{Kw}_a\varphi$ as *'agent a knows whether φ'*. *Ignorance* is defined as the negation of knowing whether, i.e. $\mathbf{I}_a\varphi \overset{def}{\equiv} \neg\mathbf{Kw}_a\varphi$.

Definition 4.3.2 (Truth interpretation of $\mathcal{L}_{\mathbf{Kw}}$ formulas in Kripke semantics)**.**
Let $\mathcal{M} = \langle \mathcal{W}, \{\mathcal{R}_a \mid a \in Ag\}, \mathcal{V}\rangle$ be a Kripke model and $w \in \mathcal{W}$ a world. The truth interpretation for formula $\varphi \in \mathcal{L}_{\mathbf{Kw}}$ in the pointed model $(\mathcal{M}, w)$ is the same as in def. 2.2.4 for boolean connectives and propositional variables. In addition, *knowing whether* is interpreted as follows:

$$(\mathcal{M}, w) \models \mathbf{Kw}_a\varphi \text{ iff for all } v, v' \in \mathcal{W} \text{ such that } w\mathcal{R}_a v \text{ and } w\mathcal{R}_a v' \text{ then } (\mathcal{M}, v) \models \varphi \text{ iff } (\mathcal{M}, v') \models \varphi.$$

As a consequence, $\mathbf{I}_a\varphi$ is interpreted as follows:

$$(\mathcal{M}, w) \models \mathbf{I}_a\varphi \text{ iff there exists } v, v' \in \mathcal{W} \text{ such that } w\mathcal{R}_a v \text{ and } w\mathcal{R}_a v' \text{ and } (\mathcal{M}, v) \models \varphi \text{ and } (\mathcal{M}, v') \models \neg\varphi.$$

Example 4.2 (*Non-distributivity of* $\boldsymbol{Kw}_a$)*.*
We use the model $\mathcal{M}_{dist}$ defined below to prove that $\mathbf{Kw}_a$ does not distribute over conjunction, i.e.

$$(\mathcal{M}_{dist}, w_1) \not\models \mathbf{Kw}_a(p \wedge q) \to (\mathbf{Kw}_a p \wedge \mathbf{Kw}_a q).$$

For $i \in \{1, 2\}$, $(\mathcal{M}_{dist}, w_i) \models \neg(p \wedge q)$, then $(\mathcal{M}_{dist}, w_i) \models \mathbf{Kw}_a(p \wedge q)$. However, $(\mathcal{M}_{dist}, w_1) \models \neg\mathbf{Kw}_a p$ and $(\mathcal{M}_{dist}, w_1) \models \neg\mathbf{Kw}_a q$, because they have different valuations at w_1 and w_2. Thus, $(\mathcal{M}_{dist}, w_1) \not\models \mathbf{Kw}_a p \wedge \mathbf{Kw}_a q$.

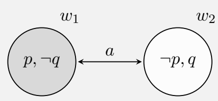

Figure 4.1: Counter example for distributivity of $\mathbf{Kw}_a$ over conjunction, $\mathcal{M}_{dist}$.

This model proves that $\mathbf{Kw}_a$ does not distribute over disjuction as well, i.e.

$$\not\models \mathbf{Kw}_a(p \vee q) \to (\mathbf{Kw}_a p \vee \mathbf{Kw}_a q).$$

We start by proving that the *logic of knowing whether over the class* $\mathbb{K}$, $\mathbb{K}_{\mathcal{L}_{\mathbf{Kw}}}$, is a modal logic (def. 3.2.1). From its interpretation of boolean operators, it is easy to see that it contains all propositional tautologies and it is closed under modus ponens. Below we prove that it is closed under uniform substitution, as well. We denote by $\varphi[\psi/p]$ a *uniform substitution* of φ, i.e. we replace all occurrences of p in φ by ψ. We can prove that given a counter-example $\mathcal{M}$ for the validity of $\varphi[\psi/p]$, then we can change $\mathcal{M}$ such that p is true at all worlds of $\mathcal{M}$ that satisfy ψ. The new model is a counter-example for the validity of φ.

Proposition 4.3.1 ($\mathbb{K}_{\mathcal{L}_{\mathbf{Kw}}}$ is closed under uniform substitution [FWvD15])**.**
Let $\varphi, \psi \in \mathcal{L}_{\mathbf{Kw}}$ *and* $p \in At$. *If* $\models \varphi$ *then* $\models \varphi[\psi/p]$.

Proof. We prove its contrapositive form, i.e. that for all $\varphi, \psi \in \mathcal{L}_{\mathbf{Kw}}$ and $p \in At$, if $\not\models \varphi[\psi/p]$ then $\not\models \varphi$. By definitions 2.2.5 and 2.2.4, we need to show that, $(*)$ if there exists a pointed Kripke model $(\mathcal{M}, w)$ such that $(\mathcal{M}, w) \models \neg\varphi[\psi/p]$ then $(\mathcal{M}', w') \models \neg\varphi$.

Assume arbitrary $\psi \in \mathcal{L}_{\mathbf{Kw}}$, $p \in At$, $\mathcal{M} = \langle \mathcal{W}, \{\mathcal{R}_a \mid a \in Ag\}, \mathcal{V}\rangle$ and $w \in \mathcal{W}$. Consider a Kripke model $\mathcal{M}'$ that is the same as $\mathcal{M}$ except for the valuation function $\mathcal{V}'$ that assigns to the variable p all worlds that satisfy ψ, i.e. $\mathcal{V}'(p) = \{v \mid v \in \mathcal{W} \text{ and } (\mathcal{M}, v) \models \psi\}$. We can prove, by induction in $\mathcal{L}_{\mathbf{Kw}}$ formulas, that $(\mathcal{M}, w) \models \varphi$ iff $(\mathcal{M}', w) \models \varphi[\psi/p]$ and use this result to prove $(*)$. □

The logic $\mathbb{K}_{\mathcal{L}_{\mathbf{Kw}}}$ is not a normal modal logic (def. 3.2.1), because the formula $\mathbf{Kw}_a(p \to q) \to (\mathbf{Kw}_a p \to \mathbf{Kw}_a q)$ is not valid. The Kripke model $\mathcal{M}_{normal}$ (fig. 4.2) is a counter example for the validity of this formula. It is easy to check that, for $i \in \{1, 2\}$, $(\mathcal{M}_{normal}, w_i) \models p \to q$ and $(\mathcal{M}_{normal}, w_i) \models \neg p$. Thus, $(\mathcal{M}_{normal}, w_1) \models \mathbf{Kw}_a(p \to q)$ and $(\mathcal{M}_{normal}, w_1) \models \mathbf{Kw}_a p$. However, q has a different valuation in world w_1 and w_2, so $(\mathcal{M}_{normal}, w_1) \not\models \mathbf{Kw}_a q$.

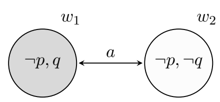

Figure 4.2: Counter example for $\mathbb{K}_{\mathcal{L}_{\mathbf{Kw}}}$ being a normal modal logic, $\mathcal{M}_{normal}$.

Example 4.3 (*How to know whether* φ?)*.*
The axiom $\mathcal{K}$ $(\mathbf{K}_a(\varphi \to \psi) \to (\mathbf{K}_a\varphi \to \mathbf{K}_a\psi))$, together with the application of modus ponens, allow us to derive formulas with $\mathbf{K}_a$ as the topmost operator. We can think of this axiom as describing a strategy for an agent to get to know that

a given φ holds. For example, if Cat knows that *'if Bert is on holidays then he is Lisbon'*, then she only needs to know that he is on holidays to know that he is in Lisbon.

The valid formula **(*)** $\mathbf{Kw}_a(\varphi \to \psi) \land \mathbf{Kw}_a(\neg\varphi \to \psi) \to \mathbf{Kw}_a\psi$, has a similar role in the logic of knowing whether. Consider the example below, from [Fan16]:

> As another example, suppose Jay submitted a paper to a conference a couple of months ago and he learns that his friend b also submitted a paper. Now Jay wonders whether his submission is accepted. But it is embarrassing to ask directly the programme chair who has decisions about submitted papers. What can he do? Jay can ask the chair the following two questions: 'Is b's submission or mine accepted?' 'If b's submission is accepted, then is mine accepted too?' No matter whether the chair says 'Yes' or 'No', Jay will know whether his submission is accepted and then his ignorance disappears.

Let j stand for *'Jay's submission is accepted'* and b for *'b's submission is accepted'*. We consider the instance of **(*)**: $\mathbf{Kw}_{Jay}(b \to j) \land \mathbf{Kw}_{Jay}(\neg b \to j) \to \mathbf{Kw}_{Jay}j$. After the programme chair's answers Jay knows whether b's submission or his is accepted ($\mathbf{Kw}_{Jay}(b \lor j)$, which is equivalent to $\mathbf{Kw}_{\text{Jay}}(\neg b \to j)$); and he knows whether, if b's submission is accepted then his is accepted ($\mathbf{Kw}_{\text{Jay}}(b \to j)$). Thus, by the validity of **(*)**, it follows that Jay knows whether his submission is accepted ($\mathbf{Kw}_{Jay}j$).

The fact that $\mathbf{Kw}_a\varphi$ holds at world w captures only the agent's epistemic state with respect to φ at that world. It does not force the interpretation of φ, at the accessible worlds from w, to a specific value. The counter-examples presented so far are based on this property. We illustrate it with the following valid formula:

$$\models \mathbf{Kw}_a\varphi \leftrightarrow \mathbf{Kw}_a\neg\varphi. \tag{4.1}$$

Example 4.4 (*Indistinguishable models*).
It is interesting to note that we cannot use formulas with knowing whether (or ignorance) as topmost operators to distinguish worlds that have at most one accessible world. Consider the Kripke models below:

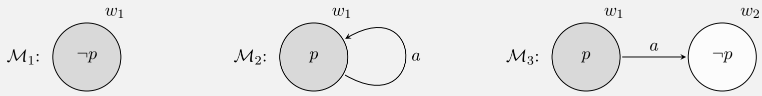

Figure 4.3: Kripke models with a trivial interpretation for all formulas in $\mathcal{L}_{\mathbf{Kw}}$.

For all formulas $\varphi \in \mathcal{L}_{\mathbf{Kw}}$ and $i \in \{1, 2, 3\}$: $(\mathcal{M}_i, w_1) \vDash \mathbf{Kw}_a\varphi$ and $(\mathcal{M}_i, w_1) \vDash \neg\mathbf{I}_a\varphi$.

The example above show that there are models which cannot be distinguished by $\mathcal{L}_{\mathbf{Kw}}$ formulas, even though they are clearly distinguishable using $\mathcal{L}_{\mathbf{K}}$ formulas. A formula with knowing whether as topmost operator vacuously holds at w, if w has at most one accessible world. This idea is explored below to prove that, for some class of models, $\mathcal{L}_{\mathbf{Kw}}$ is less expressive than $\mathcal{L}_{\mathbf{K}}$.

Before we prove the result we review the concept of two formulas being equivalent in a class of frames, and relative expressive power of two logic language over the same class of frames.

Definition 4.3.3 (Equivalence over a class of frames)**.**
Let $\mathbb{F}$ be a class of frames and φ, φ' be modal formulas. We say that φ *is equivalent to* φ' *over* $\mathbb{F}$, denoted by $\varphi \equiv_{\mathbb{F}} \varphi'$, iff for all $\mathcal{F} \in \mathbb{F}$ and for all pointed models $(\mathcal{M}, w)$ generated by $\mathcal{F}$: $(\mathcal{M}, w) \vDash \varphi$ iff $(\mathcal{M}, w) \vDash \varphi'$.

Definition 4.3.4 (Relative expressive power [vDvdHK07])**.**
Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be two modal languages that are interpreted in the same class of frames $\mathbb{F}$.

- $\mathcal{L}_2$ *is at least as expressive as* $\mathcal{L}_1$, denoted by $\mathcal{L}_1 \preceq_{\mathbb{F}} \mathcal{L}_2$, iff for all $\varphi \in \mathcal{L}_1$ there is a formula $\varphi_2 \in \mathcal{L}_2$ such that $\varphi_1 \equiv_{\mathbb{F}} \varphi_2$;
- $\mathcal{L}_1$ and $\mathcal{L}_2$ are *equally expressive*, denoted as $\mathcal{L}_1 \equiv_{\mathbb{F}} \mathcal{L}_2$, iff $\mathcal{L}_1 \preceq \mathcal{L}_2$ and $\mathcal{L}_2 \preceq \mathcal{L}_1$;
- $\mathcal{L}_2$ *is more expressive than* $\mathcal{L}_1$, denoted by $\mathcal{L}_1 \prec_{\mathbb{F}} \mathcal{L}_2$, iff $\mathcal{L}_1 \preceq_{\mathbb{F}} \mathcal{L}_2$ but $\mathcal{L}_2 \not\equiv_{\mathbb{F}} \mathcal{L}_1$.

Proposition 4.3.2 (Relative expressivity of $\mathcal{L}_{\mathbf{K}}$ and $\mathcal{L}_{\mathbf{Kw}}$ [FWvD15])**.**
The basic epistemic language is more expressive than the knowing whether language over the class of all frames, serial frames, symmetric frames, transitive frames and euclidean frames. This means that, for all $\mathbb{F} \in \{\mathbb{K}, \mathbb{D}n, \mathbb{B}n, 4n, 5n\}$*, then* $\mathcal{L}_{\mathbf{Kw}} \prec_{\mathbb{F}} \mathcal{L}_{\mathbf{K}}$.

The basic epistemic language is equally expressive to the knowing whether language over the class of reflexive frames, i.e. $\mathcal{L}_{\mathbf{Kw}} \equiv_{\mathbb{T}n} \mathcal{L}_{\mathbf{K}}$.

Proof. We start by proving that for all $\mathbb{F} \in \{\mathbb{K}, \mathbb{D}n, \mathbb{B}n, 4n, 5n, \mathbb{T}n\}$, then $\mathcal{L}_{\mathbf{Kw}} \preceq_{\mathbb{F}} \mathcal{L}_{\mathbf{K}}$. Using def. 4.1.1 we define inductively a translation of any $\mathcal{L}_{\mathbf{Kw}}$ formula to an equivalent

$\mathcal{L}_{\mathbf{K}}$ formula, as follows:

$$tr(\varphi) = \begin{cases} \varphi & \text{if } \varphi \in At, \\ tr(\varphi') & \text{if } \varphi = \neg\varphi', \\ tr(\varphi') \wedge tr(\varphi'') & \text{if } \varphi = \varphi' \wedge \varphi'', \\ \mathbf{K}_a tr(\varphi') \vee \mathbf{K}_a \neg tr(\varphi') & \text{if } \varphi = \mathbf{Kw}_a \varphi'. \end{cases}$$

It is easy to prove that for all $\varphi \in \mathcal{L}_{\mathbf{Kw}}$, $tr(\varphi) \in \mathcal{L}_{\mathbf{K}}$ and $\varphi \equiv_{\mathbb{K}} tr(\varphi)$. In addition, as $\varphi \equiv_{\mathbb{K}} tr(\varphi)$, then $\varphi \equiv_{\mathbb{F}} tr(\varphi)$ for all $\mathbb{F} \in \{\mathbb{K}, \mathbb{D}n, \mathbb{B}n, 4n, 5n, \mathbb{T}n\}$

Now we prove that, for all $\mathbb{F} \in \{\mathbb{K}, \mathbb{D}n, 4n, 5n\}$, then $\mathcal{L}_{\mathbf{K}} \not\preceq_{\mathbb{F}} \mathcal{L}_{\mathbf{Kw}}$. The goal is to present two pointed models $(\mathcal{M}, w)$ and $(\mathcal{M}', w')$ with accessibility relations that are serial, transitive and euclidean, such that we can distinguish them using a $\mathcal{L}_{\mathbf{K}}$ formula but we cannot do it with a $\mathcal{L}_{\mathbf{Kw}}$ formula. This means that, there exists $\varphi \in \mathcal{L}_{\mathbf{K}}$ such that $(\mathcal{M}, w_1) \vDash \varphi$ and $(\mathcal{M}', w_1') \nvDash \varphi$, but for all $\varphi' \in \mathcal{L}_{\mathbf{Kw}}$ we have $(\mathcal{M}, w_1) \vDash \varphi'$ iff $(\mathcal{M}', w_1') \vDash \varphi'$. We use the idea, explored in example 4.4, that $\mathcal{L}_{\mathbf{Kw}}$ formulas cannot distinguish worlds with at most one accessible world, to build such models.

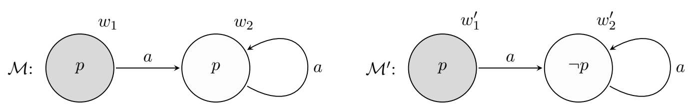

Figure 4.4: Kripke models to prove that $\mathcal{L}_{\mathbf{K}} \not\preceq_{\mathbb{F}} \mathcal{L}_{\mathbf{Kw}}$, for all $\mathbb{F} \in \{\mathbb{K}, \mathbb{D}n, 4n, 5n\}$.

We can use $\mathbf{K}_a p$ to distinguish $(\mathcal{M}, w_1)$ from $(\mathcal{M}', w_1')$. We can prove by induction on $\mathcal{L}_{\mathbf{Kw}}$ formulas, that $(\mathcal{M}, w_1) \vDash \varphi'$ iff $(\mathcal{M}', w_1') \vDash \varphi'$. The propositional case and the boolean operators cases follow straightforwardly from the fact that w_1 and w_1' have the same valuation function. For the modal operator case, we use the fact that w_1 and w_1' have only one accessible world, to prove that $(\mathcal{M}, w) \vDash \mathbf{Kw}_a\varphi$ and $(\mathcal{M}', w') \vDash \mathbf{Kw}_a\varphi$, for all $\varphi \in \mathcal{L}_{\mathbf{Kw}}$.

The models in figure 4.5 can be used to prove the same result for symmetric frames, i.e. that $\mathcal{L}_{\mathbf{K}} \not\preceq_{\mathbb{B}n} \mathcal{L}_{\mathbf{Kw}}$.

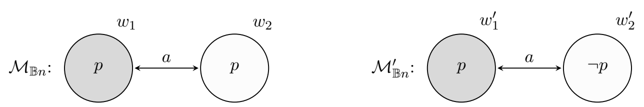

Figure 4.5: Kripke models to prove that $\mathcal{L}_{\mathbf{K}} \not\preceq_{\mathbb{B}n} \mathcal{L}_{\mathbf{Kw}}$.

To prove $\mathcal{L}_{\mathbf{Kw}} \preceq_{\mathbb{T}n} \mathcal{L}_{\mathbf{K}}$ we use the equivalence, in reflexive frames, displayed below to define an appropriate translation from $\mathcal{L}_{\mathbf{Kw}}$ formulas to $\mathcal{L}_{\mathbf{K}}$ formulas.

$$\mathbf{K}_a\varphi \equiv_{\mathbb{T}n} \mathbf{Kw}_a\varphi \wedge \varphi. \qquad \square$$

In [FWvD14] they presented the idea of $\mathbf{K}_a$ being *almost definable* with respect to $\mathbf{Kw}_a$. This is based on the observation that, for an agent a, if at w there are two $\mathcal{R}_a$ accessible worlds that can be distinguished using a contingent formula, then $\mathbf{K}_a$ can be locally defined in terms of $\mathbf{Kw}_a$.

Proposition 4.3.3 (Almost-definability schema (AD) [FWvD14])**.**
Let φ, χ be formulas in a modal language with both $\boldsymbol{K}_a$ and $\boldsymbol{Kw}_a$ modalities, then:

$$\vDash \boldsymbol{I}_a\chi \to (\boldsymbol{K}_a\varphi \leftrightarrow \boldsymbol{Kw}_a\varphi \wedge \boldsymbol{Kw}_a(\chi \to \varphi)).$$

Proof. Assume that $\vDash \mathbf{I}_a\chi$. Then, by definition 4.3.2 and validity (def. 2.2.5):

$$\text{for all pointed models } (\mathcal{M}, w) \text{ there exists } v, v' \in \mathcal{W} \text{ such that } w\mathcal{R}_a v, w\mathcal{R}_a v' \text{ and } (\mathcal{M}, v) \models \chi \text{ and } (\mathcal{M}, v') \models \neg\chi. \quad (*)$$

Consider an arbitrary pointed model $(\mathcal{M}, w)$.

$\Rightarrow$: Assume that $(\mathcal{M}, w) \vDash \mathbf{K}_a\varphi$. By definition 2.2.4, for all $v \in \mathcal{W}$ if $w\mathcal{R}_a v$ then $(\mathcal{M}, v) \models \varphi$. By propositional reasoning, for all $v \in \mathcal{W}$ if $w\mathcal{R}_a v$ then $(\mathcal{M}, v) \models \chi \to \varphi$. By definition 4.3.2 and propositional reasoning, $(\mathcal{M}, w) \vDash \mathbf{Kw}_a\varphi \wedge \mathbf{Kw}_a(\chi \to \varphi)$.

$\Leftarrow$: Assume that $(\mathcal{M}, w) \vDash \mathbf{Kw}_a\varphi \wedge \mathbf{Kw}_a(\chi \to \varphi)$. By definition 4.3.2 and propositional reasoning, for all $v, v' \in \mathcal{W}$ if $w\mathcal{R}_a v$ and $w\mathcal{R}_a v'$ then :

(i) $(\mathcal{M}, v) \models \varphi$ iff $(\mathcal{M}, v') \models \varphi$ and
(ii) $(\mathcal{M}, v) \models \chi \to \varphi$ iff $(\mathcal{M}, v') \models \chi \to \varphi$.

By our initial assumption**(*)** and **(ii)**, there exists $v, v' \in \mathcal{W}$ such that $w\mathcal{R}_a v$ and $w\mathcal{R}_a v'$ and $(\mathcal{M}, v) \models \varphi$ and $(\mathcal{M}, v') \models \varphi$. By **(ii)** and definition 2.2.4, $(\mathcal{M}, w) \vDash \mathbf{K}_a\varphi$. □

4.3.2 Minimal Non-contingency Logic $\mathcal{KW}_n$

Definition 4.3.5 (Minimal non-contingency logic $\mathcal{KW}_n$ [FWvD15])**.**
The *Hilbert system for the minimal non-contingency logic* $\mathcal{KW}_n$ over $\mathcal{L}_{\mathbf{Kw}}$, denoted by $\mathcal{KW}_n$, is defined by the following axioms and inference rules schemes, for each $a \in Ag$:

Taut. all instances of classic propositional logic tautologies;

Consistency. $\mathbf{Kw}_a(A \to B) \wedge \mathbf{Kw}_a(\neg A \to B) \to \mathbf{Kw}_a B$;

Distribution. $\mathbf{Kw}_a A \to \mathbf{Kw}_a(A \to B) \vee \mathbf{Kw}_a(\neg A \to C)$;

Equivalence. $\mathbf{Kw}_a A \leftrightarrow \mathbf{Kw}_a \neg A$;

Modus Ponens. $\dfrac{A \qquad A \to B}{B}\,{\scriptstyle MP}$;

Kw-*Necessitation.* $\dfrac{A}{\mathbf{Kw}A}\ \mathbf{Kw}Nec$;

Replacement. $\dfrac{A \leftrightarrow B}{\mathbf{Kw}A \leftrightarrow \mathbf{Kw}B}\ Repl$.

The consistency axiom is explained in example 4.3 and the equivalence axiom is the formula in 4.1. The distribution axiom tell us what we can derive from a knowing whether formula. Assume that an agent a knows whether φ. Then, he either knows φ and as a consequence he knows that $\neg\varphi$ implies any formula; or he knows $\neg\varphi$, which means that he knows that φ implies any formula.

We define below, based on the canonical model definition from previous chapter (def. 3.2.7) and the almost almost-definability schema (def. 4.3.3), the canonical model of $\mathcal{KW}_n$. This is used to prove the completeness of $\mathcal{KW}_n$ with respect to $\mathbb{K}n$.

Definition 4.3.6 (Canonical model for $\mathcal{KW}_n$)**.**
The canonical model for $\mathcal{KW}_n$ is the Kripke model $\mathcal{M}^{\mathcal{KW}} = \langle \mathcal{W}^{\mathcal{KW}}, \mathcal{R}_a^{\mathcal{KW}}, \mathcal{V}^{\mathcal{KW}} \rangle$ defined as follows:

- $\mathcal{W}^{\mathcal{KW}}$ is the set of all maximal consistent sets for $\mathcal{KW}$;
- $\mathcal{R}_a^{\mathcal{KW}}$ is a binary relation over $\mathcal{W}^{\mathcal{KW}}$, such that $(\Gamma, \Delta) \in \mathcal{R}_a^{\mathcal{KW}}$ iff:
  - $\neg\mathbf{Kw}_a\chi \in \Gamma$, and
  - for all φ and ψ: if $\mathbf{Kw}_a\varphi \wedge \mathbf{Kw}_a(\chi \to \varphi)$ then $\varphi \in \Delta$;
- $\mathcal{V}^{\mathcal{KW}}(\Gamma)(p) = true$ iff $p \in \Gamma$.

Theorem 4.3.1 (Soundness and completeness of $\mathcal{KW}_n$ for $\mathbb{K}n$ [FWvD15])**.**
The system $\mathcal{KW}_n$ is sound and complete with respect to the class of all frames $\mathbb{K}n$.

The proposition below tell us that there are frame properties that cannot be defined (see def. 3.1.1) using $\mathcal{L}_{\mathbf{Kw}}$ formulas. In particular, this means that we cannot use the *completeness-via-canonicity* proof technique, as in section 3.2.4, to prove completeness of a system for the logic of knowing whether over the class of transitive frames.

Proposition 4.3.4 (Undefinable frame properties [Zol99])**.**
The frame properties of seriality, reflexivity, transitivity, symmetry, and euclidicity are not definable in $\mathcal{L}_{\mathbf{Kw}}$.

Definition 4.3.7 (System $\mathcal{KWS}5_n$ [FWvD15])**.**
We define the system $\mathcal{KWS}5_n$ by extending $\mathcal{KW}_n$ with the following axioms schemes:

KwT. $\mathbf{Kw}_aA \wedge \mathbf{Kw}_a(A \to B) \wedge A \to \mathbf{Kw}_aB$;

Weak Kw5. $\neg\mathbf{Kw}_aA \to \mathbf{Kw}_a\neg\mathbf{Kw}_aA$.

Theorem 4.3.2 ($\mathcal{KWS}5_n$ Soundness and completeness w.r.t. $\mathbb{S}5n$ [FWvD15]).
The system $\mathcal{KWS}5_n$ is sound and complete for the class of frames $\mathbb{S}5n$.

4.4 Knowing Whether and Epistemic States

In this section, we address the *cardinality of the epistemic state space problem.* A *state of knowledge or epistemic state* is a maximally consistent set of epistemic formulas. We prove that, given two agents and a fixed propositional variable, the number of epistemic states is the same of the *continuum.*

This result was proved before using information structures, first in [Aum99] and later in [HHS96]. The proof in [Aum99] uses the notion of *knowing that.* In [HHS96], they present a shorter and more elegant proof using only *knowing whether* formulas to define the epistemic states. This operator simplifies high-order reasoning, because it does not force a truth value on the sentence it refers to. We follow the same strategy as in [HHS96] and prove this result using the logic of knowing whether.

Example 4.5 (*Differences in high-order reasoning*).
The interpretation of nested *knowing that* operators is not straightforward, because the levels are not independent. Note, for example, that $\vdash_{\mathcal{S}5n} \mathbf{K}_a\mathbf{K}_bC \to \mathbf{K}_bC$, because the formula is an instance of axiom **T** (see table 3.1).

The same does not apply to *knowing whether*, i.e.

$$\nvdash_{\mathcal{S}5n} \mathbf{Kw}_a\mathbf{Kw}_bC \to \mathbf{Kw}_bC.$$

For the Kripke model below, $\mathcal{M}$, we have:

$$(\mathcal{M}, w_1) \vDash \mathbf{Kw}_a\mathbf{Kw}_bp \text{ , but } (\mathcal{M}, w_1) \nvDash \mathbf{Kw}_bp.$$

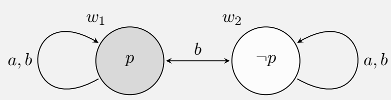

Figure 4.6: Kripke model to prove that different levels of *knowing whether* are independent.

The fact that there are uncountable many epistemic states has some interesting implications. For instance, it means that is it not possible to name all of them. Therefore, we need to be cautious when we assume that and agent is capable of communicating their epistemic state or when we assume that an agent knows its epistemic model. Moreover, given the soundness and completeness result presented in the previous chapter and the fact that there is only a countable number of finite models, then there exists epistemic states that cannot be satisfied by a finite Kripke model.

4.4.1 Countable and Cardinality of the Continuum

We review some notions about the cardinality of sets, adapted from [HJ99]. We follow the usual definition of size of a set S, denoted by $|S|$. The cardinality of a set A is less than or equal to the cardinality of B, denoted by $|A| \leq |B|$, if there is a one-to-one (injective) mapping of A into B. The set A has the same cardinality has the set B, denoted by $|A| = |B|$, if $|A| \leq |B|$ and $|B| \leq |A|$. If $|A| \leq |B|$ but $|B| \nleq |A|$, then we say that $|A| < |B|$.

Definition 4.4.1 (Countable, uncountable and cardinality of the continuum)**.**
A set S is *countable*, if $|S| = |\mathbb{N}|$. A set S is *uncountable*, if $|\mathbb{N}| < |S|$. A set S has the *cardinality of the continuum*, if $|S| = |\mathbb{R}|$.

The following two theorems are well-known results in set theory.

Theorem 4.4.1.
The set $\mathbb{R}$ *is uncountable, i.e.* $|\mathbb{N}| < |\mathbb{R}|$.

Theorem 4.4.2 (Sets with cardinality of the continuum [HJ99])**.**
$|\mathcal{P}(\mathbb{N})| = |\{0,1\}^{\mathbb{N}}| = |\mathbb{R}|$.

All languages that we work with in this thesis are countable.

Proposition 4.4.1 (Modal language is countable)**.**
Let $\mathcal{L}$ *be a modal language, then* $|\mathcal{L}| = |\mathbb{N}|$.

Proof. All modal formulas are finite. In addition, there is a finite number of operators and a countable number of propositional variables. Thus, we can define an encoding of modal formulas to natural numbers that is both a one-to-one (injection) and onto (surjection) function from $\mathcal{L}$ to $\mathbb{N}$. □

4.4.2 Cardinality of Epistemic State Space

We are interested in the number of epistemic states, when there is at least two agents. In our setting, epistemic states correspond to $\mathcal{S}5_n$ maximally consistent sets (MCS).

Definition 4.4.2 (State space defined by $\mathcal{X}$)**.**
Let $\mathcal{L}$ be a modal language and $\mathcal{X}$ a system. The *state space defined by* $\mathcal{X}$, $\mathbb{S}^{\mathcal{X}}$, is the set of all $\mathcal{X}$-MCS.

We want to prove that $|\mathbb{S}^{\mathcal{S}5_n}| = |\mathbb{R}|$. The fact that $|\mathbb{R}|$ is an upper bound for the cardinality of epistemic states follows trivially from $\mathbb{S}^{\mathcal{S}5_n} \subseteq \mathcal{P}(\mathcal{L}_{\mathbf{K}})$. Thus, we need to show that the cardinality of the *continuum* establishes, as well, a lower bound. Our strategy is to define a one-to-one mapping from the set of infinite binary sequences, $\{0,1\}^{\mathbb{N}}$, to the set of all epistemic states. We consider, in particular, the set of all epistemic states defined by $\mathcal{KWS}5_n$. This is not problematic, because by the definition of $\mathbf{Kw}_a$ (def. 4.1.1) all $\mathcal{L}_{\mathbf{Kw}}$ formulas can be translated to $\mathcal{L}_{\mathbf{K}}$ formulas. Thus, by Lindenbaum's lemma for $\mathcal{K}$ (lemma 3.2.1), there is an $\mathcal{S}5_n$-MCS for each $\mathcal{KWS}5_n$-MCS.

Our mapping works as follows. We consider two agents, *odd* and *even*, and a propositional variable, p. Each binary sequence is translated to a formula with alternated $\mathbf{Kw}_{odd}$ and $\mathbf{Kw}_{even}$ operators. The digits in a sequence define whether the operator is negated. Then, given a $s \in \{0,1\}^{\mathbb{N}}$, we collect all translations of initial segments of s in a set. We formalize this below.

Definition 4.4.3 (Translation of a digit to a know whether formula)**.**
We define the translation of digits in $\{0,1\}$ as follows:

$$\lceil 1 \rceil_a \varphi = \mathbf{Kw}_a \varphi \quad \text{and} \quad \lceil 0 \rceil_a \varphi = \neg \mathbf{Kw}_a \varphi.$$

Definition 4.4.4 (Translation of a binary string to a sequence of $\mathcal{L}_{\mathbf{Kw}}$ formulas)**.**
Let $s \in \{0,1\}^{\mathbb{N}}$. Let *odd* and *even* be two agents. We define $S_n(s)$, for all $n \in \mathbb{N}$, inductively as follows:

$$S_0(s) = \begin{cases} p & \text{if } s_0 = 1, \\ \neg p & \text{if } s_0 = 0, \end{cases}$$

$$S_n(s) = \begin{cases} \lceil s_n \rceil_{odd} S_{n-1}(s) & \text{if } n \geq 1 \text{ is odd}, \\ \lceil s_n \rceil_{even} S_{n-1}(s) & \text{if } n \geq 1 \text{ is even}. \end{cases}$$

The sequence of $\mathcal{L}_{\mathbf{Kw}}$ formulas generated by s is defined as $S(s) = \{S_n(s) \mid \text{for all } n \in \mathbb{N}\}$.

Example 4.6 (*Inconsistent set with $\mathcal{L}_{\mathbf{K}}$ formulas*).
We cannot use the mapping described above for $\mathcal{L}_{\mathbf{K}}$ formulas, because it would generate inconsistent sets. In [HHS96], they show that the following set is inconsistent:

$$\{A,\ \mathbf{K}_1 A,\ \neg\mathbf{K}_2\mathbf{K}_1 A,\ \neg\mathbf{K}_1\neg\mathbf{K}_2\mathbf{K}_1 A,\ \mathbf{K}_2\neg\mathbf{K}_1\neg\mathbf{K}_2\mathbf{K}_1 A\}.$$

The third and the fifth element of this set make it inconsistent. We prove it below.

$$\vdash_{\mathcal{S}5} \neg(\mathbf{K}_2\neg\mathbf{K}_1\neg\mathbf{K}_2\mathbf{K}_1 A \wedge \neg\mathbf{K}_2\mathbf{K}_1 A)$$

$\mathcal{S}5$-**proof** :

1.$\mathbf{K}_2\mathbf{K}_1 A \rightarrow \mathbf{K}_1 A$	*Axiom T*
2.$\neg\mathbf{K}_1 A \rightarrow \neg\mathbf{K}_2\mathbf{K}_1 A$	$PL_1(1)$ *(from example 3.2)*
3.$\mathbf{K}_1\neg\mathbf{K}_1 A \rightarrow \mathbf{K}_1\neg\mathbf{K}_2\mathbf{K}_1 A$	$N\mathcal{K}(2)$
4.$\neg\mathbf{K}_1 A \rightarrow \mathbf{K}_1\neg\mathbf{K}_1 A$	*Axiom* 5
5.$\neg\mathbf{K}_1 A \rightarrow \mathbf{K}_1\neg\mathbf{K}_2\mathbf{K}_1 A$	$MP(4,3)$
6.$\neg\mathbf{K}_1\neg\mathbf{K}_2\mathbf{K}_1 A \rightarrow \mathbf{K}_1 A$	$PL_1(5)$ *(from example 3.2)*
7.$\mathbf{K}_2\neg\mathbf{K}_1\neg\mathbf{K}_2\mathbf{K}_1 A \rightarrow \mathbf{K}_2\mathbf{K}_1 A$	$N\mathcal{K}(6)$
8.$\neg\mathbf{K}_2\neg\mathbf{K}_1\neg\mathbf{K}_2\mathbf{K}_1 A \vee \mathbf{K}_2\mathbf{K}_1 A$	*Prop. reasoning(7)*
9.$\neg(\mathbf{K}_2\neg\mathbf{K}_1\neg\mathbf{K}_2\mathbf{K}_1 A \wedge \neg\mathbf{K}_2\mathbf{K}_1 A)$	*Prop. reasoning(8)*

It is interesting to note that only the axiom **T** and **5** are used to prove this result.

We use the model defined below to prove that all infinite binary strings s generate a consistent $S(s)$. The worlds in this model are all the infinite binary strings. We connect worlds with respect to the epistemic information they encode. Thus, two worlds s, s' are connected with an agent's acessibility relation, if the agent cannot distinguish $S(s)$ from $S(s')$.

Definition 4.4.5 (Kripke model for binary sequences)**.**
The Kripke model $\mathcal{M}^{\{0,1\}} = \langle \mathcal{W}^{\{0,1\}}, \{\mathcal{R}_{odd}, \mathcal{R}_{even}\}, \mathcal{V}^{\{0,1\}}\rangle$ is defined as follows:

- $\mathcal{W}^{\{0,1\}} = \{0,1\}^{\mathbb{N}}$;
- $\mathcal{R}_{odd}$ and $\mathcal{R}_{even}$ are a binary relations over $\mathcal{W}$, such that:
  $s\mathcal{R}_{odd}s'$ iff for all odd $k \geq 1$: $s_k = s'_k$ and if $s_k = 1$ then $s_{k-1} = s'_{k-1}$;
  $s\mathcal{R}_{even}s'$ iff for all even $k \geq 1$: $s_k = s'_k$ and if $s_k = 1$ then $s_{k-1} = s'_{k-1}$;
- $\mathcal{V}^{\{0,1\}}(s)(p) = \mathit{true}$ iff $s_0 = 1$.

We use the fact that $\mathbf{Kw}_a\neg\varphi \leftrightarrow \mathbf{Kw}_a\varphi$ is valid, to prove that the model defined above satisfies all the sequence of $\mathcal{L}_{\mathbf{Kw}}$ formulas defined by a binary sequence. This validity tell us that we can disregard the tail of each sequence.

Lemma 4.4.3 (Know whether property)**.**
Let $s \in \{0,1\}^{\mathbb{N}}$ *and* $(\mathcal{M}, w)$ *a pointed Kripke model.*

If $n > 0$ *is odd, then* $(\mathcal{M}, w) \vDash S_n(s)$ *iff* $(\mathcal{M}, w) \vDash \lceil s_n \rceil_{odd} S_{n-1}(1^{n-1})$.

If $n > 0$ *is even, then* $(\mathcal{M}, w) \vDash S_n(s)$ *iff* $(\mathcal{M}, w) \vDash \lceil s_n \rceil_{even} S_{n-1}(1^{n-1})$.

Proof. We can prove this by induction on n using $\vDash \mathbf{Kw}_a\neg\varphi \leftrightarrow \mathbf{Kw}_a\varphi$. □

We prove below that all binary sequences generate a consistent set.

Lemma 4.4.4.
Let $s \in \{0,1\}^{\mathbb{N}}$ *and* $n \geq 0$*. Then,* $(\mathcal{M}^{\{0,1\}}, s) \vDash S_n(s)$*.*

Proof. We prove the statement by induction on n. Consider an arbitrary $s \in \{0,1\}^+$. The base case, $(\mathcal{M}^{\{0,1\}}, s) \vDash S_0(s)$, follows from definition of $\mathcal{M}^{\{0,1\}}$. We assume, as induction hypothesis **(IH)** that: for all $s \in \{0,1\}^+$, $(\mathcal{M}^{\{0,1\}}, s) \vDash S_n(s)$.

Case n **is even:** By lemma 4.4.3, the **(IH)** is equivalent to:

$$\text{for all } s \in \{0,1\}^+ : \ (\mathcal{M}^{\{0,1\}}, s) \vDash \lceil s_n \rceil_{even} S_{n-1}(1^{n-1}). \quad \textbf{(IH*)}$$

We want to prove for all $s \in \{0,1\}^+$ that $(\mathcal{M}^{\{0,1\}}, s) \vDash S_{n+1}(s)$. Consider an arbitrary $s \in \{0,1\}^+$.

Case $s_{n+1} = 1$**:** We want to prove that $(\mathcal{M}^{\{0,1\}}, s) \vDash \mathbf{Kw}_{odd} S_n(s)$. This is equivalent to prove that, for all $s\mathcal{R}_{odd}s'$ and $s\mathcal{R}_{odd}s''$:

$$(\mathcal{M}^{\{0,1\}}, s') \vDash S_n(s) \text{ iff } (\mathcal{M}^{\{0,1\}}, s'') \vDash S_n(s).$$

Consider arbitrary s, s'' such that $s\mathcal{R}_{odd}s'$ and $s\mathcal{R}_{odd}s''$. By definition of $\mathcal{R}_{odd}$, **(i)** $s_{n+1} = s'_{n+1} = s''_{n+1} = 1$ and **(ii)** $s_n = s'_n = s''_n$. In addition, by **(IH*)**:

$$(\mathcal{M}^{\{0,1\}}, s') \vDash \lceil s'_n \rceil_{even} S_{n-1}(1^{n-1}) \ \text{ and } \ (\mathcal{M}^{\{0,1\}}, s'') \vDash \lceil s''_n \rceil_{even} S_{n-1}(1^{n-1}).$$

Thus, by **(ii)**: $(\mathcal{M}^{\{0,1\}}, s') \vDash S_n(s)$ and $(\mathcal{M}^{\{0,1\}}, s'') \vDash S_n(s)$.

Case $s_{n+1} = 0$**:** We want to prove that $(\mathcal{M}^{\{0,1\}}, s) \vDash \neg\mathbf{Kw}_{odd} S_n(s)$. This is equivalent to prove that, there exists $s\mathcal{R}_{odd}s'$ and $s\mathcal{R}_{odd}s''$ such that:

$$(\mathcal{M}^{\{0,1\}}, s') \vDash S_n(s) \text{ and } (\mathcal{M}^{\{0,1\}}, s'') \vDash \neg S_n(s).$$

Let $s' = s$ and s'' be the same as s' except for the digit in the position n that is flipped, i.e. **(iii)** $s''(n) = |s'(n) - 1|$. As $s_{n+1} = 0$ and s, s' and s'' are identical except for the digit in the position n, then $s\mathcal{R}_{odd}s'$ and $s\mathcal{R}_{odd}s''$. By **(IH*)**:

$$(\mathcal{M}^{\{0,1\}}, s') \vDash \lceil s'_n \rceil_{even} S_{n-1}(1^{n-1}) \ \text{ and } \ (\mathcal{M}^{\{0,1\}}, s'') \vDash \lceil s''_n \rceil_{even} S_{n-1}(1^{n-1}).$$

By definition 4.4.3 and **(iii)**:

$$(\mathcal{M}^{\{0,1\}}, s') \vDash \lceil s'_n \rceil_{even} S_{n-1}(1^{n-1}) \ \text{ and } \ (\mathcal{M}^{\{0,1\}}, s'') \vDash \neg\lceil s'_n \rceil_{even} S_{n-1}(1^{n-1}).$$

Thus, by $s' = s$: $(\mathcal{M}^{\{0,1\}}, s') \vDash S_n(s)$ and $(\mathcal{M}^{\{0,1\}}, s'') \vDash \neg S_n(s)$.

Case n **is odd:** Analogous. □

Lemma 4.4.5.
The accessibility relations in $\mathcal{M}^{\{0,1\}}$ are equivalence relations.

Proof. It is easy to see that $\mathcal{R}_{odd}$ and $\mathcal{R}_{even}$ are reflexive, i.e. that for all $s \in \{0,1\}^{\mathbb{N}}$, then $s\mathcal{R}_{odd}s$ and $s\mathcal{R}_{even}s$.

We prove now that $\mathcal{R}_{odd}$ is euclidean, the case for $\mathcal{R}_{even}$ is analogous. Assume that $s\mathcal{R}_{odd}s'$ and $s\mathcal{R}_{odd}s'$, for some $s, s', s'' \in \mathcal{W}^{\{0,1\}}$. Then, by def. 4.4.5, for all odd $k \geq 1$: $s_k = s'_k = s''_k$. We can, then, prove that it follows that $s'\mathcal{R}_{odd}s''$. □

Proposition 4.4.2.
$|\mathbb{R}| = |\mathbb{S}^{\mathcal{KWS}5_2}|$.

Proof. $|\mathbb{R}| \geq |\mathbb{S}^{\mathcal{KWS}5_2}|$: Given that $\mathbb{S}^{\mathcal{KWS}5_2} \subseteq \mathcal{P}(\mathcal{L}_{\mathbf{K}})$, then it follows from proposition 4.4.1 and theorem 4.4.2, that $|\mathbb{S}^{\mathcal{KWS}5_2}| \leq |\mathbb{R}|$.

$|\mathbb{R}| \leq |\mathbb{S}^{\mathcal{KWS}5_2}|$: In definition 4.4.4 we define a translation from infinite binary strings to $\mathcal{KWS}5_2$-consistent sets, as proved by lemma 4.4.4 and lemma 4.4.5.

Consider to distinct $s, s' \in \{0,1\}^{\mathbb{N}}$. Then, by definition 4.4.4 there exists $n \in \mathbb{N}$, such that $S_n(s) = \neg S_n(s')$. By a Lindenbaum's lemma for $\mathcal{KWS}5_2$, there exist $\mathcal{KWS}5_2$-MCS $M = S(s)$ and $M' = S(s)$. Thus, we can deduce by maximally of M and M' that if $S(s) = S(s')$ then $s = s'$. This means that, our translation defines an one-to-one mapping from $\{0,1\}^{\mathbb{N}}$ to $\mathbb{S}^{\mathcal{KWS}5_2}$. □

We finish this section with the example below. We show that without negation we have finite models for infinite sets with alternated iterations of *knowing that*, *knowing whether* or *ignorance*.

Example 4.7 (*Epistemic states without negation*).
Consider a Kripke model $\mathcal{M} = \langle\{w\}, \{\mathcal{R}_1, \mathcal{R}_2\}, \mathcal{V}\rangle$, such that $(\mathcal{M}, w) \vDash A$. Then, this is a model for the following infinite sets:

$$\{(\mathbf{K}_2\mathbf{K}_1)^n A,\ \mathbf{K}_1(\mathbf{K}_2\mathbf{K}_1)^n A \mid n \in \mathbb{N}\} \quad \text{and}$$
$$\{(\mathbf{Kw}_2\mathbf{Kw}_1)^n A,\ \mathbf{Kw}_1(\mathbf{Kw}_2\mathbf{Kw}_1)^n A \mid n \in \mathbb{N}\}$$

The finite model below satisfies the set: $\{(\mathbf{I}_2\mathbf{I}_1)^n A,\ \mathbf{I}_1(\mathbf{I}_2\mathbf{I}_1)^n A \mid n \in \mathbb{N}\}$.

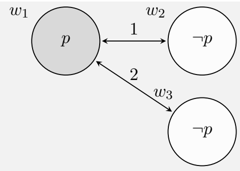

Figure 4.7: Kripke model for iterated ignorance.

CHAPTER 5

Conclusion and Future Work

In this thesis we reviewed the foundations of epistemic logic and introduced the logic of knowing whether and ignorance. Our motivation was the analysis of epistemic scenarios. We proved, in addition, that there are *continuum* many epistemic states, when we consider at least two agents.

There exists a vast literature on epistemic logic, which covers a myriad of systems built to accommodate many of its possible applications. Given its fast and fruitful development, it is not surprising that discussions related to its foundations and implicit assumptions are scattered and not developed into much detail. This poses a problem to newcomers. Moreover, such considerations are important to understand puzzles' solutions and to discuss paradoxes.

In this work we clarify the assumptions behind our decisions. In addition, even though our epistemic models are elements of $\mathbb{S}5n$, we did not narrow our analysis to these models and gave an example of a theory of knowledge that only requires reflexive models.

We use contingent modal logic to capture the notion *knowing whether*, and its dual *ignorance*. This is a *non-normal logic*. As a consequence, the logic of *knowing whether* is less expressive that *basic epistemic logic* for some classes of frames. Additionally, there are classes of frames that are not definable by contingent formulas. Therefore, there was no straightforward approach to the axiomatization of *knowing whether* in terms of the system $\mathcal{K}_n$ or $\mathcal{S}5_n$. It was, then, important to understand the role of each axiom and rule in $\mathcal{S}5_n$ to be able to compare systems for contingent logics in the literature. We use the completeness proof of $\mathcal{S}5$ with respect to the class $\mathbb{S}5$, to develop such intuitions. This completements the usual presentation of this topics in textbooks, as normally those proofs are left to the reader. In addition, while proving completeness results we worked extensively with maximally consistent sets which were used later to prove the result about the cardinality of the epistemic state space.

In our result we only proved that there are *continuum*-many epistemic states. This implies that there are epistemic states that cannot be adequately represented by a finite Kripke model. However, it does not tell us whether there are models with uncountable many connected worlds. This is an interesting topic for future work.

List of Figures

List of Tables

Bibliography

[Aum76] Robert J Aumann. Agreeing to disagree. *The Annals of Statistics*, 4(6):1236–1239, 1976.

[Aum99] Robert J Aumann. Interactive epistemology I: Knowledge. *International Journal of Game Theory*, 28(3):263–300, 1999.

[BdRV01] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal logic*. Cambridge University Press, New York, NY, USA, 2001.

[BMS98] Alexandru Baltag, Lawrence S. Moss, and Slawomir Solecki. The logic of public announcements, common knowledge, and private suspicions. In *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge*, TARK '98, pages 43–56, San Francisco, CA, USA, 1998. Morgan Kaufmann Publishers Inc.

[Che80] Brian F. Chellas. *Modal Logic: An Introduction*. Cambridge University Press, 1980.

[CL16] Alexandre Costa-Leite. Interplays of knowledge and non-contingency. *Logic and Logical Philosophy*, 25(4):521–534, 2016.

[Fan16] Jie Fan. Removing your ignorance by announcing group ignorance: a group announcement logic for ignorance. *Studies in Logic*, 9(4):4–33, 2016.

[FHVM95] Ronald Fagin, Joseph Y Halpern, Moshe Y Vardi, and Yoram Moses. *Reasoning about knowledge*. MIT Press, 1995.

[FWvD14] Jie Fan, Yanjing Wang, and Hans van Ditmarsch. Almost necessary. *Advances in Modal Logic*, 10:178–196, 2014.

[FWvD15] Jie Fan, Yanjing Wang, and Hans van Ditmarsch. Contingency and knowing whether. *The Review of Symbolic Logic*, 8(1):75–107, 2015.

[Ger07] Jelle Gerbrandy. The surprise examination in dynamic epistemic logic. *Synthese*, 155(1):21–33, 2007.

[GG97] Jelle Gerbrandy and Willem Groeneveld. Reasoning about information change. *Journal of Logic, Language and Information*, 6(2):147–169, Apr 1997.

[GG06] Paul Gochet and Pascal Gribomont. Epistemic logic. In Dov M. Gabbay and John Woods, editors, *Logic and the Modalities in the Twentieth Century*, volume 7 of *Handbook of the History of Logic*, pages 99–195. North-Holland, 2006.

[GS11] Nina Gierasimczuk and Jakub Szymanik. A note on a generalization of the muddy children puzzle. In *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge*, pages 257–264. ACM, 2011.

[H$^+$95] IL Humberstone et al. The logic of non-contingency. *Notre Dame Journal of Formal Logic*, 36(2):214–229, 1995.

[HHS96] Sergiu Hart, Aviad Heifetz, and Dov Samet. "Knowing Whether,""Knowing That," and The Cardinality of State Spaces. *Journal of Economic Theory*, 70(1):249–256, 1996.

[HII10] Wesley Halcrow Holliday and Thomas Frederick Icard III. Moorean phenomena in epistemic logic. *Advances in Modal Logic*, 8, 2010.

[Hin62] J. Hintikka. *Knowledge and belief: an introduction to the logic of the two notions*. Contemporary Philosophy. Cornell University Press, 1962.

[HJ99] K. Hrbacek and T. Jech. *Introduction to Set Theory*. Taylor & Francis, 1999.

[Hol13] Wesley H Holliday. Epistemic logic and epistemology. *Handbook of Formal Philosophy*, 2013.

[HS15] Vincent Hendricks and John Symons. Epistemic logic. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2015 edition, 2015. `https://plato.stanford.edu/archives/fall2015/entries/logic-epistemic/`.

[Kuh95] Steven T. Kuhn. Minimal non-contingency logic. *Notre Dame J. Formal Logic*, 36(2):230–234, 04 1995.

[Kyl15] Brent G Kyle. The new and old ignorance puzzles: How badly do we need closure? *Synthese*, 192(5):1495–1525, 2015.

[LM13] Pierre Le Morvan. Why the standard view of ignorance prevails. *Philosophia*, 41(1):239–256, 2013.

[LM15] Pierre Le Morvan. On the ignorance, knowledge, and nature of propositions. *Synthese*, 192(11):3647–3662, Nov 2015.

[MFM+16] Tim Miller, Paolo Felli, Christian J Muise, Adrian R Pearce, and Liz Sonenberg. 'Knowing whether' in Proper Epistemic Knowledge Bases. In *AAAI*, pages 1044–1050, 2016.

[MH95] J.-J. Ch. Meyer and W. van der Hoek. *Epistemic logic for AI and computer science*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 1995.

[Moo77] Robert C Moore. Reasoning about knowledge and action. In *Proceedings of the 5th international joint conference on Artificial intelligence-Volume 1*, pages 223–227. Morgan Kaufmann Publishers Inc., 1977.

[MR66] Hugh Montgomery and Richard Routley. Contingency and non-contingency bases for normal modal logics. *Logique et analyse*, 9(35/36):318–328, 1966.

[Pac13] Eric Pacuit. Dynamic epistemic logic II: Logics of informaiton change. *Philosophy Compass*, 8(9):815–833, 2013.

[Par02] Rohit Parikh. States of knowledge. *Electronic Notes in Theoretical Computer Science*, 67(Supplement C):49 – 60, 2002. WoLLIC'2002, 9th Workhop on Logic, Language, Information and Computation.

[Par03] Rohit Parikh. Levels of knowledge, games, and group action. *Research in Economics*, 57(3):267–281, 2003.

[Pee10] Rik Peels. What is ignorance? *Philosophia*, 38(1):57–67, 2010.

[PK92] Rohit Parikh and Paul Krasucki. Levels of knowledge in distributed systems. *Sadhana*, 17(1):167–191, 1992.

[Pla89] J. Plaza. Logics of public communications. *Proceedings of the Fourth International Symposium on Methodologies for Intelligent Systems: Poster Session Program*, pages 201–216, 1989.

[Rei01] Raymond Reiter. *Knowledge in action: logical foundations for specifying and implementing dynamical systems*. MIT press, 2001.

[Sor17] Roy Sorensen. Epistemic paradoxes. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2017 edition, 2017. https://plato.stanford.edu/archives/spr2017/entries/epistemic-paradoxes/.

[Ste17] Matthias Steup. Epistemology. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2017 edition, 2017. `https://plato.stanford.edu/archives/fall2017/entries/epistemology/`.

[vB01] Johan van Benthem. *Correspondence Theory*, pages 325–408. Springer, Dordrecht, 2001.

[vB04] Johan van Benthem. What one may come to know. *Analysis*, 64(282):95–105, 2004.

[vB10] Johan van Benthem. *Modal logic for open minds*. Center for the Study of Language and Information, 2010.

[vB14] Johan van Benthem. *Logical dynamics of information and interaction*. Cambridge University Press, New York, NY, USA, 2014.

[vDF16] Hans van Ditmarsch and Jie Fan. Propositional quantification in logics of contingency. *Journal of Applied Non-Classical Logics*, 26(1):81–102, 2016.

[vdHL04] Wiebe van der Hoek and Alessio Lomuscio. *A Logic for Ignorance*, pages 97–108. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[vDHvdHK15] H. van Ditmarsch, J.Y. Halpern, W. van der Hoek, and B.P. Kooi. *Handbook of Epistemic Logic*. College Publications, 2015.

[vDK15] Hans van Ditmarsch and Barteld Kooi. *One hundred prisoners and a light bulb*. Springer, 2015.

[VDVdHI12] Hans Van Ditmarsch, Wiebe Van der Hoek, and Petar Iliev. Everything is knowable–how to get to know whether a proposition is true. *Theoria*, 78(2):93–114, 2012.

[vDvdHK07] Hans van Ditmarsch, Wiebe van der Hoek, and Barteld Kooi. *Dynamic Epistemic Logic*. Springer Publishing Company, Incorporated, 1st edition, 2007.

[vW51] G. H. von Wright. *An Essay in Modal Logic*. Amsterdam: North-Holland Pub. Co., 1951.

[Wan18] Yanjing Wang. *Beyond Knowing That: A New Generation of Epistemic Logics*, pages 499–533. Springer, 2018.

[Wil00] Timothy Williamson. *Knowledge and its limits*. Oxford University Press, 2000.

[Zol99] Evgeni E. Zolin. Completeness and definability in the logic of noncontingency. *Notre Dame J. Formal Logic*, 40(4):533–547, 10 1999.