Knowledge Sharing among Ideal Agents

Alessio Lomuscio

A thesis submitted
to the School of Computer Science
of The University of Birmingham
for the degree of Doctor of Philosophy

School of Computer Science
The University of Birmingham
Birmingham B15 2TT
U.K.

May 1999

A Laura

Abstract

Multi-agent systems operating in complex domains crucially require agents to interact with each other. An important result of this interaction is that some of the private knowledge of the agents is being shared in the group of agents. This thesis investigates the theme of knowledge sharing from a theoretical point of view by means of the formal tools provided by modal logic.

More specifically this thesis addresses the following three points.

First, the case of hypercube systems, a special class of interpreted systems as defined by Halpern and colleagues, is analysed in full detail. It is here proven that the logic S5WD$_n$ constitutes a sound and complete axiomatisation for hypercube systems. This logic, an extension of the modal system S5$_n$ commonly used to represent knowledge of a multi-agent system, regulates how knowledge is being shared among agents modelled by hypercube systems. The logic S5WD$_n$ is proven to be decidable. Hypercube systems are proven to be synchronous agents with perfect recall that communicate only by broadcasting, in separate work jointly with Ron van der Meyden not fully reported in this thesis.

Second, it is argued that a full spectrum of degrees of knowledge sharing can be present in any multi-agent system, with no sharing and full sharing at the extremes. This theme is investigated axiomatically and a range of logics representing a particular class of knowledge sharing between two agents is presented. All the logics but two in this spectrum are proven complete by standard canonicity proofs. We conjecture that these two remaining logics are not canonical and it is an open problem whether or not they are complete.

Third, following a influential position paper by Halpern and Vardi, the idea of refining and checking of knowledge structures in multi-agent systems is investigated. It is shown that, Kripke models, the standard semantic tools for this analysis are not adequate and an alternative notion, Kripke trees, is put forward. An algorithm for refining and checking Kripke trees is presented and its major properties investigated. The algorithm succeeds in solving the famous muddy-children puzzle, in which agents communicate and reason about each other's knowledge.

The thesis concludes by discussing the extent to which combining logics, a promising new area in pure logic, can provide a significant boost in research for epistemic and other theories for multi-agent systems.

Contents

Preface and acknowledgements

A posteriori, beginnings and ends are always easy to recognise.

For the case of this thesis, it all started with my interview in Birmingham in September 1996. At the end of it I was asked what occupation I was hoping to have in ten years time. Back then I *knew* even less English than now and I was simply trying to guess what the natives were saying. In that case my interpretation was "What were you doing when you were ten?"; without hesitation I answered "playing football all the time".

The end of it started with a trip to London in September 1998 for another interview, this time for a job at Queen Mary and Westfield College. At the time my *knowledge* of London was very limited and I thought I would leave about fifty minutes to go from Euston Station to Mile End where QMW is located. No one had told me that travelling times in London can vary enormously depending on many conditions. Needless to say I was late.

Two things come to mind. The first is the puzzlement about the fact that I was actually offered both positions. At time of writing, I am still not sure how that happened.

The second is that the beginning and the end of the time I was involved with this research are truly marked by events in which I did not *know* something crucial that many other people around me knew. I find it peculiar that the research I carried out between these two events is actually about various degrees of how knowledge can be shared among a community of agents.

Acknowledgements. I would like to thank all the persons with which I have had discussions on some of the issues of this thesis. In particular, I would like to mention: Marco Colombetti, Robert Goldblatt, Guido Governatori, Joe Halpern, Manfred Kerber, Ron van der Meyden, Aaron Sloman and Timothy Surendonk who have all provided me with interesting remarks about my work.

A big special thanks goes to my friend Mathias Kegelmann for being such a rigorous test-bed for many of the proofs here contained, for providing valuable ideas and especially for having done all this without saying "After all this is simply this-and-that category and therefore...". Well, at least not too often!

Most of all I would like to thank my supervisor, Mark Ryan. From the very beginning of my application for a PhD to the very end of the final modifications of these pages he has offered me what I found to be an extremely qualified supervision. I am very grateful to him for this and possibly even more for all the help he has offered me on matters not at all related to this work. This has really played a fundamental role in creating the conditions to carry out this research.

My studies in Birmingham were supported by an internal grant of the School of Computer Science of the University of Birmingham; this support is gratefully acknowledged.

Chapter 1

Introduction

1.1 Agent technology

Under the caption name of "Agent" lives one of the potentially ground-breaking paradigms that have hit Computer Science since the advent of Object-Oriented Programming [Mey88, Boo86]. Agents (see [WJ95, Woo97] for short reviews, or [Je98, Bra97] for more extensive expositions), like expert systems and others, are one of the paradigms that originated in Artificial Intelligence (AI) and that were quickly taken up by many computer scientists not necessarily working in AI. Indeed, it is striking to note how many Computer Science sub-disciplines have now well-established research areas that contain the keyword "agent-based" in their titles. The international community has seen research in agents for automatic diagnosis [Sch98], agents for control systems [GL87], agents for Internet-based information retrieval [EW94], agents for telecommunication [HB99b, HB99a], agents for automatic negotiation [PSJ98], etc.

The agent-paradigm has even gone beyond the boundaries of AI and Computer Science where it was born and has influenced Psychology [CC95, Cas98], Economics [Wel93], and other disciplines. Only in Computer Science there are currently no less than ten annual or bi-annual international conferences on agents themes, and more than a handful of international journals. So many different theories, languages and architectures for agents have been proposed that agent *standards* are currently under investigation (see [FIP] for example) in order to provide a more systematic technology transfer to the industry.

Indeed, even more striking for an idea forged in academic departments, big business is also fascinated by the concept. In 1998, the pioneer agent company "Firefly" was sold to Microsoft for approximately US$40M. Similarly, NetBot, the company set up by agent pioneers Oren Etzioni and Dan Weld, was acquired by the search engine company Excite for a similar sum in early 1998. Still in 1998, even Excite was taken over for about US$7Billion and in January 1999 a major investment of News Corporation, the global media conglomerate controlled by Rupert Murdoch, in Yahoo was unveiled.

Microsoft, Sun, Hewlett Packard, Mitsubishi Electronics, British Telecom, just to name a few, all have internal ongoing research on agent-related topics.

So, what are these agents and what is all this fuss about? Many definitions of agents have been proposed [RN95, Mae95, HR95, WJ95, Ld95]; I will not argue about any of these, nor will I discuss which ones and why are perhaps the more appropriate and I simply refer the interested reader to [FG96b] for some of the most recent discussions. Instead, in line with much of the literature in this area, we will use the working definition of agent given in [WJ95] to denote a *self-contained problem solving system capable of autonomous, reactive, pro-active, social behaviour*. More precisely, as it will become clear in the next section, we will follow Dennet's (see [Den87]) "intentional's stance" approach by assuming that an agent's behaviour can usefully be described in terms of collections of mental attitudes such as *knowledge, beliefs, intentions, desires, obligations*, etc. Clearly, depending on the application only a subset of these will be employed.

As to why agents are attracting so much attention, agents are new abstraction tools and as such they are increasingly seen as a most promising approach to developing complex distributed computing systems. Applications built upon the agent paradigm (agent-based systems or simply multi-agent systems (MAS)) already cover a wide number of areas including electronic commerce [CM96], information management [Mae94], health care [HJF95], process control [JCL95], electronic games [WG96], manufacturing [Par98], etc. (for a comparative survey of these applications, see [JW95, JW98] and [Par96]). The nowadays general feeling about developing agent-based software can perhaps be summarised by using the own words of one of the many companies selling packages for agent-based programming:

> Software developers and system designers use high-level abstractions in building complex software for one reason; to manage complexity. An abstraction focuses on the important and essential properties of a problem and hides the incidental components of that problem. Agents provide a new way of managing complexity because they provide a new way of describing a complex system or process. Using agents, it is easy to define a system in terms of agent-mediated processes.
> ([AGB], front page).

With so much different intellectual and economic investment in the idea of agent-based computing, it should not be too surprising that the term "agent" is seriously running the risk of becoming a conceptually overloaded term and agent-based technology an ill-defined software architecture (and sometimes quite simply a vehicle for re-branding object-oriented software for distributed computing).

Indeed some researchers (see for example [WJ98]) have already pointed out the real risks for agent-based systems of a backslash similar to the one happened to, for example, expert systems that were once guaranteed to change the face of computing as we knew it.

My opinion on this is that after more than 5 years years since the real boom in interest in MAS (perhaps to be identified by the publication of [Sho93]) we still desperately need *to ground* our intuitions as to what properties MAS should have and how we should go about building a MAS system.

This thesis offers no contribution to the latter issue (see [KG97, Woo97] for some recent approaches) but tries to enhance our understanding on the former. Arguably the best option to try to understand and explore a complex concept as the one of agent is to use a formal language. Among many that have been proposed (Z for example [Ld95] but see [dFL$^+$97] for a short discussion) here I chose formal logic to carry out my analysis. The reasons for this are threefold. Firstly, logic is now a fully-fledged formal technique that has been pursued with great success for many years for its own sake[1] and with respect to Computer Science[2].

[1] For example propositional logic, first order logic, second order logic, modal logics, linear logics, etc.

[2] Complexity theory, non-monotonic logics, automatic theorem-proving, model-checking, logic program-

Secondly, logic has been proven to be particularly suitable as formal model of key characteristics of agents such as their knowledge, beliefs, intentions, etc. and some results on these topics are already available. Thirdly, and perhaps most importantly, logic-based methods are now clearly valuable tools for the task of *validation* and *verification* [MP91, MP95, CES86] of computing systems[3]. This is indeed crucial if we are to deliver on our promises of using agents to build safe and reliable systems to fly airplanes, to control space missions, to operate automatic transactions on the Internet, to diagnose automatically telecommunication networks and so on. In fact in all these areas the techniques now currently used to certify software (e.g. testing [Mye79, Bei84, Bei90]) have been proven in some cases quite unsuccessful; the Intel Pentium division bug [Pra95], the explosion of Arianne V [Dow97] being perhaps the most distressing.

In this thesis I will use formal logic as a working technique to reason about key characteristics of MAS. By presenting some technical results achieved by using this tool, my ultimate aim is to show that some very basic questions on MAS theories are still unanswered and that logic is a good tool for such an investigation. More specifically, this work will focus on interactions in the private knowledge of agents in a MAS. Quite surprisingly, little work has been done in this area so far and my intention is to contribute to this by carrying out a case study analysis at different level of abstraction (logical, semantical, and with a more low-level descriptive language).

Given this assumption, the content of this work is bound to be quite technical and this is why in the next section we will give a quick overview at some of the technical tools that we will be using through this thesis. This is needed for two reasons. First, because it will give formal logic the up-front role that it is going to have in the rest of this work. Secondly, and more importantly, because it will introduce the frame of reference for the discussion that it will follow on MAS theories and on the reasons why this thesis is an attempt to contribute to them.

1.2 Modal logic

If one is to use formal logic to reason about MAS the obvious question to ask is which logic to adopt. As we will observe in Section 1.3, one of the key mental states of an agent is its beliefs and it can be observed that first-order logic [End72], the natural candidate, is actually not the best option available. For example, for representing beliefs, one could think about formulae like

$$believes(tweety, wrote(lewis\text{-}carrol, alice\text{-}in\text{-}wonderland)).$$

This formula raises two types of problems. The first one is syntactic: the formula above is not a legal first-order formula because the predicate $wrote$ occurs in the scope of the predicate $believes$. This could be overcome (at the expense of a much heavier formalism) by using one of the extensions that have been presented[4] to allow for this kind of expressions. The second

ming, functional programming, etc.

[3]This is *not* handwaving. Big commercial enterprises like Intel, British Telecom and others are now currently investing much financial resources into the use of formal methods (model checking for example) to valididate and verify hardware and software. The reason they are doing this is that the development of formal tools for validation and verification are currently seen as more economical than testing on a mid-long term perspective.

[4]See for example [Mon63], [Tho80a] and the more recent [Tur90], and [Dav93].

one is semantical and is bound to the fact that states like belief, desire, etc. are *referentially opaque*. To see what we mean by this consider a reasonable interpretation in which we have the equality of terms $lewis\text{-}carrol = charles\text{-}dogdson$. By substitution of terms we may now derive

$$believes(tweety, wrote(charles\text{-}dodgson, alice\text{-}in\text{-}wonderland);$$

but this is counterintuitive as agent 'Tweety' may be in a situation in which he does not believe that Lewis Carroll is actually a pseudonym for Charles Dodgson.

Modal logic offers a neat solution to this and other problems related to representing mental states and it also comes with a semantics which is more intuitive for this task. This is the reason why this thesis builds upon some techniques developed in propositional modal logic, whose knowledge I will have to assume from the reader. The aim of the rest of this section is not to present an introduction to modal logic, but to fix the notation we will be using throughout the thesis and to provide references for some of the known results that we will be using later on. Proofs of the theorems reported in this section can be found in most of the references below.

The first systematic approach to modal logic can be found in [LL59]. More recent and very good introductions to the subject are [Che80, HC84, Boo93, Pop94, HC96] and the more concise [Gol92]. For the similarities of themes and motivations [MH95] is also an excellent reference and it has provided a basis for this work. Among these, [Pop94] and [MH95] are the only one taking a full multi-modal perspective from the beginning. Recently, more advanced works have been presented (see [CZ97], [Sur98]) or are in preparation (see [BdRV99]) showing how active the field is, especially with respect to Computer Science. Modal propositional logic in itself builds upon propositional calculus whose standard references are [Men64] and the more recent [Gam91, Ham78].

1.2.1 Syntax

As standard, sets $X, Y, \ldots$ will be denoted by Italic capital letters. If X is a set, $|X|$ denotes its cardinality, and id_X the identity relation on it. If $\sim$ is an equivalence relation on X and $x \in X$, then $X/\sim$ is the set of equivalence classes of X , and $[x]_\sim$ is the equivalence class containing x.

We assume a countable set $P = \{p, q, \ldots\}$ of *propositional atoms*, and a finite set $A = \{1, \ldots, n\}$ of *agents*. For most of this work, our formal language $\mathcal{L}$ is given by the usual grammar:

$$\phi \ ::= \ p \mid \neg\phi \mid \phi_1 \wedge \phi_2 \mid \Box_i \phi$$

where $p \in P, i \in A$. We will make it clear when we are operating in a richer language.

The other propositional and modal connectives can be defined in the usual way, in particular:

$$\Diamond_i \phi \ ::= \neg\Box_i\neg\phi$$

Everything which has not been defined or discussed here is to be interpreted in line with the standard tradition.

1.2.2 Semantics

Throughout this thesis we will only be using the traditional "possible-worlds" semantics proposed in the modern form by Kripke in [Kri63] and the one of "interpreted systems" first put forward by Halpern and Fagin in [HF89]. Other approaches are possible, for example algebraic semantics (which dates back to [McK41]) was the first semantics for modal logic.

Like in Section 1.2.1 as throughout this work we assume a set $A = \{1, \dots, n\}$ of agents. We first recall the basic definitions of Kripke semantics.

Kripke frames were first proposed in [Car46] and later developed in [Bay58, Hin57, Kan57, Mon60, Pri62]. The formalisation presented here and widely used nowadays was first presented in [Kri59, Kri63] and later advocated in [Lem77][5].

Definition 1.1 (Kripke frames and Kripke models). *A* frame *F is a tuple $F = (W, R_1, \dots, R_n)$, where W is a non-empty set of points or worlds $W = \{w, \dots\}$ and $R_i, i \in A$, are binary relations on W. If all the relations are equivalence relations, the frame is an* equivalence frame *and we write $\sim_i$ for R_i.*

A model *M is a tuple $M = (W, R_1, \dots, R_n, \pi)$, where $(W, R_1, \dots, R_n)$ is its underlying frame and $\pi : P \to 2^W$ is an interpretation for the atoms. An* equivalence model *is a model whose underlying frame is an equivalence frame.*

Committing an abuse of notation, given a frame $F = (W, R_1, \dots, R_n)$ and an interpretation π, we will sometimes denote $M = (W, R_1, \dots, R_n, \pi)$ as $M = (F, \pi)$. Also we will denote $M = (W, R_1, \dots, R_n, \pi)$ as $M = (W, \{R_i\}_{i \in A}, \pi)$ and $F = (W, R_1, \dots, R_n)$ as $F = (W, \{R_i\}_{i \in A})$. If $A = \{1, \dots, n\}$ is clear from the context, we will sometimes simply write $F = (W, \{R_i\}_i)$ and $M = (W, \{R_i\}_i, \pi)$.

Satisfaction on Kripke structures is defined as follows:

Definition 1.2 (Satisfaction). *The satisfaction of a formula ϕ in a world w of a model M, formally $M \models_w \phi$, is inductively defined as follows:*

$M \models_w p$	*if*	$w \in \pi(p)$,
$M \models_w \neg\phi$	*if*	$M \not\models_w \phi$,
$M \models_w \phi \wedge \psi$	*if*	$M \models_w \phi$ *and* $M \models_w \psi$,
$M \models_w \Box_i \psi$	*if*	*for each* $w' \in W$ *we have* $w R_i w'$ *implies* $M \models_{w'} \psi$.

Satisfaction for the other logical connectives can be defined in the usual way.

Validity on Kripke structures is defined as follows:

Definition 1.3 (Validity). *A formula ϕ is valid on a model $M = (W, R_1, \dots, R_n, \pi)$, formally $M \models \phi$, if for any point $w \in W$ we have $M \models_w \phi$. A formula ϕ is valid on a frame $F = (W, R_1, \dots, R_n)$ if for any interpretation π we have $(F, \pi) \models \phi$. A class of models $\mathcal{M}$ validates a formula ϕ if for any model $M \in \mathcal{M}$ we have $M \models \phi$. A formula ϕ is valid on a class of frames $\mathcal{F}$ if for any frame $F \in \mathcal{F}$ we have $F \models \phi$.*

We now turn our attention to interpreted systems; we follow the definition in the form given in [FHMV95]. Given a set $A = \{1, \dots, n\}$, consider n sets of local states, one for every agent of the MAS and a set of states for the environment. We denote by L_i the non-empty set of local states possible for agent i, and by L_e the non-empty set of possible states for the environment. Elements of L_i will be denoted by $l_i, m_i, \dots$ Elements of L_e will be denoted by $l_e, m_e, \dots$

[5]The term "frame" was first suggested by Dana Scott and used by Segerberg in [Seg68].

Definition 1.4 (Global states of interpreted systems). A set of global states for an interpreted system *is a subset of the Cartesian product* $S \subseteq L_1 \times \cdots \times L_n \times L_e$.

A global state represents the situation of all the agents and of the environment at a particular instant of time. The idea behind considering a subset is that some of the tuples that originate from the Cartesian product might not be possible because of explicit constraints present in the MAS. By considering functions (runs) from the natural numbers to the set of global states, it is possible to represent the temporal evolution of the system. An interpreted system is a set of functions on the global states with a valuation for the atoms of the language.

Although in this work we will discuss temporal evolution of MAS, we will not consider runs of interpreted systems explicitly. So, for us the key definition will be the one of global states (Definition 1.4) and an interpreted system will simply be defined on global states. We call these *static interpreted systems*.

Definition 1.5 (Static interpreted systems). *A* static interpreted system *is a tuple* $IS = (S, \pi)$, *where* S *is a set of global states as in Definition 1.4 and* $\pi : P \to 2^S$ *is an interpretation for the atoms.*

Any static interpreted system is immediately suitable to interpret a propositional language on its set of global states. We now see how they can be adapted to interpret a multimodal language.

Consider a family of first-order predicates $\{P_i\}_{i \in A}, P_i \subseteq S \times S$; we can use these to specify when two global states are related for agent i.

Definition 1.6. *Given a static interpreted system* $IS = (S, \pi)$, *consider the Kripke model* $M_{IS} = (F, \pi)$, *where the frame* $F = (S, R_1, \ldots, R_n)$ *is defined by:* sR_is' *if* $P_i(s, s')$. *The model* M_{IS} *is called the model* generated *by* IS.

This construction can be regarded as a way of seeing interpreted systems as a special class of Kripke models.

When IS is clear from the context we will omit the subscript IS from M. If $M_{IS} = (W, \{R_i\}_i, \pi)$ is the model generated by IS, then $F_{IS} = (W, \{R_i\}_i)$, the frame underlying M_IS, is also called the frame generated by the set S of global states.

In the applications (logics for knowledge for example), the family of predicates $\{P_i\}_{i \in A}$ is fixed in advance; if that is the case we can then define validity on static interpreted systems by relying on the generated Kripke models.

Definition 1.7. *A formula* ϕ *is valid on a static interpreted system* IS, $IS \models \phi$, *if* $M_{IS} \models \phi$. *A formula* ϕ *is valid on a set of global states* S, $S \models \phi$, *if for any interpretation* π *we have that* $(S, \pi) \models \phi$. *A formula* ϕ *is valid on a class of static interpreted systems* $\mathcal{IS}$ *if for every* $IS \in \mathcal{IS}$ *we have* $IS \models \phi$. *A formula* ϕ *is valid on a class of global states* $\mathcal{S}$ *if for every* $S \in \mathcal{S}$ *we have that* $S \models \phi$.

1.2.3 Proof theory

As is known, modal logic was historically introduced simply by giving an axiomatic account of various modal logic systems, but without relating these to any semantics. In particular the intuitive possible-worlds semantics that we described in Section 1.2.2 was introduced some twenty years after the first modal systems were introduced (the system S3 described in

[Lew18]). Modal systems were simply described as sets of formulae that could be deduced from axioms and rules of inference. This Hilbert-style way of expressing a logic is sometimes referred to as "syntactic approach" in contrast to the "semantic approach" described in the previous subsection.

Definition 1.8. *A* normal modal system L_n*, where* n *is the number of modal boxes of the logic, is the set of formulae that can be deduced from a set of axioms and inference rules. The set of axioms includes the following:*

Taut	*Any propositional tautology*
K	$\Box_i(p \Rightarrow q) \Rightarrow (\Box_i p \Rightarrow \Box_i q)$

The set of inference rules include the following:

Uniform Substitution	*The result of uniformly replacing any propositional variables* $p_1, \ldots, p_n$ *in a theorem by any formula* $\phi_1, \ldots, \phi_n$ *is itself a theorem*[6]*,*
Modus Ponens	*If* ϕ *and* $\phi \Rightarrow \psi$ *are theorems so is* ψ*,*
Necessitation	*If* ϕ *is a theorem, so is* $\Box_i \phi$*, for any* $i \in A$*.*

Sometimes we will use the short-cuts US for the rule of uniform substitution and similarly MP and Nec for modus ponens and necessitation. In the following we will also extensively use the term "logic" for normal modal system. We write $\vdash_{L_n} \phi$ to mean that the formula ϕ is a theorem of the logic L_n, and, as expected, we write $\not\vdash_{L_n} \phi$ to mean that it is not the case that $\vdash_{L_n} \phi$.

Definition 1.9 (Maximal consistency). *A formula* ϕ *is* L_n-consistent *(or simply* consistent *if* L_n *is clear from the context) if* $\not\vdash_{L_n} \neg\phi$*. A finite set of formulae* $\{\phi_1, \ldots, \phi_m\}$ *is* L_n*-consistent if* $\not\vdash_{L_n} \neg(\bigwedge_{i=1,\ldots,m} \phi_i)$*. An infinite set of formulae* Φ *is* L_n*-consistent if any finite subset of* Φ *is* L_n*-consistent.*

A set Φ *of formulas is* maximal *if for every formula* ϕ *we have that either* $\phi \in \Phi$ *or* $\neg\phi \in \Phi$*.*

A set Φ *is* L_n-maximal consistent *if it is both L-consistent and maximal.*

Formulae and sets of formulae are inconsistent if it is not the case that they are consistent. The importance of consistent sets is mainly due to the fact that every set admits a maximal consistent extension:

Lemma 1.10 (Lindembaum's Lemma). *Let* L_n *be a normal modal logic. Given an* L_n*-consistent set of formulae* Φ*, there is a maximal* L_n*-consistent set* Γ *such that* $\Phi \subseteq \Gamma$*.*

The maximal extension Γ is not necessarily unique.

We now introduce the logic systems that will be used as reference throughout this thesis. The weakest normal modal logic (i.e. the one containing the smallest number of theorems) is the logic K_n. The logic K_n is obtained from Definition 1.8 by stipulating that no extra axioms and inference rules are included.

Stronger logics can be defined by enriching the list of axioms and inference rules. Figure 1.1 will serve as a reference for the names of the axioms we will be using in this work.

[6] In the following we will use the expression $\phi[\psi_1/p_1, \ldots, \psi_n/p_n]$ to denote the formula which results from ϕ by replacing $p_1, \ldots, p_n$ uniformly by $\psi_1, \ldots, \psi_n$.

Axiom	*Axiom name*
$\Box_i\phi \Rightarrow \phi$	T
$\Box_i\phi \Rightarrow \Diamond_i\phi$	D
$\Box_i\phi \Rightarrow \Box_i\Box_i\phi$	4
$\Diamond_i\phi \Rightarrow \Box_i\Diamond_i\phi$	5
$\Diamond_i\Box_i\phi \Rightarrow \Box_i\Diamond_i\phi$	G1
$\Box_i\Diamond_i\phi \Rightarrow \Diamond_i\Box_i\phi$	M
$\phi \Rightarrow \Box_i\Diamond_i\phi$	B
$\phi \Leftrightarrow \Box_i\phi$	Triv

Figure 1.1: The traditional reference names for some important axioms.

Logic	*Name of the logic*
K_n + T	T_n
K_n + D	KD_n
K_n + 4	$K4_n$
K_n + Triv	$Triv_n$
T_n + 4	$S4_n$
$S4_n$+ G1	$S4.2_n$
T_n + 5	$S5_n$
D_n + 4	$KD4_n$
$KD4_n$ + 5	$KD45_n$

Figure 1.2: The traditional reference names for some important logics.

We will use only modal logics that can be expressed by using Definition 1.8 in which no extra inference rules are included. So, we will be able to express any logic of interest L_n as the system K_n to which we add a set of axioms Φ. In doing this we will sometimes write $L_n = K_n + \Phi$. Often the set Φ will be a singleton. Historically some logics have been given particular names, that we will also use and that we report in Figure 1.2.

It should be noted that a certain logic can be defined in many ways by adding different axioms to the system K_n and Figure 1.2 is only a possible definition. For example the logic $S5_n$ above defined as $S5_n = T_n + 5$ can also be defined as $S5_n = S4_n + B$. Many more systems are known and this list serves only as a reference.

We will also make use a standard shortcut: If a system L can be axiomatised by taking a logic M and adding an axiom whose name is A, by MA we will denote the logic L = M + A.

1.2.4 Methodologies

Over the last thirty years, many formal techniques have been developed for the study of modal logics grounded on Kripke semantics, such as completeness proofs via canonical models [Mak66, Kap66, Lem77], decidability via filtrations [Lem77], and more recently completeness and decidability transfer via combining logics [KW91, Gab96a]. Indeed part of this heritage of techniques constitutes the pillars upon which most of this thesis is built upon.

To fix the notation I present here some results that we will use later on. The reader is assumed familiar with these and they serve as a reference only. For the proofs the reader is referred to any of the general references for modal logic mentioned before.

Definition 1.11 (Isomorphism of frames). *Two frames $F = (W, R_1, \ldots, R_n)$, $F' = (W', R'_1, \ldots, R'_n)$ of a class $\mathcal{F}$ of frames are* isomorphic *($F \cong_{\mathcal{F}} F'$) if and only if:*

- *There exists a bijection $b : W \to W'$,*
- *For all $s, t \in W$, and all $i \in A$, sR_it if and only if $b(s)R'_ib(t)$.*

Isomorphic frames validate the same formulae.

Lemma 1.12. *If F and F' are isomorphic frames then for all ψ we have $F \models \psi$ if and only $F' \models \psi$.*

An equivalent notion of isomorphism can be defined on models. A result equivalent to Lemma 1.12 holds for that case.

Isomorphism is a very strong property. A weaker interesting property is *p-morphism*. We can define these both at the level of frames and at the level of Kripke models.

Definition 1.13 (P-morphism). *A* frame p-morphism *from $F = (W, R_1, \ldots, R_n)$ to $F' = (W', R'_1, \ldots, R'_n)$ is a function $p : W \to W'$ such that:*

1. *the function p is surjective,*
2. *for all $u, v \in W$ and each $i = 1 \ldots n$, if uR_iv then $p(u)R'_ip(v)$,*
3. *for each $i = 1 \ldots n$ and $u \in W$ and $v' \in W'$, if $p(u)R'_iv'$ then there exists $v \in W$ such that uR_iv and $p(v) = v'$.*

If $M = (W, R_1, \ldots, R_n, \pi)$ and $M' = (W', R'_1, \ldots, R'_n, \pi')$ are Kripke structures, then a model p-morphism *from M to M' is a mapping $p : W \to W'$ that is a frame p-morphism from $(W, R_1, \ldots, R_n)$ to $(W', R'_1, \ldots, R'_n)$ that satisfies $q \in \pi'(p(w))$ if and only if $q \in \pi(w)$ for all propositions q and points $w \in W$.*

If there is a p-morphism from F to F', F' is also said to be a p-morphic image of F.

The following result (see for example [Gol92] for the mono modal case) shows that p-morphisms preserve satisfaction and validity for the language $\mathcal{L}$.

Lemma 1.14. *If p is a model p-morphism from M to M' then for all worlds w of M and formulae $\phi \in \mathcal{L}$, we have $M \models_w \phi$ if and only if $M' \models_{p(w)} \phi$. Thus ϕ is valid in M if and only if ϕ is valid in M'. If p is a frame p-morphism from F to F' then for all $\phi \in \mathcal{L}$, if $F \models \phi$ then $F' \models \phi$.*

The following are two other basic concepts that we will use.

Definition 1.15 (Reachable points). *Given a frame $F = (W, \{R_i\}_i)$ and two points $x, y \in W$, we say that y is* reachable in k steps *from x if there are $w_1, w_2, \ldots w_{k-1} \in W$ and $i_1, i_2, \ldots i_k$ in A such that $x\, R_{i_1}\, w_1\, R_{i_2}\, w_2\, \ldots\, R_{i_{k-1}}\, w_{k-1}\, R_{i_k}\, y$. We also say that y is* reachable *from x if there is some k such that y is reachable from x in k steps.*

Definition 1.16 (Connected model). *A frame is connected if for every $x, y \in W$ we have that x is reachable from y. A model is connected if its underlying frame is.*

Given a point on model, the connected sub-model that contains that point is usually called the generated model.

Property $\mathcal{P}$ of the frames	*Property name*	*Axiom A*	*Name*
$\forall w\ \exists w'\ wR_iw'$	Serial	$\Box_i\phi \Rightarrow \Diamond_i\phi$	D
$\forall w\ wR_iw$	Reflexive	$\Box_i\phi \Rightarrow \phi$	T
$\forall ww'w''\ wR_iw'$ and $w'R_iw'' \Rightarrow wR_iw''$	Transitive	$\Box_i\phi \Rightarrow \Box_i\Box_i\phi$	4
$\forall ww'\ wR_iw' \Rightarrow w'R_iw$	Symmetric	$\phi \Rightarrow \Box_i\Diamond_i\phi$	B
$\forall ww'w''\ wR_iw'$ and $wR_iw'' \Rightarrow w'R_iw''$	Euclidean	$\Diamond_i\phi \Rightarrow \Box_i\Diamond_i\phi$	5
$\forall ww'w''\ wR_iw'$ and $wR_iw'' \Rightarrow$ $\exists w'''(w'R_iw'''$ and $w''R_iw''')$	Convergent	$\Diamond_i\Box_i\phi \Rightarrow \Diamond_i\Box_i\phi$	G1
$\forall w\ wR_iw$ and $\forall w'\ (wR_iw' \Rightarrow w = w')$	Reflexive dead-end	$\phi \Leftrightarrow \Box_i\phi$	Triv

Figure 1.3: Correspondences between property of the frames and validity of axioms (Theorem 1.19).

Definition 1.17. *Given a model $M = (W, R_1, \ldots R_n, \pi)$ and a point $w \in W$, the model $M_w = (W', R'_1, \ldots, R'_n, \pi')$* generated *by w from M is defined as follows:*

- *The set W' contains w and all the points reachable from w.*
- *If $w', w'' \in W'$ and $w'R_iw''$ then $w'R'_iw''$, for any $i \in \{1, \ldots, n\}$.*
- *The interpretation is defined by: $\pi'(p) = \pi(p) \cap W'$.*

Clearly, given any equivalence model the model generated by any point of it is connected.

The following result (see, e.g. [HC84] page 80) makes precise the claim that satisfaction of a formula of $\mathcal{L}$ at a world depends only on connected worlds.

Lemma 1.18. *For all worlds w of a model M and for all formulae $\psi \in \mathcal{L}$ we have $M \models_w \phi$ if and only if $M_w \models_w \phi$.*

One of the basic insights into modal logic is provided by what is usually referred to as *correspondence theory*. This provides a one-to-one relation between the validity of a certain axiom on a class of frames and properties of the relations of these frames.

Theorem 1.19 (Correspondences). *For the properties $\mathcal{P}$ and axioms A indicated in Figure 1.3 we have that: a frame F has the property $\mathcal{P}$ if and only if $F \models A$.*

Some axioms do not correspond to any first-order property and conversely some first-order properties (irreflexivity for example) cannot be captured by any axiom. Correspondence theory studies under which circumstances syntax and semantics are equally powerful. See [Ben84] for details.

As in any logic, two major concepts relate the syntax and semantics in modal logics: *soundness* and *completeness*.

Definition 1.20 (Soundness and completeness). *Given a logic L_n and a class of frames $\mathcal{F}$, L_n is* sound *with respect to $\mathcal{F}$ if for any formula $\phi \in \mathcal{L}$ we have that $\mathrm{L}_n \vdash \phi$ implies $\mathcal{F} \models \phi$. A logic L_n is* complete *with respect to a class of frames $\mathcal{F}$ if for any formula $\phi \in \mathcal{L}$ we have that $\mathcal{F} \models \phi$ implies $\mathrm{L}_n \vdash \phi$.*

Note that completeness is defined with respect to a class of frames, not with respect a class of models. Completeness can be proven in a number of ways, the easiest of which is by using what is often called the *canonical model*.

Definition 1.21 (Canonical model). *Given a logic* L_n*, the* canonical model $M_C^{\mathrm{L}_n} = (W, \{R_i\}_i, \pi)$ *is a model built as follows.*

- *The set* W *is made of all the maximal* L_n*-consistent sets of formulae,*
- *the family of relations* $\{R_i\}_i$ *on* W^2 *is defined by*

$$wR_iw' \text{ if for all formulae } \alpha \in \mathcal{L} \text{ we have that } (\Box_i\alpha \in w \text{ implies } \alpha \in w').$$

- *The interpretation* $\pi : P \to 2^W$ *for the atoms is defined as* $w \in \pi(p)$ *if* $p \in w$*.*

When it is clear that we are referring to the canonical model of a logic L_n, we will simply denote it as M. The canonical model has the property of validating all and only the theorems of the logic.

Theorem 1.22. *For any formula* $\phi \in \mathcal{L}$, $M_C^{\mathrm{L}_n} \models \phi$ *if and only if* $\vdash_{\mathrm{L}_n} \phi$*.*

The importance of the canonical model is due to the fact that some logics are not only described by the canonical model but also by the frame of the canonical model, called the *canonical frame*. When this circumstance holds we call the logic *canonical*.

It can be proved that under certain circumstances completeness of a logic L with respect to a class of frames $\mathcal{F}$ can be proved simply by reasoning about the canonical frame.

Theorem 1.23 (Completeness via the canonical model). *Let* L_n *be a logic and let* $\mathcal{F}$ *be a class of frames. If the frame* $F_C^{\mathrm{L}_n}$ *underlying the canonical model* $M_C^{\mathrm{L}_n}$ *for* L_n *is in the class* $\mathcal{F}$ *then the logic* L_n *is complete with respect to* $\mathcal{F}$*.*

By using Theorem 1.23 and considerations on soundness, it is not hard to show the following:

Theorem 1.24 (Completeness of basic logics). *For the properties* $\mathcal{P}$ *and logics* L_n *shown in Figure 1.4 we have that: the logic* L_n *is sound and complete with respect to the class of frames that have property* $\mathcal{P}$*.*

The last meta-property we need to introduce is *decidability*.

Definition 1.25 (Decidability). *A logic* L_n *is* decidable *if there is an effective procedure that in a finite number of steps determines whether given a formula* $\phi \in \mathcal{L}$ *it is the case that* $\vdash_{\mathrm{L}_n} \phi$ *or that* $\nvdash_{\mathrm{L}_n} \phi$*.*

Decidability is related to two other properties properties as follows.

Theorem 1.26. *If a logic* L_n *is finitely axiomatisable and has the finite model property then* L_n *is decidable.*

All the logics we have seen so far and that we will work with are finitely axiomatisable (i.e. they can be axiomatised by considering a *finite* number of axiom schemas) and so, a good way for us to prove decidability will be to prove that the logic has the *finite model property*. That is defined as follows:

Property $\mathcal{P}$ of the frames	*Logic L_n*
-	K_n
Reflexive	T_n
Serial	KD_n
Transitive	4_n
Reflexive and transitive	$S4_n$
Reflexive, transitive and convergent	$S4.2_n$
Reflexive, symmetric, transitive	$S5_n$
Serial and transitive	$KD4_n$
Serial, transitive and Euclidean	$KD45_n$
Reflexive dead-end[7]	$Triv_n$

Figure 1.4: Completeness table for some basic modal systems discussed in Theorem 1.24.

Definition 1.27 (Finite model property). *A logic* L_n *is said to have the* finite model property *(or* fmp *in short) if for any formula $\phi \in \mathcal{L}$ we have that $\nvdash_{L_n} \phi$ implies that there is a finite model M for* L_n *such that $M \not\models \phi$.*

A logic can be proved to have the fmp in a number of different ways: algebraically as in [McK41], [Ber49], by the use of a "mini-canonical" model as in [HC96], etc. Here we will use the another standard technique which is better suited for this case: *filtrations* (first presented in [Lem77]). It is more convenient to describe this technique while we will be using it, i.e. in Chapter 3.

For now all we need to mention is that all the logics that we discussed so far can be proven to have the fmp.

Theorem 1.28 (Decidability of basic logics). *All the logics in Figure 1.2 have the fmp and so they are decidable.*

1.2.4.1 Two useful lemmas about $S5_n$

In this thesis we will extensively use the modal logic $S5_n$. We report here two lemmas that we will extensively use later.

Lemma 1.29. *For any $\phi, \psi \in \mathcal{L}$ we have $\vdash \phi \Rightarrow \psi$ implies $\vdash \Box_i\phi \Rightarrow \Box_i\psi$ and $\vdash \Diamond_i\phi \Rightarrow \Diamond_i\psi$.*

Proof. If $\vdash \phi \Rightarrow \psi$, by necessitating by $\Box_i$ (the logic $S5_n$ is normal; see Definition 1.8 and Figure 1.2) we have $\vdash \Box_i(\phi \Rightarrow \psi)$. By axiom K we have $\vdash \Box_i\phi \Rightarrow \Box_i\psi$.

If $\vdash \phi \Rightarrow \psi$, by contraposition and the first part of this lemma we have $\vdash \Box_i\neg\psi \Rightarrow \Box_i\neg\phi$. By contraposition again and by using the definition $\Diamond_i\chi \equiv \neg\Box_i\neg\chi$ we have $\vdash \Diamond_i\phi \Rightarrow \Diamond_i\psi$. □

Lemma 1.30. *For any $\phi \in \mathcal{L}$, we have $\vdash \Box_i\phi \Leftrightarrow \Box_i\Box_i\phi \Leftrightarrow \Diamond_i\Box_i\phi$ and $\vdash \Diamond_i\phi \Leftrightarrow \Box_i\Diamond_i\phi \Leftrightarrow \Diamond_i\Diamond_i\phi$ where $i \in A$.*

[7] This is defined as follows. Given any world w we have that wR_iw and that for any world w' wR_iw' implies $w = w'$.

Proof. We prove the first chain of bi-implications. By axiom 4 $\vdash \Box_i\phi \Rightarrow \Box_i\Box_i\phi$. By axiom T and Lemma 1.29 we have $\vdash \Box_i\Box_i\phi \Rightarrow \Box_i\phi$. For the second part of the bi-implication, assume $\Box_i\phi$, then by taking the contrapositive of axiom T:$p \Rightarrow \Diamond_i p$ and substituting $p = \Box_i p$ we have $\vdash \Box_i\phi \Rightarrow \Diamond_i\Box_i\phi$. The implication $\vdash \Diamond_i\Box_i\phi \Rightarrow \Box_i\phi$ is simply an instance of the contrapositive of axiom 5. $\Box$

1.3 Agent theories

In this section we will give a quick overview at which characteristics of agents can effectively be expressed by means of a modal language - it should be stressed that this section is by no means an exhaustive review of the work in this area as this has already been carried out elsewhere (see [WJ95]). My wish, instead, is to give the flavour of these formalisms in order to prepare the grounds for discussing the topic of this thesis in the last section.

The basic assumption of logicians working on agents, as for any other researcher working on MAS formalisms for that matter, is that it is both reasonable and useful to *ascribe* mentalistic attitudes to agents[8]. The philosophical background to such an attitude was given by the influential [Den87], in which Daniel Dennet coined the term "intentional stance" to describe the attitude of describing and predicting the behaviour of complex systems by means of intensional attitudes such as beliefs and desires.

Interestingly, Dennet was not the first to put forward such a proposal and a similar approach was also suggested nearly twenty years before by one of the fathers of AI, John McCarthy. Back in 1979, he wrote:

> To ascribe *beliefs, free will, intentions, consciousness, abilities*, or *wants* to a machine is legitimate when such an ascription expresses the same information about the machine that it expresses about a person. It is useful when the ascription helps us understand the structure of the machine, its past or future behaviour, or how to repair or improve it. It is perhaps never logically required even for humans, but expressing reasonably briefly what is actually known about the state of the machine in a particular situation may require mental qualities or qualities isomorphic to them. Theories of belief, knowledge, and wanting can be constructed for machines in a simpler setting than for humans, and later applied to humans. Ascription of mental qualities is most straightforward for machines of known structure such as thermostats and computer systems, but it is most useful when applied to entities whose structure is incompletely known.
> ([McC79], page 1. Cited in [Woo92, Sho93]).

The above is exactly the approach taken in this thesis, in line with much of the literature on MAS. We look at agents as systems that can be described by assuming an "intentional stance" and we aim to model key characteristics of them. MAS formal theories are then to be regarded as *specifications*.

If specifications are to be useful, they need to be *verifiable*. Loosely speaking verification means "testing that the specification meets the desired requirements". In this thesis we take the view that formal logic provides a good tool for specifying and reasoning about MAS; why is it so? The answer comes by considering three key issues: syntax, semantics and

[8]We will randomly use female and male gender for the term agent.

proof-theory. The syntax of a specification is just the language that we use to describe it and as such it cannot provide many insights into the system. In most of the cases, this will be the multi-modal language that we described in Section 1.2.1 (we will make clear when we are using extensions of it).

The semantics of a MAS theory will be based on Kripke style possible worlds or interpreted systems (see Section 1.2.2). Semantics provides a way of describing the application which is "relatively" grounded on it. Suppose we have a class of MAS that can be fully described by a class of frames $\mathcal{F}$ and we want to test whether a particular MAS of this class, described by a finite conjunction of formulas $\Gamma = \gamma_1 \wedge \cdots \wedge \gamma_n$, satisfies a property that can be written as ϕ. This can be done by checking the entailment $\Gamma \models_{\mathcal{F}} \phi$, i.e. whether for any model M built on a frame in $\mathcal{F}$ and for any world w we have that $M \models_w \Gamma \Rightarrow \phi$.

Proof theory (which we will use in the shape of Section 1.2.3) can also be used to check the properties of a MAS, but it takes a different perspective. The class of MAS under analysis is specified by a set of axiom schemas and inference rules that describe a logic L_n; a particular MAS in this class is again identified by a conjunction of formulas Γ. In this case, verifying that the specification of the particular MAS meets a requirement ϕ amounts to proving in the logic L_n that ϕ follows from Γ, i.e. that $\Gamma \vdash_{\mathrm{L}_n} \phi$, or in other words that $\vdash_{\mathrm{L}_n} \Gamma \Rightarrow \phi$.

Throughout this thesis we will be often be proving soundness and completeness of a logic with respect to a certain semantics. This guarantees that these two methods for reasoning about a MAS are equally powerful, i.e. that $\vdash_{\mathrm{L}_n} \Gamma \Rightarrow \phi$ if and only if $\models_{\mathcal{F}} \Gamma \Rightarrow \phi$.

Although the two methods are logically equivalent we need to be careful here as there are two schools of thought on how one should in principle verify a MAS.

There are those who think that given a MAS one could in principle be able to have a model of the system (for example as a class of frames of modal logic). According to these completeness is about being able to reason about a MAS by using a Hilbert style proof system that is equivalent to the semantical description. This unlocks the possibility of using automatic theorem proving techniques (Isabelle [Pau94] is one of them) to test the specification.

A radically different way of seeing the process is to start from the axiomatisation. One could choose the axiom schemas that suit best his or her needs to describe the system and then test whether a formula (representing a property) is a theorem of the logic (representing the system) by testing validity of the formula on the corresponding semantical class, for example by using semantical refutation techniques like tableaux ([Fit83]).

People feel (and argue) quite passionately about these issues and this thesis is not about solving this long-going dispute (but we come back to this point in Section 1.4.2). More neutrally I would take the view that completeness gives us a collection of new methods for proving properties of the MAS *whichever* side we come from. Another advantage is that thanks to completeness we have two expressive tools to prove properties and sometimes simply by switching to the more appropriate one, proof-theoretical or semantical, hard problems become easier to solve. For example, it is thanks to completeness that comparing two or more logics and deciding whether they are different or equivalent is now an accessible problem (note that this used to be a very hard problem before works such as [Lem77] became known) and it is very relevant for MAS theories.

We should see the process of generating a good specification for a MAS as an *incremental* process. One way is to start from the semantics, find a complete axiomatisation for it, and then carry out experiments on the axiomatisation by using automatic theorem proving on it. This will probably reveal that some theorems that we did not want to be true are actually true and some other formulae that we assumed would hold actually do not. By completeness the semantics we started from is either not the intended one or, if matching the application, it reveals some unwanted properties about the application. We can then go back to the semantics and change it accordingly. But the opposite path[9], is in principle equally possible and it is up to us to decide what to start with.

The argument above is to show that if we want agent theories to be useful for the specification of MAS, we need to provide *sound and complete axiomatisations* for them. This is the reason why the theme of completeness will play such a crucial role in this work.

The rest of this section is organised as follows. First I will present logics modelling a single aspect of agency. After this we will move to the multi-agent case where we will discuss some general issues and analyse a few important multi-agent theories.

1.3.1 Single-agent theories

In this subsection we use modal operators to model intentional aspects of agency. The whole sub-section is devoted to the specification of a single agent. Syntax, semantics and proof theory are the ones described in Section 1.2 but limited to *one* modal box.

1.3.1.1 Epistemic logics

The design of an agent of knowledge is a central issue in agents theory, as knowledge is a key property of any intelligent system (see for example [RN95], Section 1).

The most widely used system to model knowledge was proposed by Hintikka in his famous book [Hin62] and since then used by many others.

The syntax is the one presented in Section 1.2, in which the modal box $\Box$ is here written as K. In this setting a formula $K\phi$ reads as "the agent *knows* ϕ". The operator L the dual of K, is defined as in corresponding to the modal diamond $\Diamond$ (page 12). A formula $L\phi$ is read as "the agent considers ϕ to be possible according to his knowledge".

The interpretation is given on Kripke models (Definition 1.1) following Definition 1.2. In this case the fact that two states w_0, w_1 are related by the accessibility relation intuitively means "the states w_0, w_1 are epistemically possible according to the agent's knowledge". The favoured approach uses an equivalence relation for the accessibility relation. Let us see why.

First the relation should be reflexive. In fact, if not an agent in the state w_0, the actual state, could regard w_0 as being impossible. So the agent would be able to have *false* knowledge. This is contrary to our intuition of knowledge as "true belief".

Let us now see why the relation should be symmetric. Suppose an agent at a state w_0 considers w_1 as possible but in w_1 she considers w_0 not to be possible. Then she would know in w_0 that w_0 cannot be a possible state, but we have just noticed in the previous paragraph that this would be absurd. So w_0 has to be possible from w_1 whenever w_1 is possible from w_0.

Finally transitivity. Suppose an agent in w_0 considers w_1 to be epistemically possible and that in w_1 she considers w_2 as epistemically possible. Then if she did not regard w_2 as possible when in w_0, it would mean that in w_0, she would know that w_2 cannot be the case. But then, she would also know this at w_1 (because it is considered possible from w_0) and so w_2 would not be regarded as possible from w_1.

[9] That is, from the axiomatisation find a complete semantics, check this one for satisfaction of desired properties with semantical methods and if necessary change the axiomatisation accordingly.

There is another way of showing that an equivalence relation is adequate for the case of knowledge. Recall Definition 1.4 of static interpreted systems and how these generate Kripke models (Definition 1.6). Consider two global states[10]: $s = (l_e, l_1)$ and $s' = (l'_e, l'_1)$; where l_e, l'_e are local states for the environment and l_1, l'_1 are states for agent 1. Suppose the agent is in the state s (which implies that its local state is l_1), when can we say that s' is epistemically possible for the agent in consideration? Well, we can say that the agent will regard s' to be possible when its local states in the two global states are the same. In fact if $l_1 = l'_1$ then up to the agent's knowledge there is no difference between the two states s and s'. So, for the case of knowledge we can fix the predicate $\{P_i\}_{i \in A}$ of Definition 1.6 to be exactly the equality of local states, i.e. $(l_e, l_1)R_1(l'_e, l'_1)$ if $l_1 = l'_1$. So, any generated frame will be an equivalence frame. This is the approach taken in [FHMV95]; see this reference for more details.

So, whether we favour Kripke models or interpreted systems as semantic foundations for logic for MAS, it is reasonable to consider the accessibility relation on the frames to be an equivalence relation. As we observed in Theorem 1.24, the logic which is complete with respect to equivalence frames is S5 = T + 5. Note that that axiom 4 can be proven in S5. So, it is convenient to think of S5 as defined as follows:

Taut	$\vdash_{S5} t$, where t is any propositional tautology
K	$\vdash_{S5} K(p \Rightarrow q) \Rightarrow (Kp \Rightarrow Kq)$
T	$\vdash_{S5} Kp \Rightarrow p$
4	$\vdash_{S5} Kp \Rightarrow KKp$
5	$\vdash_{S5} Lp \Rightarrow KLp$
US	If $\vdash_{S5} \phi$, then $\vdash_{S5} \phi[\psi_1/p_1, \ldots, \psi_n/p_n]$
MP	If $\vdash_{S5} \phi$ and $\vdash_{S5} \phi \Rightarrow \psi$, then $\vdash_{S5} \psi$
Nec	If $\vdash_{S5} \phi$, then $\vdash_{S5} K\phi$

In the context of knowledge, the axiom K states the closure of the agent's knowledge under implication. Axiom T asserts the truth of anything that is known by the agent. Axiom 4 expresses positive introspection (in the sense of [Kon86]) in the agent knowledge, meaning that if the agent knows a fact then he or she knows to know it. Axiom 5 represents a negative introspection property, because it states that the agent knows that he or she does not know something whenever he or she does not know it. The inference rule of necessitation asserts that the agent knows any valid formulae, in particular all the propositional tautologies.

The properties described above are very strong. It would not be reasonable for us to think that any real agent can master these abilities while operating in any scenario and we should regard agents modelled by S5 (or S5-agents) as *ideal agents*. This is so, because they behave like *perfect reasoners*. Indeed we should interpret any formula $K\phi$ as "ϕ follows from the information that the agent holds", thus an agent would have explicit knowledge of ϕ if he or she had infinite computational resources. In this idealisation an S5-agent is logically omniscient and with complete introspection properties. For a review of philosophical arguments for and against S5 as a logic for knowledge see [Len78].

The logic S5 with its extensions is extensively studied because it provides an important theoretical base for the study of knowledge. In order to model more realistic scenarios the system S5 has been modified to incorporate more practical needs.

[10]Here we are dealing with the case $n = 1$.

If it is positive (respectively negative) introspection bothering us for its philosophical implications (how can an agent be aware of *anything* he or she knows and does not know?!), then S5 can simply be replaced by the weaker KT4 (respectively KT5). Or we could simply consider KT should we want to drop both. We know from Theorem 1.24 that these logics are still complete with respect to an appropriate class of frames.

If we are uneasy with what is usually called "the logical omniscience problem" (i.e. the agents knowing all valid formulae together with his or her knowledge being closed under implication), we would like to drop axiom K and the inference rule of necessitation altogether. This is not possible in any normal modal system (see Definition 1.8) and we would need to modify the semantics heavily. A number of frameworks have been proposed, the most known of which are perhaps [Thi92] and [FHV95]. Neither of these last two systems reject modal logic on its own: the former uses a partial semantics for interpreting the formulae, the latter has a non-standard negation built in the semantics in order to block some of the critical schemas of inference, that provoke omniscience.

S5 is a successful logic for modelling idealised knowledge also because, apart from being complete, is also decidable (see Theorem 1.28).

Some proposals [Gar84, LM94, LC96, Lom95] have also been put forward on how to extend systems similar to S5 to the first order case in order to provide more expressive specifications. A major expressive capability in first order epistemic logic is the ability of representing statements like "knowing who" (rather than the usual "knowing that" of the propositional case described here), but this has to be paid in terms of the heavy formal machinery needed and in the difficulty of proving completeness results.

Of course modal logic is by no means the only logic technique to model knowledge and alternative proposals have been put forward (for example [Lev84, Kon86, Tur90] or even first order [McC78]). Since in this thesis we only deal with modal logics we will not discuss these here.

1.3.1.2 Doxastic logics

The representation of beliefs of an agent is equally (some would argue more) important as the representation of the agent's knowledge. The commonly accepted difference between the two concepts is that while the beliefs of an agent may be false, this is not the case for her knowledge. This is why the logic commonly used for this task is the logic KD45 which is basically the logic S5 in which the axiom T: $\Box\phi \Rightarrow \phi$ is replaced by the weaker D: $\Box\phi \Rightarrow \neg\Box\neg\phi$, that simply guarantees the coherence of the agent's beliefs.

As we showed in Figure 1.2, the logic KD45 can be axiomatised as below in which modal formulae $B\phi$ in this case are meant to be read as "the agent believes that ϕ":

Taut	$\vdash_{KD45}$ t where t is any propositional tautology
K	$\vdash_{KD45} B(p \Rightarrow q) \Rightarrow (Bp \Rightarrow Bq)$
D	$\vdash_{KD45} Bp \Rightarrow \neg B\neg p$
4	$\vdash_{KD45} Bp \Rightarrow BBp$
5	$\vdash_{KD45} \neg Bp \Rightarrow B\neg Bp$
US	If $\vdash_{KD45} \phi$, then $\vdash_{KD45} \phi[\psi_1/p_1, \ldots, \psi_n/p_n]$
MP	If $\vdash_{KD45} \phi$ and $\vdash_{KD45} \phi \Rightarrow \psi$, then $\vdash_{KD45} \psi$
Nec	If $\vdash_{KD45} \phi$, then $\vdash_{KD45} B\phi$

The semantics is given through Kripke structures and, for reasons similar to the one presented on page 23 for the epistemic case, the binary relation (which intuitively means "being possible according to the agent's beliefs)" is in this case chosen to be serial, transitive and Euclidean. Indeed the logic KD45 is complete with respect to serial, Euclidean and transitive frames, and decidable (see Theorem 1.24 and Theorem 1.28). The methods and the aims of extending KD45 to first order logic are equivalent to the ones for S5; see the references above.

The logic KD45, just as S5, suffers from the logical omniscience problem and in order to overcome this difficulty some systems have been presented. The most famous is probably the one presented in [Lev84] in which Levesque distinguishes between *implicit* and *explicit* beliefs. Implicit beliefs, i.e. beliefs that an agent may not be aware of, still enjoy logical omniscience (indeed they are modelled just like KD45) but explicit believes are based on a model similar to the one in situation semantics [BP84]. Interestingly it has been noted in [MH95] that even Levesque's explicit beliefs suffer from logical omniscience when seen from a relevance logic perspective [Dun86]. Other treatments that depart further more from the traditional approach presented here are [FH88], [Jas91], [Jas93], and [HM89]. Discussing these is beyond the scopes of this introduction.

1.3.1.3 Logics of intention

Philosophers have long been concerned with the mental state of intention. It has been considered for long time that the intention of an agent could be expressed as a combination of desires and believes, but after the influential work of Bratman [Bra87] and [Bra90] intentions are generally considered as an irreducible mental state.

Intentions involve actions, hence change, differently from belief and knowledge that concern static states. For this reason some authors ([Cas75] and others) have suggested that intentions should be specified with logics for actions, e.g. dynamic logic. Others ([CL90] and [Bra83]) have claimed that this would make it impossible to integrate different mental states in a single framework.

For the scopes of this section here we report a simple modal treatment of intentions that constitutes a fragment of the BDI system presented in [Rao96] to define a comprehensive formalism for MAS. The language is the usual modal propositional logic, where a modal formula $I\phi$ reads "the agent intends that ϕ"; on the semantical side a world is related to another one, if the latter is a possible way of achieving the agent's intention at the former world.

The system used in that work, the modal logic KD, is axiomatised as follows:

Taut	$\vdash_{KD} t$, where t is any propositional tautology
K	$\vdash_{KD} I(p \Rightarrow q) \Rightarrow (Ip \Rightarrow Iq)$
D	$\vdash_{KD} Ip \Rightarrow \neg I \neg p$
US	If $\vdash_{KD} \phi$, then $\vdash_{KD} \phi[\psi_1/p_1, \ldots, \psi_n/p_n]$
MP	If $\vdash_{KD} \phi$ and $\vdash_{KD} \phi \Rightarrow \psi$, then $\vdash_{KD} \psi$
Nec	If $\vdash_{KD} \phi$, then $\vdash_{KD} I\phi$

KD models an ideal agent of intention. This is because the modal schema D guarantees the consistency of the agent's intentions. It is therefore not admissible for an agent to intend ϕ and $\neg\phi$ at the same time. The logic KD is sound and complete with respect to serial frames (see Theorem 1.24) and enjoys decidability (see Theorem 1.28).

The modal logic KD has also been used in [Col96] to model intentional comunication.

1.3.1.4 Time

Although time is not a mental state, it is an essential component of any model aiming at representing the temporal evolution of agent's and environment's change.

The formalisation of temporal evolution has been for a long time topic of research among logicians and the literature of the field is cumbersome (excellent references are [MP92, CE81, GHR93, Ben83, Eme90]). For the purpose of this document we only point out the most common usages of modal logic for representing temporal evolution in agent theories.

First we report the mainstream approach (the one presented in [GHR93] for example). This is given by a syntax defined on a propositional language enriched by two modal operators, $\Box_+, \Box_-$, which are read as "always in the future" and "always in the past" respectively. The boxes $\Box_+$ and $\Box_-$ generate through the usual relation their duals: $\Diamond_+\phi = \neg\Box_+\neg\phi$, and $\Diamond_-\phi = \neg\Box_-\neg\phi$. Formulae $\Diamond_+\phi$ are read as "sometimes in the future", while $\Diamond_-\phi$ are read as "sometimes in the past".

The semantics is given through Kripke frames $F = (W, R_f, R_p)$, in which the worlds represent state of affairs at particular time instants and the two binary relations R_f, R_p are bound by the property

$$\text{For any } w, w' \in W \; wR_f w' \text{ if and only if } w' R_p w,$$

that captures the intuition of the the two (forward and backwards) flows (sequences of states related by one of the relations) of time being one the converse of the other. Further properties can be imposed on these relations. Often it is meaningful to consider R_p, R_f, as partial orderings, but many more issues arise like whether the relations have to be linear or branching, strict, total, etc. Indeed much of the literature of modal logic arose in the effort of having a good formal model of temporal evolution. All these issues are very important but not so relevant in the context of this thesis and so we simply present the weakest axiomatisation for linear time (from [GHR93], page 92):

Taut	Any propositional tautology
$K_{\Box_+}$	$\Box_+(p \Rightarrow q) \Rightarrow (\Box_+ p \Rightarrow \Box_+ q)$
$K_{\Box_-}$	$\Box_-(p \Rightarrow q) \Rightarrow (\Box_- p \Rightarrow \Box_- q)$
$R_{\Box_-}$	$p \Rightarrow \Box_-\Diamond_+ p$
$R_{\Box_+}$	$p \Rightarrow \Box_+\Diamond_- p$
$4_{\Box_-}$	$\Diamond_- p \Rightarrow \Diamond_-\Diamond_- p$
$4_{\Box_+}$	$\Diamond_+ p \Rightarrow \Diamond_+\Diamond_+ p$
US	If $\vdash \phi$, then $\vdash \phi[\psi_1/p_1, \ldots, \psi_n/p_n]$
MP	If $\vdash \phi$ and $\vdash \phi \Rightarrow \psi$, then $\vdash \psi$
Nec	If $\vdash \phi$, then $\vdash \Box_-\phi$ and $\vdash \Box_+\phi$

It can be observed that axiom 4 is present for both modal operators representing the transitivity of temporal evolution.

We now turn our attention to a slightly different model of temporal evolution, often used to model MAS and due to Fagin, Halpern, Vardi and Moses; we present it as it appears in [FHMV95] and [HF89]. Recall Definition 1.4 where we reported the definition of a set of global states for interpreted systems. We have already seen that interpreted systems can be used to model knowledge (see page 24), we now show how they can also be used to model time. To describe it we use the full notion of interpreted sytem without limitating it to the static case as we did in Definition 1.5.

Given a set of global states S, consider a set of *runs* $R = \{r : \mathbb{N} \rightarrow S\}$, where $\mathbb{N}$ is the set of natural numbers. Thus a run is simply a sequence of globals states $r = (s_0, \ldots s_n, \ldots)$. If $r(m) = (l_e, l_1, \ldots, l_n)$ is the global state identified by run r at time m, then $r_e(m) = l_e$ and $r_i(m) = l_i$, for $i = 1, \ldots, n$. An interpreted system suitable to model time can then be defined as $IS = (\mathcal{R}, \pi)$, where π is an interpretation function suitable to interpret a set of propositional variables, and $\mathcal{R}$ is a set of runs R as defined above.

The syntax commonly used in this context is slightly different from the one presented above and it is defined from the usual set of propositional variables enriched by two modal operators: $\bigcirc, \mathcal{U}$, the first being unary and the second binary. We do not present here operators for the past. A formula $\bigcirc\phi$ should be read as "at the next step ϕ", while a formula $\phi \mathcal{U} \psi$ means "ϕ until ψ". The interpretation for these connectives is standard and the operators of $\Box$ ("always true"), and $\Diamond$ ("sometime true") can be defined in terms of $\mathcal{U}$, given the equivalence $\Diamond\phi = true \mathcal{U} \phi$.

Satisfaction for this language is defined as follows:

$IS \models_{r,m} p$	if	$w \in \pi(p)$
$IS \models_{r,m} \neg\phi$	if	$IS \not\models_{r,m} \phi$
$IS \models_{r,m} \phi \wedge \psi$	if	$IS \models_{r,m} \phi$ and $IS \models_{r,m} \psi$
$IS \models_{r,m} \bigcirc\phi$	if	$IS \models_{r,(m+1)} \phi$
$IS \models_{r,m} \phi \mathcal{U} \psi$	if	$IS \models_{r,m'} \psi$ for some $m' \geq m$ and
		$IS \models_{r,m''} \phi$ for all m'' such that $m \prec m'' \prec m'$

Satisfaction for the other logical connectives can be defined in the usual way.

This formal machinery does not deal with continuous time, nor with the indeterminism of branching time, but it can be extended appropriately to deal with them (see for example [EH85]).

An interpreted system suitable to model temporal evolution can be axiomatised by the following:

$\vdash t$ Where t is any propositional tautology
$\vdash \bigcirc p \wedge \bigcirc(p \Rightarrow q) \Rightarrow \bigcirc q$
$\vdash \bigcirc(\neg p) \Rightarrow \neg \bigcirc p$
$\vdash p \mathcal{U} q \Leftrightarrow q \vee (p \wedge \bigcirc(p \mathcal{U} q))$
If $\vdash \phi$, then $\vdash \phi[\psi_1/p_1, \ldots, \psi_n/p_n]$
If $\vdash \phi$, then $\vdash \bigcirc\phi$
If $\phi' \Rightarrow \neg\psi \wedge \bigcirc\phi'$, then $\phi' \Rightarrow \neg(\phi \mathcal{U} \psi)$

The above axiomatisation is known to be sound and complete for linear flows of time [GPSS80].

We have now terminated with our description of modal theories that deal with one aspect of agency of an agent. We have seen that a modal language is well-suited to represent the idealisation of the concepts involved. We leave the single-agent case and we turn our attention to the multi-agent case.

1.3.2 Multi-agent theories

So far we have discussed theories addressing one of the mental states that we can ascribe (in the sense of [McC79] and [Den87]) to a system composed by one agent and the environment. Distributed Artificial Intelligence, though, needs theories that take into account *several* agents operating in their environment (sometimes in cooperation, sometimes in conflict, sometimes independently one for another). This is why the theories of mental states for an agent of Section 1.3.1 have been extended to the multi-agent case.

In order to discuss properly the multi-agent case, three issues need to be considered: multiplicity of the agents, group properties and interaction properties.

1.3.2.1 Multiplicity of the agents

In order to deal with the multiplicity of the agents, a MAS theory has to be able to represent not just one mental state (as we saw in Section 1.3.1) but *several* mental states of the same type, one for every agent in the set A. Technically, this can be achieved by considering the multi-modal extensions of the modal logics in use for the mental state in consideration.

So, in order to represent a group of agents, we need to use the syntax, semantics and proof theory as they were defined in a multi-modal context over a set A of agents in Section 1.2. As it was reported there, the properties of the logics that are relevant here (mainly completeness and decidability) are not affected by moving to the multi-agent case.

For the case of knowledge, the multi-modal extension of [Hin62] was first proposed by Halpern and Moses in [HM90]; a good analysis of the logic appears in [HM92a]. Similarly, all the logics that we reported in the previous section can be extended to the multi-agent case. For example the logic KD45$_n$ has been extensively used for the representation of beliefs.

1.3.2.2 Group properties

Group properties arise naturally when we investigate a group of agents as a whole instead of as a collection of individuals. Let us consider the interesting case of knowledge; we will follow the approach that appears in [FHMV95].

We have seen how to reason about the knowledge of any agent in a set $A = \{1, \ldots, n\}$: we can use a language with n modal boxes $K_i, i \in A$ and we now know why the system S5$_n$ is a suitable candidate. As stated in Theorem 1.24, the logic S5$_n$ is complete with respect to equivalence frames.

We now extend this machinery by considering particular types of knowledge of the group. In doing so we will extend the syntax of Section 1.2.1 with new modal operators and we will give their semantics. Three notions of group knowledge are particularly important: "everybody knows", distributed knowledge and common knowledge. All of these can be expressed in modal logic by extending the logic S5$_n$.

"Everybody in a group G knows ϕ" is a state of knowledge of a group $G \subseteq A$ of agents that arises when all the agents have the knowledge about some fact ϕ. Formally, we can extend our syntax by introducing a modal box E_G, whose interpretation is defined in terms of the $K_i, i \in G$:

$$M \models_w E_G\phi \text{ if } M \models_w K_i\phi \text{ for any } i \in G.$$

Not surprisingly, the logic S5$_n^E$ =S5$_n$ + $\{E_G\phi \Leftrightarrow \bigwedge_{i\in G} K_i\phi\}$ is sound and complete with respect to equivalence frames.

We now turn our attention to common knowledge. Common knowledge was first proposed in [Lew69] and later formalised in [MSHI78] and [Mil81] and [Leh84]. A fact ϕ is common knowledge among a group of agents $G \subseteq A$, formally $C_G\phi$, when every agent knows ϕ and every agent knows that every agent knows ϕ and so on. In fact it can be argued that any number of finite conjunctions of everybody knows is not enough to capture the intuition of certain circumstances. For example suppose that in a group of agents an announcement, say the formula ϕ, is made. Under the assumption of all the agents being perfect reasoners able to receive the message one can easily persuade himself that the knowledge of the group after that the announcement is made becomes stronger than any chain of "everyone knows". In fact, it would not only be that everyone knows ϕ but also that everyone knows that everyone knows ϕ and that everyone knows that everyone knows that everyone knows ϕ and so on. Common knowledge is particularly useful when modelling situations in which the agents share a *strong core* of knowledge. Typically, protocols in a distributed system are commonly known.

The syntax of the system S5$_n^E$ defined above can be extended to accommodate common knowledge. Indeed, common knowledge of a fact ϕ among a group $G \subseteq A$ of agents can be expressed as a modal operator $C_G\phi$, whose interpretation can be defined as follows:

$$M \models_w C_G\phi \text{ if } M \models_w E_G^k\phi \text{ for any } k \in \{1, \dots\},$$

where $E^k\phi$ is a shorthand for $E_G \dots E_G\phi$, k times.

The obtained logic that we call S5$_n^{E,C}$ can be presented as follows:

Taut	$\vdash_{S5_n^{E,C}} t$, where t is any propositional tautology
K	$\vdash_{S5_n^{E,C}} K(p \Rightarrow q) \Rightarrow (Kp \Rightarrow Kq)$
T	$\vdash_{S5_n^{E,C}} Kp \Rightarrow p$
4	$\vdash_{S5_n^{E,C}} Kp \Rightarrow KKp$
5	$\vdash_{S5_n^{E,C}} Lp \Rightarrow KLp$
E	$\vdash_{S5_n^{E,C}} E_G\phi \Rightarrow \bigwedge_{i\in G} K_i\phi$
C	$\vdash_{S5_n^{E,C}} C_G\phi \Rightarrow E_G(\phi \wedge C_G\phi)$
US	If $\vdash_{S5_n^{E,C}} \phi$, then $\vdash_{S5_n^{E,C}} \phi[\psi_1/p_1, \dots, \psi_n/p_n]$
MP	If $\vdash_{S5_n^{E,C}} \phi$ and $\vdash_{S5_n^{E,C}} \phi \Rightarrow \psi$, then $\vdash_{S5_n^{E,C}} \psi$
Nec	If $\vdash_{S5_n^{E,C}} \phi$, then $\vdash_{S5_n^{E,C}} K\phi$
Ind	If $\vdash_{S5_n^{E,C}} \phi \Rightarrow E_G(\phi \wedge \psi)$, then $\vdash_{S5_n^{E,C}} \phi \Rightarrow C_G\psi$

Having done so, it is possible to prove completeness of the logic S5$_n^{E,C}$ with respect to equivalence frames. This was originally done in [KP81].

We now discuss an extension of S5$_n^{E,C}$ that includes an operator for distributed knowledge. Distributed knowledge of a fact ϕ among a group G of agents arises when the agents of the group can obtain knowledge of ϕ (that is not necessarily known by any agent) by confronting their mental states and eliminating epistemic alternatives that are not deemed possible by some agent. One can see distributed knowledge as the knowledge of a wise man, external to the group G, that knows what the agents consider possible and, therefore, can infer some knowledge about the community as a whole.

Distributed knowledge, sometimes also called "implicit knowledge" (as for example in [HM92b]) cannot be defined in terms of "everybody knows" or common knowledge. Still, we can express it by using another modal operator D_G whose interpretation we can define in terms of the accessibility relations of the agents in G. In particular, we say that a fact ϕ is distributely known among a group G of agents at world w if ϕ holds at all the worlds that are related to w via *every* accessibility relation of the agents in G, i.e.

$$M \models_w D_G\phi \text{ if for any } w' \in W \text{ we have } w(\bigcap_{i \in G} \sim_i)w' \text{ implies } M \models_{w'} \phi.$$

The reason for considering the intersection of the accessibility relations lies on the observation that if some agent knows that an epistemic alternative is not possible this knowledge will permit the others to rule out this epistemic alternative. Hence, in a distributed setting they only need considering the alternatives that all of them consider possible.

Given the above definition, the two following axioms will hold for an axiomatisation of distributed knowledge:

$$D_{\{i\}}\phi \Rightarrow K_i\phi, \text{ for any } i \in \{1, \ldots, n\}, \tag{1.1}$$

$$D_G\phi \Rightarrow D_{G'}\phi, \text{ for any } G \subseteq G'. \tag{1.2}$$

Proving completeness for a language that includes distributed knowledge was independently achieved in [HM92b] and [FHV92]. The result is the modal logic $S5_n^D$ that can be axiomatised by taking the logic $S5_n$ and adding the S5 axiomatisation for the D operator plus the two axioms 1.1, 1.2 above.

Completeness for a language including all the operators of group knowledge is presented in [HM97]. The logic $S5_n^{E,C,D}$, obtained by taking the union of the axiomatisations of $S5_n^D$ and $S5_n^{E,C}$, is still complete with respect to equivalence frames.

Other group properties of knowledge that we do not report here are also studied; for example, see [HLM96] for a different notion of distributed knowledge. The complete axiomatisation for common knowledge described above can be relaxed from S5-operators to KD45-operators obtaining a complete axiomatisation for mutual belief (see for example [Col93]).

Considering a group of agents, there are other group properties that we may consider, like joint intentions. The treatments in these cases are usually similar to the case described above and we do not present them here.

1.3.2.3 Interactions between different mental states of the same agent

In Section 1.3.1 we have discussed that many mental states are needed to describe an intelligent system. We have observed that there are subtle differences between the concepts of knowledge and belief, between desire and intention, etc. If we are to specify a fully-fledged intelligent system, in general we do not want all these modules to be independent from each other. Rather, we would like to be able to specify some sort of dependency among these characteristics. For example, it would not be reasonable to let an agent a believe facts that a already knows to be false. Equally, it would seem unreasonable to let agent a intend to bring about something which is not believed to be possible by a.

In order to specify interesting MAS theories it is not enough to consider many single aspect theories and "paste" them together as independent modules, but we need express

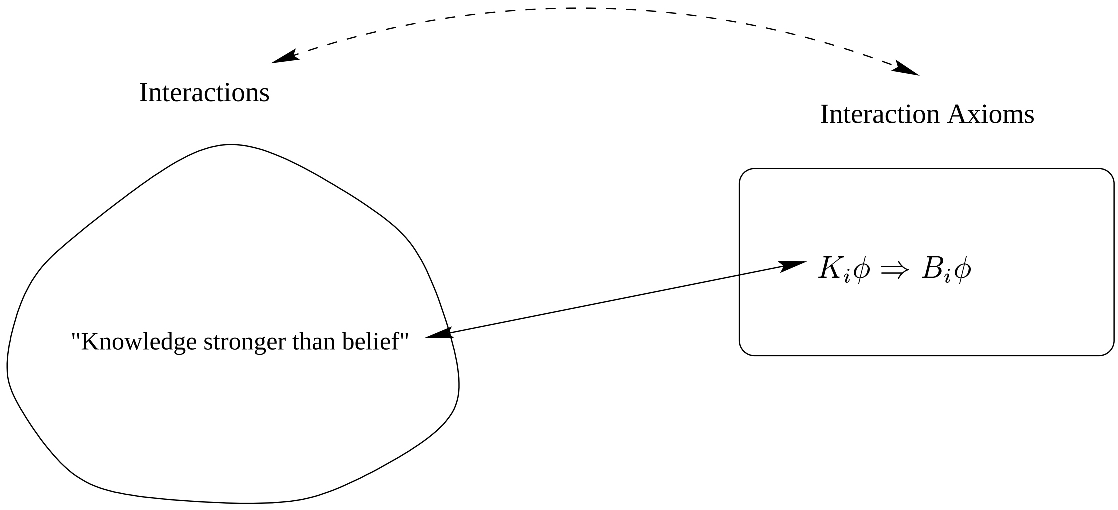

Figure 1.5: A representation of interactions and interaction axioms. The solid line connects a relation between two mental attitudes to its formalisation in the language of modal logic. The dashed line shows the general relation that exists between interactions between mental states and interaction axioms. Interaction axioms are the formal tools we will employ to represent interactions between mental states.

appropriate *interactions* between the different mentalistic components. By interaction we informally mean a form of *binding* between the mental states of the agents, a relation between the mentalistic components. For example, an interaction between knowledge and belief (as we shall see briefly) may be that "knowledge is stronger than belief", meaning that everything which is known is also believed.

Modal logic, once again, proves to be powerful enough to capture formally our needs. If we imagine using more than one modal operator for every agent to express the mental states in interest we can represent the interactions between different mental states as *interaction axioms*, i.e. formulae containing different modal operators. In the example above, the interaction between knowledge and belief can be expressed by the implication $K_i\phi \Rightarrow B_i\phi$, where K_i and B_i are the operators for knowledge and belief respectively referring to agent i. See Figure 1.5.

In the next three examples we explore a few interesting cases of interaction from the literature.

1.3.2.3.1 Case 1: Knowledge and belief. Kraus and Lehmann presented in [KL88] a combined system for reasoning about both knowledge and beliefs of a set $A = \{1, \ldots, n\}$ of agents (this system was then revisited by van der Hoek and Meyer in [Hoe93] where more interactions are studied). They propose a multi-modal logic for knowledge and belief with group properties, and a very reasonable collection of interaction axioms. The language is multi-modal as in Section 1.2 but defined on *two* families of modal operators, K_i and B_i (i is as always in a set A of agents), enriched by the modal operators of "everybody knows" (E_K), "everybody believes" (E_B), common knowledge ($\mathcal{C}_K$) and common belief ($\mathcal{C}_B$). Common knowledge is defined as Section 1.3.2.2[11]; the operator of common belief is defined as common knowledge but interpreted on the accessibility relation for belief.

More precisely, the semantics is based on possible worlds as in Section 1.2 but with two families of relations. Formally, frames for this system are tuples of the form $F = (W, \{k_i\}_{i\in A}, \{b_i\}_{i\in A})$, where $\{k_i\}_{i\in A}$ is a family of equivalence relations used to interpret the knowledge operators $K_i, E_K, \mathcal{C}_K$. The family $\{b_i\}_{i\in A}$ is composed by Euclidean, serial and transitive binary relations used to interpret the belief operators $B_i, E_B, \mathcal{C}_B$.

The two families of relations satisfy the following two extra properties:

1. for any i in A, $b_i \subseteq k_i$.
2. for any i in A, $k_i \circ b_i \subseteq b_i$.

The relation 1 represents the fact that knowledge should be stronger than belief, while relation 2 represents the introspection of knowledge over belief.

Kraus and Lehmann prove that these constrains on the class of frames correspond to the following sound and complete axiomatisation:

$\vdash t$, where t is a propositional tautology
$\vdash K_i(\phi \Rightarrow \psi) \Rightarrow (K_i\phi \Rightarrow K_i\psi)$
$\vdash K_i\phi \Rightarrow \phi$
$\vdash \neg K_i\phi \Rightarrow K_i\neg K_i\phi$
$\vdash \mathcal{C}_K(\phi \Rightarrow \psi) \Rightarrow (\mathcal{C}_K\phi \Rightarrow \mathcal{C}_K\psi)$
$\vdash \mathcal{C}_K\phi \Rightarrow K_i\phi$
$\vdash \mathcal{C}_K\phi \Rightarrow K_i\mathcal{C}_K\phi$
$\vdash \mathcal{C}_K(\phi \Rightarrow E_K\phi) \Rightarrow (\phi \Rightarrow \mathcal{C}_K\phi)$
$\vdash B_i(\phi \Rightarrow \psi) \Rightarrow B_i\phi \Rightarrow B_i\psi$
$\vdash \neg B_i \perp$
$\vdash \mathcal{C}_B(\phi \Rightarrow \psi) \Rightarrow (\mathcal{C}_B\phi \Rightarrow \mathcal{C}_B\psi)$
$\vdash \mathcal{C}_B\phi \Rightarrow E_B\phi$
$\vdash \mathcal{C}_B\phi \Rightarrow E_B\mathcal{C}_B\phi$
$\vdash \mathcal{C}_B(\phi \Rightarrow E_B\phi) \Rightarrow (E_B\phi \Rightarrow \mathcal{C}_B\phi)$
$\vdash K_i\phi \Rightarrow B_i\phi$
$\vdash B_i\phi \Rightarrow K_iB_i\phi$
$\vdash \mathcal{C}_K\phi \Rightarrow \mathcal{C}_B\phi$
If $\vdash \phi$ and $\vdash \phi \Rightarrow \psi$, then $\vdash \psi$
If $\vdash \phi$, then $\vdash \mathcal{C}_K\phi$

The system is complex and can be seen as composed by several layers. The basic level is inherited from propositional logic and contains Modus Ponens and the classical tautologies. Next we can recognise two levels, one referring to knowledge and one to belief. The two fragments are, as we would expect, the logic KD45$_n$ for belief and the logic S5$_n$ for knowledge. The last three axioms represent the interactions between knowledge and belief. The first and the third of these make knowledge stronger than belief, the second regulates the introspection over beliefs. It is worth noting that $K_i\phi \Rightarrow B_i\phi$ (respectively $B_i\phi \Rightarrow K_iB_i\phi$) corresponds[12] to property 1 (respectively 2) of the relation above.

Other studies for interaction between knowledge and belief have been proposed; see for example [FH94], [MS93], and [Voo92].

[11] But here we are using $G = A$.
[12] In the sense of Theorem 1.19.

1.3.2.3.2 Case 2: Knowledge and time. MAS evolve over time. Hence MAS theories need to specify the temporal evolution of the aspects of agency under investigation. In the second part of Section 1.3.1.4, we saw how to use the semantics of interpreted systems to model temporal evolution. In Section 1.3.1.1 we also saw the use of this semantics to model epistemic states of a MAS. Here we briefly present how these two formal tools can be merged together to provide a model for epistemic change. This can serve as a specification of static knowledge of a MAS about a changing world, as a model for MAS dynamic knowledge of a static world or, more generally, as a specification for dynamic epistemic states of a MAS about a changing world. The work reported here is presented in [HV86], [Mey94], [FHV92], and the recent [HMV97], where some logics that formalise a class of agents whose knowledge changes over time are discussed. The complexity of some of these logics is studied in [HV89].

There are obviously many ways in which knowledge and time can interact. Interpreted systems are more suitable for modelling intuitive properties of MAS, because of the explicit way of representing local and global states. This becomes particularly useful in the case of knowledge and time where intuitive classes of MAS can be modelled in terms of interpreted systems.

Recall from Section 1.3.1.4 that a set of global states of an interpreted system (Definition 1.4) can also be seen as a set of pairs (r, m), where $r \in R$ is a run and $m \in \mathbb{N}$ is a time instant. If we define knowledge on this structure as we did in Section 1.3.1.1 for the static case, we obtain the following. Two global states $(r, m), (r', m')$ of an interpreted systems are equivalent to agent i (formally $(r, m) \sim_i (r', m')$) if i's local state is the same in the two states, i.e. if $r_i(m) = r'_i(m')$. In other words in this way we build the generated model of the interpreted system, similarly to what we did in Definition 1.6 for the static case.

Halpern and colleagues identify important classes of multi-agent systems with respect to the relation between the agents' knowledge and time. Here we mention (a-)synchronicity, perfect recall, unique initial state, and no learning; a few other examples can be found in their papers.

An interpreted system IS is *synchronous* if for all agents i, and points $(r, m), (r', m') \in \mathcal{R}$ if $(r, m) \sim_i (r', m')$ then $m = m'$. This is a standard assumption in many real systems and what it amounts to is that the agents have access to a shared clock, actions being taken in rounds. Equivalently we can say that it is common knowledge among the agents that the system is synchronous.

Another interesting class is made by those agents that enjoy *perfect recall*. Intuitively perfect recall agents never forget the local states they have been in; in particular, once they know something they will never forget it. Formally, define agent i's local state sequence at point (r, m) to be the sequence of local states that i has gone through in run r up to time m. We can then say that an interpreted system IS has perfect recall if for all agents i, and points $(r, m), (r', m') \in \mathcal{R}$ if $(r, m) \sim_i (r', m')$ then the agent i has the same local state sequence in (r, m) and (r', m'). It can be argued that (see [FHMV95] page 130) perfect recall is a reasonable assumption for real MAS operating over short periods of time.

A further characteristics of temporal evolution of knowledge is *no learning*. Intuitively no learning is the dual of perfect recall and is the condition of agents never ruling out previously considered possible epistemic alternatives. Formally, define agent i's future local state sequence at point (r, m) to be the set of local states that i will go through at run r starting from the point (r, m). An interpreted system IS has no learning capabilities if for any pair of points points $(r, m), (r', m') \in R$, if $(r, m) \sim_i (r', m')$ then i has the same future local state sequence at (r, m) and (r', m').

The last important case we report here is of interpreted systems having a *unique starting state*. Formally, we say that an interpreted system IS has a unique starting state if for all runs $r, r' \in \mathcal{R}$, we have $(r, 0) = (r', 0)$.

The language used by Halpern and colleagues to axiomatise these semantic conditions is a modal language for knowledge plus two operators to represent time: until ($\mathcal{U}$) and next ($\bigcirc$). Indeed it is the union of the languages we employed for the study of knowledge in Section 1.3.1.1 and in the second part of Section 1.3.1.4.

Let us consider the logic system $\mathrm{S5}_n$ with respect to a family of modal operators representing knowledge $K_i, i \in A$ and the axiomatic system presented on page 28. Call $\mathrm{S5}_n^U$ the the combination between the two.

The authors investigate completeness with respect to the following list of axioms:

1. $K_i \bigcirc \phi \Rightarrow \bigcirc K_i \phi$,

2. $K_i\phi_1 \wedge (K_i\phi_2 \wedge \neg K_i\phi_3) \Rightarrow \neg K_i \neg((K_i\phi_1)\,\mathcal{U}\,((K_i\phi_2)\,\mathcal{U}\,\neg\phi_3)))$,

3. $K_i\phi_1\,\mathcal{U}\,K_i\phi_2 \Rightarrow K_i(K_i\phi_1\,\mathcal{U}\,K_i\phi_2)$,

4. $\bigcirc K_i\phi \Rightarrow K_i \bigcirc \phi$.

The important result that we cite here is the following:

Theorem 1.31. *[HMV97]*

1. $S5_n^U$ *is a sound and complete axiomatisation with respect to the class of asynchronous systems, synchronous systems, systems with a unique initial state and synchronous systems with unique initial state.*

2. $S5_n^U$ *+ 2 is a sound and complete axiomatisation with respect to systems with perfect recall and systems with perfect recall with a unique starting state.*

3. $S5_n^U$ *+ 1 is a sound and complete axiomatisation with respect to systems with perfect recall and synchronicity and systems with perfect recall, synchronicity and unique initial state.*

4. $S5_n^U$ *+ 3 is a sound and complete axiomatisation with respect to systems with no learning.*

5. $S5_n^U$ *+ 2 + 3 is a sound and complete axiomatisation with respect to systems with no learning and perfect recall.*

6. $S5_n^U$ *+ 4 is a sound and complete axiomatisation with respect to systems with no learning and synchronicity.*

7. $S5_n^U$ *+ 1 + 4 is a sound and complete axiomatisation with respect to systems with no learning, perfect recall and synchronicity.*

8. $S5_n^U$ *+ 1 + 4 +* $\{K_i\phi \equiv K_1\phi\}$ *is a sound and complete axiomatisation with respect to systems with no learning, synchronicity and unique initial state and with respect to systems with no learning, synchronicity, unique initial state and perfect recall.*

All the logics defined above are also decidable and results on their complexity are presented in [HMV97].

The above is an important class of results for interactions between knowledge and time. By adding group properties of knowledge it is possible to enrich the axiomatisation. It should be noted though that in some cases by adding common knowledge completeness may be lost and an accurate analysis is required.

1.3.2.3.3 The BDI framework. The BDI framework [RG91, Rao96, RG98] is a rich and powerful logical framework that has been developed for more than five years mainly by Anand Rao and Michael Georgeff at the Australian Artificial Intelligence Institute. The BDI model focuses on three components of an agent: its beliefs, desires and intentions.

The BDI framework is defined on the branching time logic CTL*. For the purposed of this work here we only describe the static fragment of BDI logics, so no temporal evolution will be present (for further references see [RG98]).

The language is a propositional modal language with three families of modal operators $B_i, D_i, I_i, i \in A$. The belief operator B_i is the one described in Section 1.3.1.2, the intention operator $I_i\phi$ is the one described in Section 1.3.1.3. Formulae $D_i\phi$ are read as "agent i desires to bring about ϕ".

The semantics is given through standard Kripke frames. Three families of accessibility relations are defined, one for belief, $\{b_i\}$, one for intention, $\{i_i\}$, and one for desire, $\{d_i\}$. The binary relations b_i are Euclidean, transitive and serial, the relations for intention i_i and desires d_i are serial.

Although Rao and Georgeff suggest that axioms need to be tailored to the specific application, they do suggest a family of logics for which they prove completeness. All these are based on a logic that contains a KD45 fragment for belief and two KD fragments for desire and intention. The key assumption of the BDI framework are that:

> "Intentions are stronger than desires and desires are stronger than beliefs". ([Rao96], page 7)

The three families of relations are linked by the following two conditions:

1. For any $i \in A$ $b_i \subseteq d_i$;
2. For any $i \in A$ $d_i \subseteq i_i$.

Interpretation is defined in the usual way described in Section 1.2.

The following is a sound and complete axiomatisation of the modal logic with respect to the class of frames described above:

Taut	$\vdash t$, where t is any propositional tautology
K_B	$\vdash B_i(p \Rightarrow q) \Rightarrow (B_i p \Rightarrow B_i q)$
D_B	$\vdash B_i p \Rightarrow \neg B_i \neg p$
4_B	$\vdash B_i p \Rightarrow B_i B_i p$
5_B	$\vdash \neg B_i p \Rightarrow B_i \neg B_i p$
K_I	$\vdash I_i(p \Rightarrow q) \Rightarrow (I_i p \Rightarrow I_i q)$
D_I	$\vdash I_i p \Rightarrow \neg I_i \neg p$
K_D	$\vdash D_i(p \Rightarrow q) \Rightarrow (D_i p \Rightarrow D_i q)$
D_D	$\vdash D_i p \Rightarrow \neg D_i \neg p$

IB1	$\vdash I_i p \Rightarrow D_i p$
IB2	$\vdash D_i p \Rightarrow B_i p$
US	If $\vdash \phi$, then $\vdash \phi[\psi_1/p_1, \ldots, \psi_n/p_n]$
MP	If $\vdash \phi$ and $\vdash \phi \Rightarrow \psi$, then $\vdash \psi$
Nec_B	If $\vdash \phi$, then $\vdash B_i\phi$
Nec_D	If $\vdash \phi$, then $\vdash D_i\phi$
Nec_I	If $\vdash \phi$, then $\vdash I_i\phi$

It would be reductive to say that BDI logics are confined to the above framework. In fact for the static case many more logics are studied in [Rao96], where more interaction axioms are studied. [RG98] contains sound and complete axiomatisations for the dynamic case, where the authors allow for the syntax to express temporal CTL*-style as well as mentalistic operators. We do not discuss this issue here.

1.4 Knowledge sharing among a community of agents

In the previous section we reviewed some of the best known MAS theories and we stressed the need of expressing *interactions* between mental attitudes of the agents. We observed that in order to build fully-fledged MAS the different mental attitudes of agents cannot be specified as being independent from each other as we need to express interactions between belief and knowledge (as in case study 1 of Section 1.3.2.3.1), between desire and intention (as in Section 1.3.2.3.3 for example), etc.

All the examples of interaction we reviewed in Section 1.3.2.3 were heterogeneous in type and homogeneous in subject. What I mean by that is that all the interactions (in the sense of Figure 1.5) involved were between *different mental attitudes* of the *same agent*. But there is another type of interaction that we have not discussed yet which is homogeneous in type and heterogeneous in subject. This is an interaction that involves the *same mental attitude* of *different agents*. An example will clarify the issue.

Consider a group of agents whose abilities involve some sort of communication[13]. Surely, communication involves information passing among the agents involved. In any scenario in which agents are co-ordinating towards a goal to be accomplished (for example selling and buying goods in an electronic commerce scenario as in [CM96]), the intentions of one agent change depending on the intentions of the others. Put it in different words, an interaction between the agents' intentions is present and so we need capture it.

We have a choice here. We could either use a fine grain of abstraction and implement every message passing activity (using the formal model presented in Section 4 of [FHMV95] for example) that produces the interactions between the intentions in the system, or, instead, we could take a high level approach and limit ourselves to express the relation that originates from this between the agents' intentions. Each approach has different advantages and disadvantages; by choosing the former we have more expressivity, by following the latter we can avoid being bogged down with too many details and it will be easier to examine properties of intention.

[13]Communication has been represented formally in [CP79], [CL90]; the key idea of these works is that communicative utterances are actions (indeed that idea can be dated back to [Aus62]). Automated massage passing systems, a class of communicating MAS modelled by a variant of interpreted systems based on actions, have been formalised in [FHMV95] as in other papers by Halpern and Vardi. For the purposes of this introduction we need not discuss these issues.

Whatever grain of detail we decide to employ, which should ultimately depend on what of these aspects we value most, we need to be able to specify this kind of heterogeneous interaction.

1.4.1 Degrees of knowledge sharing

Above we have seen an informal example of interaction between the intentions of different agents. For similar reasons, studies of interactions between other mental attitudes are equally in need.

This thesis is a study of *interactions* between the *knowledge* of agents in a MAS.

In *practical* AI the importance of interaction of the knowledge of agents is quite evident as we need theories able to express scenarios in which the agents are conceptually different. For example, consider a distributed system composed by a group of agents and the two following situations:

One agent knowing everything the others know. An agent j is the central processing unit of a distributed system of agents whose non-specialised entities transmit their knowledge to a central unit j[14].

Linear order in agents' private knowledge. Several processes with the same information at disposal but different computational power are running the same program. Under certain assumptions, it is reasonable to assume that the knowledge of the agents, seen as knowledge bases, increases with the order of computational power at disposal.

These are only two of the many scenarios we can think of in which general pattern of interactions are needed to specify knowledge in a MAS. As we shall see in this thesis, the theme of interaction between knowledge of agents is also *theoretically* quite interesting.

If there is interaction of knowledge in the group, it means that knowledge is effectively being *shared* in the group of agents. By this I do not mean that agents have exactly the same data structures and that they do not have a private core of personal knowledge because if this happened they would no longer be *autonomous* and so they would not comply with the definition of agent we gave on page 10. What I mean is that they share *some* information about the environment. As we will see there are many degrees to which knowledge can be shared in a MAS.

1.4.2 The content of this thesis

In this work I will present some results to the problem of modelling knowledge sharing by taking different perspectives.

Given our discussion so far, there are three main approaches that we could take to model a MAS in which knowledge is being shared:

1. Specify the evolution of the system,
2. Give an axiomatic account of the logic describing the system,

[14] For example, this conceptual scenario has been successfully used to obtain encouraging preliminary results on the problem of collective map making for robotic agents [dMAE$^+$97].

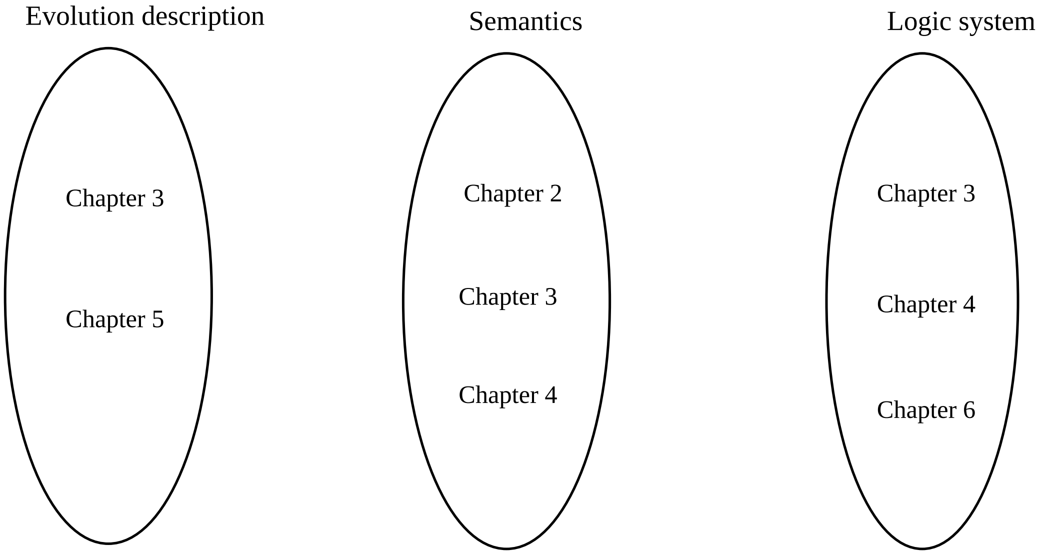

Figure 1.6: The chapters of this thesis classified according to their content.

3. Give a semantical account of the logic describing the system.

Figure 1.6 classifies the chapters of this thesis according to this classification (some chapters span over more than one of the classes).

The first approach is quite low-level. It involves formally describing the evolution of the MAS, for example by describing any form of communication that takes place and spelling out how this affects the knowledge of the agents in the system. From our previous analysis good candidates for this are variants of the framework of interpreted systems that we described in Section 1.3.1.4: the formal tool developed in [Mey96] and the notion of context in [FHMV95] seem good choices here. This is a dynamic model, i.e. rules for state transition will have to be specified.

The axiomatic approach is radically different, as it completely ignores how knowledge sharing is achieved but it only focuses on what logic describes the system at any time point. Differently from above, here we have the choice of whether or not to implement time explicitly. We can either specify the temporal evolution of knowledge or focus on its static properties. In this work we will follow this second option; given the discussion carried out so far, it is clear that extensions of the logic $S5_n$ with opportune interaction axioms are a good choice here.

The semantic approach sits in the middle of the two. It offers a description which is quite low level but avoids all the details of the full descriptive approach. Again, we have the choice of modelling temporal epistemic change (along the lines of [EV98] and [HMT94] for example) or focus on the static property of knowledge. In accordance to the previous point we will opt for a static analysis. Two candidates seem promising here: Kripke models and static interpreted systems.

There certainly is an interconnection between these three approaches, but it should be stressed that it is not the case that any description of a MAS carried out in one of the three formalisms above can be translated into one in another class. For example, there are certainly many evolution descriptions of the distributed-AI scenarios described above, quite simply because the knowledge could be sent and processed in many different ways while still producing the same macroscopic phenomenon.

Therefore if one were to investigate theoretically a MAS and would like to be able to identify descriptions of it in *all* the three classes (in order to provide a formal basis for validation and verification), he would quite simply be left with the question of which class to use *initially* to describe the MAS before attempting to translate the obtained specification into the another class. Is it better to employ a logic, a semantic, or a more low-level evolution description? I will not answer this general question in this thesis, but we will analyse the three cases.

Still, in this work one successful investigation is carried out where we will be able to identify representatives of a MAS specification in each of these classes. Crucially this will be possible by starting from the *semantics* of the MAS in analysis. Indeed in Chapter 2 we will study the semantics of a class of MAS (that we call hypercubes) for which this correspondence can be drawn. In Chapter 3 we will give their corresponding logic system by presenting their complete and decidable axiomatisation and in the last section of the same chapter we briefly present part of some joint work with Ron van der Meyden, University of Technology (Sydney) on a description of hypercubes in a low-level formal language that allows us to represent actions and communication. By drawing this correspondence we will observe that hypercubes are MAS that share information in a peculiar way (Chapter 3), and that this is the result of the agents being able to send information (and therefore to exchange knowledge) by doing broadcasting (Section 3.8).

The axiomatisation of hypercubes developed in Chapter 3 will itself raise more questions about other interaction axioms of agents sharing information. We will answer many of these questions in Chapter 4 by proving completeness and decidability of many other classes of MAS. Chapter 4 is therefore to be regarded as both in the logic and in the semantics conceptual classes.

Having done this, in Chapter 5 we will leave the proof theory and go back to a more descriptive analysis of how to represent and update epistemic scenarios of MAS. A key example here will be again about communication and how knowledge is shared following communication acts, including an analysis of how common knowledge is affected.

We will conclude in Chapter 6 by going back to logic and semantics and discuss on a much higher level whether in the future methods to produce complete and decidable axiomatisation of MAS that rely on abstract results about combination of logics, will become available and if so what impacts this will have on the area. This will mainly be a review of the literature in the area and I will propose my personal view.

The chapters of this thesis can be ordered according to the level of abstraction (see Figure 1.6). The last chapter is indeed the most abstract of all in this document since it is about techniques according to which logic specification can be produced. Following this order, Chapters 4 and 3 (being mainly about logic systems) would then follow; Chapters 2 and Chapters 5 are progressively more low-level.

There is a last but still quite important point I would like to make before moving to the technical core of this work. Although set in the cultural framework of MAS theories, most of the formal tools employed here are from modal logic and I would like to think this thesis as having a contribution to the field of modal logic itself.

In fact, by taking a radically different perspective we can see these pages as a formal investigation on some properties of multi-modal logics. For some reason this theme has not really received the attention it really deserves, even if Dana Scott nearly 30 years ago wrote

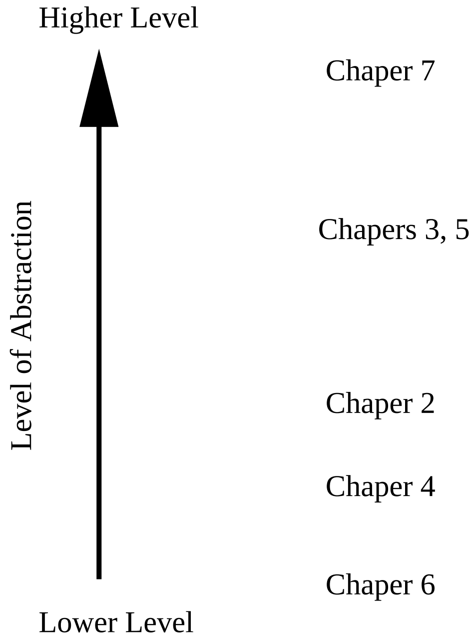

Figure 1.7: The chapters of this thesis ordered according to their level of abstraction.

the following important passage:

> [The exclusive study of mono-modal systems] is what I consider one of the biggest mistakes of all in modal logic: concentration on a system with just *one* modal operator. The only way to have any philosophically significant results in deontic or epistemic logics is to combine those operators with: tense operators (otherwise how can you formulate theories of change?); the logical operators (otherwise how can you compare the relative with the absolute?); the operators of historical or physical necessity (otherwise how can you relate the agent with its environment?); and so on and so on.
> ([Sco70], page 161).

Indeed it could be said that this thesis is about extensions of the multi-modal logic $S5_n$. Chapter 3 contains some axiomatisation problems for several special classes of equivalence frames and an example of a same modal system being complete and decidable with respect to many different classes of Kripke frames. The logic we will study there resembles the logic $S4.2_n$ and indeed it refers to a class of frames that will remind us of convergent frames. In Chapter 4 a systematic study of interaction axioms for extensions of $S5_n$ is carried out, providing some insights into how far we can push $S5_n$ before it collapses into the logic Triv. Chapter 5 can be seen as an investigation on multi-modal Kripke semantics itself and compares the effectiveness of Kripke frames to Kripke trees and provides a formal algorithm for performing updates on Kripke trees. In Chapter 6, although we review general techniques of combining logics, the focus will be on their potential for proving completeness and decidability in the multi-modal case. Finally, Chapter 2 can be seen an investigation into the relation between interpreted systems and Kripke models, the two commonly preferred semantics on which $S5_n$ is defined.

Publication note. Except where acknowledged, all the results presented in this thesis constitute original research carried out by the author. Many of these results appear in proceedings of international journals, conferences or are in preparation for submission.

More specifically, the correspondence between hypercube systems and Kripke models presented in Chapter 2 appeared in [LR98d] in the Proceedings of the AI97 Workshop on Theoretical and Practical Foundation of Intelligent Agents, Perth (Western Australia).

The main results of Chapter 3 on the axiomatisation of hypercube systems was presented as full paper [LR98b] at the European Conference of Artificial Intelligence in Brighton (UK) in August 1998; a longer version of that paper, inclusive of all the mathematical proofs, appeared as technical report [LR98a] of the School of Computer Science of the University of Birmingham. A journal submission [LMR99] focused on the characteristics of homogeneous broadcasting systems is in preparation.

The results concerning the spectrum of knowledge sharing reported in Chapter 4 will also appear, in a much shortened version, in a research paper [LR99b] in the Proceedings of the Sixth International Workshop on Agent Theories, Architectures, and Languages (ATAL-99). An extended version of that paper with decidability results is in preparation ([LHR99]).

A preliminary approach on the issues of refining and updating Kripke models, which largely constitute Chapter 5, was presented in [LR98c]. An extended version [LR99a] of that work, whose technical content is largely similar to the above mentioned chapter, appeared in April 1999 in the Journal of Artificial Intelligence for Engineering Design, Analysis and Manufacturing.

Finally, Chapter 6 builds upon one of the author's progress reports [Lom97] and parts of it appear in a forthcoming journal submission [Lom99] to the International Journal of Knowledge Engineering Review.

Chapter 2

Hypercube systems

2.1 Introduction

In Section 1.4.2 we motivated the need for investigating classes of MAS whose knowledge is related by interaction patterns. In the same section we argued that there are three conceptual areas in which we can carry out an investigation: by means of a low-level description of a system, by specifying its semantics or, more intuitively, by giving an axiomatisation. In section 1.3 we have explored some possibilities.

As we argued, it is not clear whether for a given system we can find an appropriate description for it in each of these categories. What is clear is that having one-to-one correspondences between different formalisms expressing the same MAS at different level of abstraction would be an advantage. This is so because it would enable us to choose the appropriate level of description depending on the kind of properties we want to specify, verify or, simply, reason about. Consider for example a case of agents communicating to acquire knowledge about the environment. If we wanted to check some properties of their knowledge, we would be inclined to reason on their axiomatisation, but if we needed to examine the sequence of communication actions being performed, we would instead reason on the low-level description of the system.

In this and in the following two chapters we will study hypercube systems, a particular kind of ideal agents of knowledge (Section 1.3.1.1), for which these correspondences can be drawn. We will achieve this result by starting from the semantics[1] that describes them, i.e. a variant of static interpreted systems.

In this chapter our only task is to understand how hypercube systems relate to Kripke models. The reason why we will do this is the following. Recall from Section 1.3.1.1 that two semantics are available for the case of knowledge in MAS: Kripke models and interpreted systems. The latter present the advantage of being more intuitive (as shown in the case study of Section 1.3.2.3.2), the former come with an heritage of techniques (summarised in Section 1.2.4) which are extremely useful to prove properties of the formalisation. Because of this observation, the ideal scenario of investigation is one in which we anchor our specification to interpreted systems but retain the possibility of proving properties about them by means of the more technically explored Kripke models.

We prepare the grounds for this investigation in this chapter where we formally relate

[1]Indeed this is the most commonly used method in Computer Science to define a class of systems. This is so because it is conceptually close to the system, while it already has the advantage of being formal.

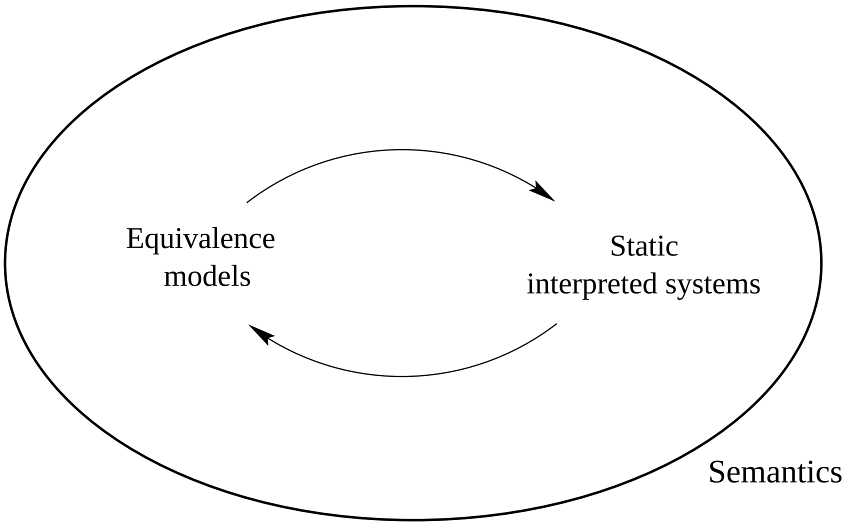

Figure 2.1: The content of this chapter. The arrows represent the correspondence that is drawn between particular classes of equivalence models and static interpreted systems.

hypercube systems to a special class of Kripke models (see Figure 2.1). The main result here is to find the semantic equivalents of hypercube systems in the class of equivalence Kripke models. This will allow us to continue our study in Chapter 3 where, by axiomatising the corresponding class of Kripke frames, we will show that hypercube systems constitute a semantic class of agents that share part of their knowledge.

The methodology employed here could possibly be extended for mapping other classes of interpreted systems to Kripke semantics, but it is worth stressing that the analysis carried out here only applies to the case of hypercube systems.

2.2 Definition of hypercube systems

Recall from Section 1.3.2.3.2 the definition of interpreted system. These are defined to give an account of the temporal evolution of a system by defining runs over global states. If we fix the time of an interpreted system to analyse static properties of its global states, we simply obtain a pair of global states with an interpretation, i.e. what we called static interpreted systems (Definition 1.5). In Section 1.3.1.1 we discussed how to represent knowledge in a static interpreted system. This essentially involves generating an equivalence model by using the construction given in Definition 1.6 for the case that the predicate $P_i, i \in A$ is the equality on local states for agent i, as we discussed in more detail on page 24.

In order to introduce hypercube systems we discuss an example of a static interpreted system. Figure 2.2 shows the underlying structures of a static interpreted system and its corresponding generated model[2]. As expected, we find that two global states are related for agent i if agent i has the same local state in the two global states. In the example we can see that at any state no agent knows the local state of the environment. In fact, for any global state each agent considers the possibility of another global state in which the environment

[2]For simplicity the reflexive links are not illustrated. Also in Figure 2.2 the relations are the transitive closure of the ones depicted.

$$L_1 = \{l_1, m_1\} \qquad S = \{(l_1, l_2, l_e), (l_1, l_2, m_e), (m_1, l_2, l_e)\}$$
$$L_2 = \{l_2, m_2\} \qquad \{(m_1, m_2, l_e), (m_1, m_2, m_e), (m_1, l_2, m_e)\}$$
$$L_e = \{l_e, m_e\}$$

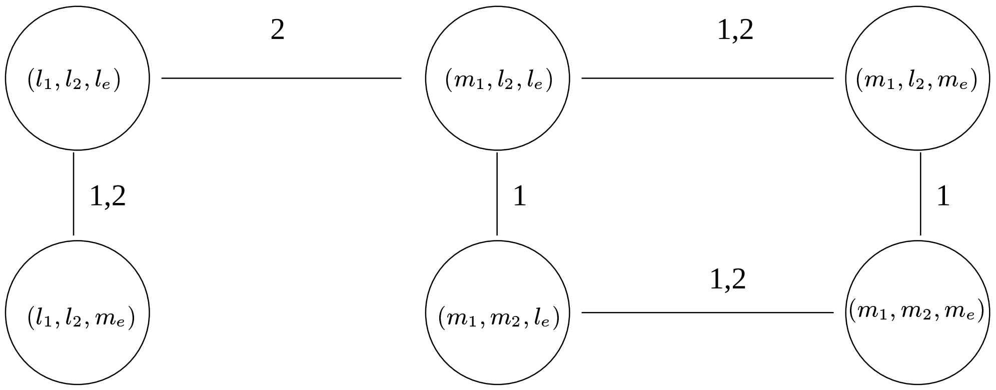

Figure 2.2: An example of the underlying structure of a static interpreted system and its corresponding generated frame. For simplicity the reflexive links are not illustrated and intended to be the transitive closure of the ones depicted.

has a different local state. This is not surprising at all, as we assumed that agents know only about their local states. Local states can encode some information about the environment but this has to be considered not totally observable.

More importantly, in the global state (l_1, l_2, l_e), the top left state in the figure, agent 1 considers two global states possible: (l_1, l_2, l_e) and (l_1, l_2, m_e). In both of these global states agent 2's configuration is the same. So in that global state agent 1 has complete information about agent 2, i.e. when in the system is in that global state agent 1 *knows* agent 2's local state.

The converse of this situation is when an agent has no information at all about another. This happens in the example for agent 2 in the global state (l_1, l_2, l_e). In this case not only agent 2 considers the possibility of another global state where agent 1 is in the local state l_1, but also agent 1 has no information about possible dependencies between agent 1's local state and the environment's. As it can be verified in the example all the global states of the Cartesian product $L_1 \times \{l_2\} \times L_e$, where the local states are defined in the figure, are effectively present.

Hypercube systems are static interpreted systems in which all the agents at all global states are in the situation of agent 2 at (l_1, l_2, l_e). Interestingly, although they have no information about each other's local states, we will note at the end of Chapter 3 that they still share some knowledge. Formally, hypercube systems result by considering the admissible state space of the MAS to be described by the full Cartesian product of its sets of local s-tates. This means that every global state is in principle possible, i.e. there are no mutually exclusive configurations between local states. With hypercubes systems we are imposing a further simplification on the notion of static interpreted system presented in Definition 1.5: in the tuples representing the configuration of the system we do not consider a slot for the environment. The presence of the environment in the notion of Fagin et al. [FHMV95] is motivated in order to keep track of the changes in the system and in general to represent everything that cannot be captured by the local states of the single agents (most importantly messages in transit, etc.). This restriction is equivalent to analysing tuples in which the environment is constant. We will reintroduce the environment in Section 3.8 where we will discuss a low-level formal model for hypercube systems, in which we will introduce actions for the agents and the environment.

So, the following are our formal definitions for hypercube system.

Definition 2.1 (Hypercube states). *A hypercube state, or hypercube, is a Cartesian product $H = L_1 \times \cdots \times L_n$, where $L_1, \ldots, L_n$ are non-empty sets of local states. The class of hypercube systems is denoted by $\mathcal{H}$.*

Definition 2.2 (Hypercube systems). *A hypercube system is a static interpreted system $HS = (H, \pi)$, where H is a hypercube state. The class of hypercube systems is denoted by $\mathcal{HS}$.*

Static interpreted systems were defined in Definition 1.5.

2.3 Maps between hypercubes and equivalence frames

The aim of this section is to relate semantically hypercube systems to Kripke models. More specifically, we would like to identify the class of Kripke models that satisfy exactly the same formulae satisfied by hypercube systems. Given the notion of validity of formulae on static interpreted systems and Kripke models of Definitions 1.7 and 1.3, it is appropriate to compare the two underlying semantic structures: hypercube states and Kripke frames.

We noted in Definition 1.6 that static interpreted systems can be seen as a special class of Kripke models. In Section 1.3.1.1 we noted that for the case of knowledge, these generated models are equivalence models. Since hypercube systems are a special case of static interpreted systems for knowledge, it is reasonable to think that they also generate a special case of equivalence Kripke models. In order to clarify the relationship, we define and analyse mappings between the two underlying structures. Specifically we proceed as follows.

- We define a map $\mathcal{H} \xrightarrow{f} \mathcal{F}$ from hypercubes to equivalence Kripke frames.
- We define a map $\mathcal{F} \xrightarrow{g} \mathcal{H}$ from equivalence Kripke frames to hypercubes.
- We analyse the compositions of the maps f and g.
- We isolate the images of f in $\mathcal{F}$.

Hypercubes and frames are always defined over a set A of n agents, which we assume as given.

Every hypercube generates a frame [FHMV95]:

Definition 2.3 (Hypercubes to frames). *The function $f : \mathcal{H} \to \mathcal{F}$ that maps the hypercube H onto the Kripke frame $f(H)$ is defined as follows:*

If $H = L_1 \times \cdots \times L_n$, $f(H) = (L_1 \times \cdots \times L_n, \sim_1, \ldots, \sim_n)$, where $\sim_i$ is defined as: $(l_1, \ldots, l_n) \sim_i (l'_1, \ldots, l'_n)$ if and only if $l_i = l'_i$.

Definition 2.3 basically expresses the construction we saw in Definition 1.6 for non-interpreted structures in the case of knowledge ($\{P_i\}_{i\in A}$ being a family of equality predicates on the local states of the agent in question as discussed above). It is clear from the definition that $f(H)$ is an equivalence frame. What is less obvious is that it is a very particular equivalence frame. In fact the following holds.

Lemma 2.4. *If H is a hypercube, and $f(H) = (W, \sim_1, \ldots, \sim_n)$ is the frame defined from it by Definition 2.3, then*

1. $\bigcap_{i\in A} \sim_i = id_W$*;*
2. *For any $w_1, \ldots, w_n$ in W there exists a world $\overline{w}$ such that $\overline{w} \sim_i w_i$, $i = 1, \ldots, n$.*

Proof. For 1, Consider any two elements $w = (l_1, \ldots, l_n), w' = (l'_1, \ldots, l'_n)$ in W such that $w(\bigcap_{i\in A} \sim_i)w'$. Then for all i in A, $(l_1, \ldots, l_n) \sim_i (l'_1, \ldots, l'_n)$. Therefore by definition, for all i in A, $l_i = l'_i$, that is $w = w'$.

For 2, consider any $w_1 = (l^1_1, \ldots, l^1_n), \ldots, w_i = (l^i_1, \ldots, l^i_n), \ldots, w_n = (l^n_1, \ldots, l^n_n)$. Let $\overline{w} = (l^1_1, \ldots, l^i_i, \ldots, l^n_n)$. By Definition 2.3, the element $\overline{w}$ is in W and for each i, $\overline{w} \sim_i w_i$. $\square$

This shows that the Kripke frames that we build from hypercubes by means of the standard technique (Definition 1.6) constitute a proper subset of all the possible reflexive, symmetric and transitive Kripke frames. To relate the two semantic classes, we have to analyse the two properties expressed in Lemma 2.4.

The first one says that in a frame generated by a hypercube there cannot be two distinct states related by all the equivalence relations. This follows from the fact that the environment is constant in our formalism.

The second property reflects the fact that hypercubes are full Cartesian products. The property expresses the circumstance that for every pair of points in the n dimensional space of the images of the hypercubes, there are $n!$ ways to connect them in two steps. Indeed, we can change $n-1$ coordinates in n possible ways and change the last one in the last step.

Given these differences between the class of hypercubes and equivalence frames, it is likely that the two semantic structures satisfy different formulae. We will come back to this point in the next chapter.

It is also possible to generate a hypercube from a frame:

Definition 2.5 (Frames to hypercubes). *The function $g : \mathcal{F} \to \mathcal{H}$ maps a frame $F = (W, \sim_1, \ldots, \sim_n)$ onto the hypercube $g(F) = W/\sim_1 \times \cdots \times W/\sim_n$.*

We now have defined maps between the two semantic structures. Our aim is to use them to identify the class of equivalence frames that are semantically equivalent, i.e. that satisfy the same formulae, to hypercubes. In order to do so, we introduce a notion of isomorphism on $\mathcal{F}$ and $\mathcal{H}$. Many notions (such as p-morphisms or bisimulations for frames) may be appropriate for this task, but for our aims we need a strong equivalence between the structures.

Consider two MAS. If we can draw a bijection between the agents of the MAS such that the local states of the corresponding agents are themselves in a bijection, then in a way we can think that one MAS can simulate the other, and so the two MAS can be thought as being equivalent. We formalise this as follows:

Definition 2.6 (Isomorphism of hypercubes). *Two hypercubes $H = L_1 \times \cdots \times L_n$, $H' = L'_1 \times \cdots \times L'_n$ are isomorphic ($H \cong_{\mathcal{H}} H'$) if $|L_i| = |L'_i|$ for $i = 1, \ldots, n$.*

To reason about equivalent frames we take the standard notion of isomorphism (Definition 1.11).

We can prove that the maps we defined preserve isomorphisms:

Lemma 2.7. *If $H \cong_{\mathcal{H}} H'$, then $f(H) \cong_{\mathcal{F}} f(H')$.*

Proof. Let $H = L_1 \times \cdots \times L_n$, and $H' = L'_1 \times \cdots \times L'_n$. Since $H \cong_{\mathcal{H}} H'$ there is a family of bijections $b_i : L_i \to L'_i$. Consider $b = b_1 \times \cdots \times b_n$. The function b is a bijection, and therefore the universes of the frames $f(H)$ and $f(H')$ are in a bijection.

Consider now $s = (l_1, \ldots, l_i, \ldots, l_n)$, $s' = (l'_1, \ldots, l'_i, \ldots, l'_n)$ such that $s, s' \in H$, and $s \sim_i s'$ on $f(H)$. Consider $b(s) = (b_1(l_1), \ldots, b_i(l_i), \ldots, b_n(l_n))$ and $b(s') = (b_1(l'_1), \ldots, b_i(l'_i), \ldots, b_n(l'_n))$. Since, by definition, $l_i = l'_i$, then $b_i(l_i) = b_i(l'_i)$ and therefore $b(s) \sim'_i b(s')$.

Let now be $b(s) \sim'_i b(s')$. Then, by definition $b_i(l_i) = b_i(l'_i)$ and then $l_i = l'_i$, that implies $s \sim_i s'$. □

Lemma 2.8. *If $F \cong_{\mathcal{F}} F'$, then $g(F) \cong_{\mathcal{H}} g(F')$.*

Proof. Consider two isomorphic frames $F = (W, \sim_1, \ldots, \sim_n)$, $F' = (W', \sim'_1, \ldots, \sim'_n)$ and consider the induced bijection $b : W \to W'$ between W and W'. We want to prove that there is a family of bijections c_i between the components of $g(F) = W/{\sim_1} \times \cdots \times W/{\sim_n}$ and $g(F') = W'/{\sim'_1} \times \cdots \times W'/{\sim'_n}$. Let $c_i : W/{\sim_i} \to W'/{\sim'_i}$ such that $c_i([w]_{\sim_i}) = [b(w)]_{\sim'_i}$.

The function c_i is well defined. In fact, let $[w_1]_{\sim_i} = [w_2]_{\sim_i}$, with $w_1, w_2 \in W$. Then $c_i([w_1]_{\sim_i}) = [b(w_1)]_{\sim'_i} = [b(w_2)]_{\sim'_i} = c_i([w_2]_{\sim_i})$.

The function c_i is injective. $c_i([w_1]_{\sim_i}) = c_i([w_2]_{\sim_i})$, then $[b(w_1)]_{\sim'_i} = [b(w_2)]_{\sim'_i}$ that is $b(w_1) \sim_i b(w_2)$, $w_1 \sim_i w_2$ and then $[w_1]_{\sim_i} = [w_1]_{\sim'_i}$.

The function c_i is surjective. Consider $[w']_{\sim'_i}$, such that $w' \in W'$ and let $w \in W$ be such that $b(w) = w'$. Then $c_i([w]_{\sim_i}) = [w']_{\sim'_i}$. □

Figure 2.3 shows the preservation of isomorphisms under f and g between frames and hypercubes as proved Lemmas 2.7 and 2.8. Since we want to import and export results from one structure into the other, this is the result we need.

2.4 Composition of maps between hypercubes and Kripke frames

We now investigate the extent to which the composition of f with g (respectively g with f) results in a hypercube (respectively frame) which is isomorphic to the one we started with. We do this for two reasons. First we want to check whether by going back and forth between the two class of structures we lose information, i.e. the structure we obtain satisfies different formulae from the original one. Secondly, this will help us to prove a general the result on the correspondence between hypercubes and equivalence frames. We operate as follows.

Given a hypercube $H = L_1 \times \cdots \times L_n$, consider the image under f of $H, f(H)$. Let $H' = (L_1 \times \cdots \times L_n)/{\sim_1} \times \cdots \times (L_1 \times \cdots \times L_n)/{\sim_n}$ be the image under g of $f(H)$. We want to investigate the relationship between H and H'.

Theorem 2.9. *For any hypercube H in $\mathcal{H}$, $H \cong_{\mathcal{H}} g \circ f(H)$.*

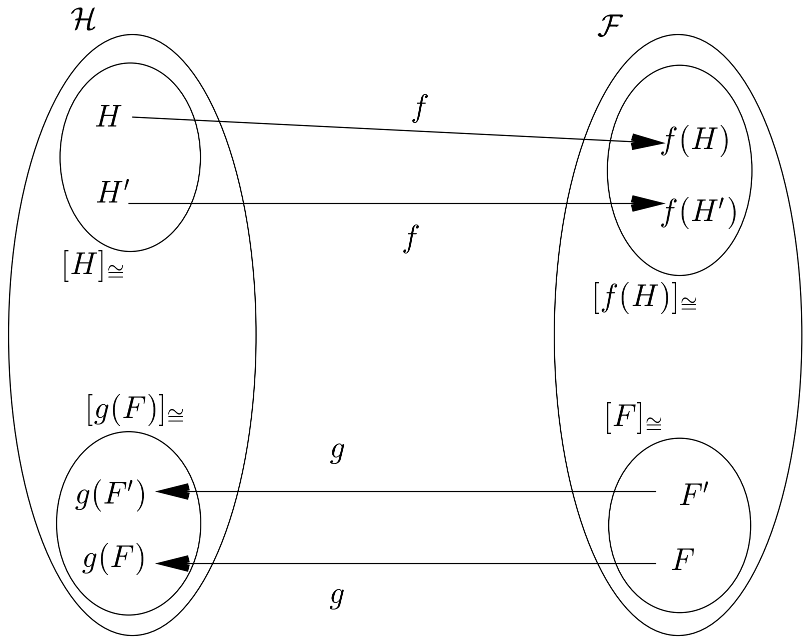

Figure 2.3: Preservation of isomorphisms under the maps.

Proof. We prove that $b_i : L_i \to (L_1 \times \cdots \times L_n)/{\sim_i}$, defined as $b_i(l_i) = [(l_1, \ldots, l_i, \ldots, l_n)]_{\sim_i}$, where $l_j, i \neq j$, is an arbitrary element in L_j, is a bijection.

The function b_i is well defined. Let $l_i = l'_i$. Then $b_i(l_i) = [(l_1, \ldots, l_i, \ldots, l_n)]_{\sim_i}$ and $b_i(l'_i) = [(l'_1, \ldots, l'_i, \ldots, l'_n)]_{\sim_i}$. But $(l_1, \ldots, l_i, \ldots, l_n) \sim_i (l'_1, \ldots, l'_i, \ldots, l'_n)$ and therefore $b_i(l_i) = b_i(l'_i)$.

The function b_i is an injection. In fact, let $b_i(l_i) = b_i(l'_i)$, then $[(l_1, \ldots, l_i, \ldots, l_n)]_{\sim_i} = [(l'_1, \ldots, l'_i, \ldots, l'_n)]_{\sim_i}$, that implies $l_i = l'_i$.

The function b_i is a surjection. Consider any $[(l_1, \ldots, l_i, \ldots, l_n)]_{\sim_i} \in (L_1 \times \cdots \times L_n)/{\sim_i}$. We have $b_i(l_i) = [(l'_1, \ldots, l_i, \ldots, l'_n)]_{\sim_i} = [(l_1, \ldots, l_i, \ldots, l_n)]_{\sim_i}$. □

See also Figure 2.4. In other words, if we start from a hypercube H, and consider the generated Kripke frame $f(H)$, it is still possible to extract all the information from the frame by applying the function g that produces another system H', which is in a bijection with the original H.

Corollary 2.10. *For any hypercube H in $\mathcal{H}$ and formula $\phi \in \mathcal{L}$, $H \models \phi$ if and only if $g \circ f(H) \models \phi$.*

Proof. It follows from Theorem 2.9, Definition 1.7, Lemma 2.7 and Lemma 1.12. □

We now investigate the other side of the relation. Consider a frame F and its image under g, $g(F)$. If we take the image under f of $g(F)$, that frame will satisfy the properties stated by Lemma 2.4 and therefore will not in general be isomorphic to F.

What we can prove is the following (See also Figure 2.4):

Lemma 2.11. *If F is a frame such that there exists a hypercube H, with $F \cong_{\mathcal{F}} f(H)$, then $F \cong_{\mathcal{F}} f \circ g(F)$.*

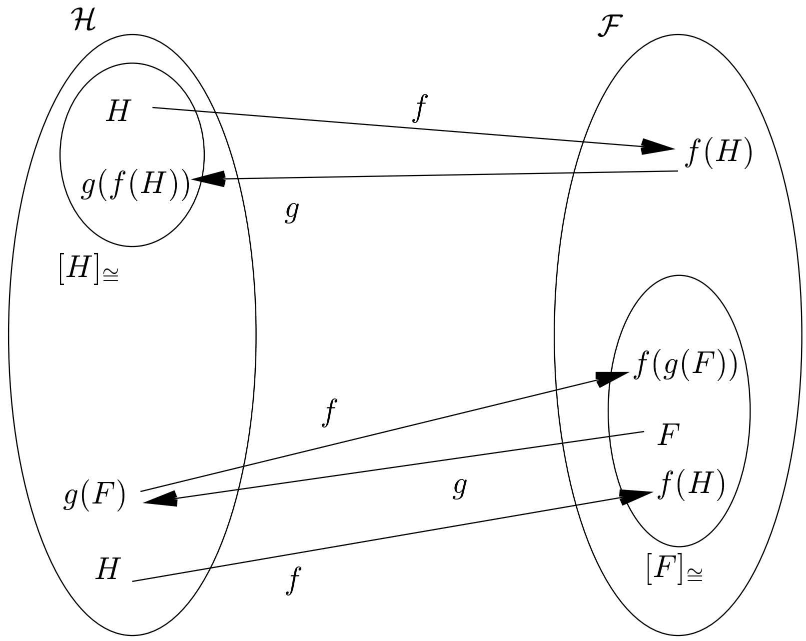

Figure 2.4: Compositions of maps between frames and hypercubes as in Theorem 2.9 and Lemma 2.11.

Proof. By Lemma 2.8, $g \circ f(S) \cong_{\mathcal{H}} g(F)$. By Theorem 2.9 and transitivity $g(F) \cong_{\mathcal{H}} S$. Consider now S and $g(F)$: by Lemma 2.7, $f(S) \cong_{\mathcal{F}} f \circ g(F)$. But by hypothesis, $f(S) \cong_{\mathcal{F}} F$, therefore $F \cong_{\mathcal{F}} f \circ g(F)$. □

Similarly to the situation on hypercubes, we have the following.

Corollary 2.12. *If F is a frame such that there exists a hypercube H, with $F \cong_{\mathcal{F}} f(H)$, then for any $\phi \in \mathcal{L}$ we have that $F \models \phi$ if and only if $f \circ g(F) \models \phi$.*

Proof. It follows from Lemma 2.11 and Lemma 1.12. □

We now proceed to identify what are these frames for which Lemma 2.11 holds. For example, consider a frame $F = (W, \sim_1, \ldots, \sim_n)$ such that $\bigcap_{i \in A} \sim_i = id_W$. The frame $f \circ g(F)$ is not in general be isomorphic to F. As an example, consider:

$$F = (\{w_1, w_2\}, \{(w_1, w_1), (w_2, w_2)\}, \{(w_1, w_1), (w_2, w_2)\}).$$

We need to restrict our attention to *both* the properties inherited from the mapping from hypercubes.

Theorem 2.13. *If $F = (W, \sim_1, \ldots, \sim_n)$ is a frame such that:*

- $\bigcap_i \sim_i = id_W$,
- $\forall w_1, \ldots, w_n, \exists \overline{w}$ *such that* $\overline{w} \sim_i w_i$, $i = 1, \ldots, n$;

then $F \cong_{\mathcal{F}} f \circ g(F)$.

Proof. Consider the frame $f \circ g(F) = (W/{\sim_1} \times \cdots \times W/{\sim_n}, \sim'_1, \ldots, \sim'_n)$ built according to Definition 2.5 and Definition 2.3. Let now h be a mapping $h : W \to W/{\sim_1} \times \cdots \times W/{\sim_n}$, defined by $h(w) = ([w]_{\sim_1}, \ldots, [w]_{\sim_n})$. We prove that h is a bijection.

Injective: suppose $h(w_1) = h(w_2)$, i.e. $([w_1]_{\sim_1}, \ldots, [w_1]_{\sim_n}) = ([w_2]_{\sim_1}, \ldots, [w_2]_{\sim_n})$. Therefore, for all i, $[w_1] \sim_i [w_2]$, but since $\bigcap_i \sim_i = id_W$, it must be $w_1 = w_2$.

Surjective: consider any element $([w_1]_{\sim_1}, \ldots, [w_n]_{\sim_n})$ in $W/{\sim_1} \times \cdots \times W/{\sim_n}$. By hypothesis on F, there exists a world $\overline{w}$ in W, such that $[\overline{w}]_{\sim_i} = [w_i]_{\sim_i}$, for each $i = 1, \ldots, n$. Therefore $([w_1]_{\sim_1}, \ldots, [w_n]_{\sim_n}) = ([\overline{w}]_{\sim_1}, \ldots, [\overline{w}]_{\sim_n}) = h(\overline{w})$.

Now we prove that $w_1 \sim_i w_2$ in F if and only if $h(w_1) \sim'_i h(w_2)$ in $f \circ g(F)$. Suppose $w_1 \sim_i w_2$, that is $[w_1]_{\sim_i} = [w_2]_{\sim_i}$; by definition of $\sim_i$, this is equivalent to $([w_1]_{\sim_1}, \ldots, [w_1]_{\sim_n}) \sim'_i ([w_2]_{\sim_1}, \ldots, [w_2]_{\sim_n})$.

This proves that F and $f \circ g(F)$ are isomorphic. □

Theorem 2.13 and Lemma 2.4 allow us to characterise the frames that are images of some hypercube. In fact, let us call $\mathcal{G}$ the class of frames that satisfy property 1 and 2 of Lemma 2.4.

Definition 2.14. *Let $\mathcal{G}$ be the class of equivalence frames that satisfy properties:*

1. $\bigcap_{i \in A} \sim_i = id_W$*;*

2. *For any $w_1, \ldots, w_n$ in W there exists a point $\overline{w}$ such that $\overline{w} \sim_i w_i$, $i = 1, \ldots, n$*

We can now prove that:

Theorem 2.15. *Given a frame F, the following are equivalent:*

1. $F \in \mathcal{G}$*,*

2. *there exists a hypercube H, such that $F \cong_{\mathcal{F}} f(H)$.*

Proof. 1 implies 2: Under these conditions by Theorem 2.13, $F \cong_{\mathcal{F}} f \circ g(F)$. That is: $H = g(F)$.

2 implies 1: By Lemma 2.4 the frame $f(H)$ has the properties expressed by proposition 1. But F is isomorphic to $f(H)$ and therefore it has those properties as well. □

Theorem 2.15 characterises the frames that we obtain by applying the map f to the class of hypercubes. Every member of this class of frames is isomorphic to a system and a frame not included in this class is not.

Given Definition 1.7, we can now prove that:

Theorem 2.16. *For all formulae ϕ, $\mathcal{H} \models \phi$ if and only if $\mathcal{G} \models \phi$.*

Proof. From right to left. If $\mathcal{G} \models \phi$, then, since $f(\mathcal{H}) \subseteq \mathcal{G}$, $f(\mathcal{H}) \models \phi$. So, by Definition 1.7 $\mathcal{H} \models \phi$.

From left to right. Assume $\mathcal{H} \models \phi$, i.e. $f(\mathcal{H}) \models \phi$, we want to show that for any $F \in \mathcal{G}$, $F \models \phi$. By Theorem 2.13 and Definition 2.14, $F \cong_{\mathcal{F}} f(g(F))$. But then $F \models \phi$ if and only if $f(g(F)) \models \phi$. But $g(F) \in \mathcal{H}$, and so $f(g(F)) \models \phi$, and so $F \models \phi$. □

We finally have the result we aimed for, i.e. we have found the semantical class that corresponds to hypercube systems.

Corollary 2.17. *The class of hypercube systems and Kripke models built on the class $\mathcal{G}$ of Kripke frames are semantically equivalent.*

Proof. For a contradiction, suppose $\mathcal{HS} \models \phi$ and $\mathcal{M}_{\mathcal{G}} \not\models \phi$, where $\mathcal{M}_{\mathcal{G}}$ is the class of models built on the class $\mathcal{G}$ of frames. By Definition 1.7, $\mathcal{M}_{\mathcal{HS}} \models \phi$, where $\mathcal{M}_{\mathcal{HS}}$ is the class of models generated by $\mathcal{HS}$. But by hypothesis there is a model M built on a frame $F \in \mathcal{G}$ such that $M \not\models \phi$. So, $\mathcal{G} \not\models \phi$. So, by Theorem 2.16 $\mathcal{H} \not\models \phi$. Therefore $\mathcal{HS} \not\models \phi$ which is absurd.

For the other direction, suppose $\mathcal{HS} \not\models \phi$ and $\mathcal{M}_{\mathcal{G}} \models \phi$, where $\mathcal{M}_{\mathcal{G}}$ is the class of models built on the class $\mathcal{G}$ of frames. So there is a hypercube system HS such that the generated model M_{HS} (see Definition 1.6 and Definition 1.7) is such that $M_{HS} \not\models \phi$. Therefore by Definition 1.7 there is a hypercube $H \in \mathcal{H}$ such that $H \not\models \phi$. So by Theorem 2.16 there is a frame $F \in \mathcal{G}$, such that $F \not\models \phi$. Therefore we have $\mathcal{G} \not\models \phi$ and so $\mathcal{M}_{\mathcal{G}} \not\models \phi$, which is contrary to our assumption. □

Theorems 2.15 and 2.16 together with Corollary 2.17 completely characterise hypercube systems.

We can now benefit from this result by axiomatising the class $\mathcal{G}$ of Kripke frames which are technically more explored in the literature. By Corollary 2.17 this will also be an axiomatisation for hypercube systems.

Chapter 3

Axiomatisation of hypercube systems

3.1 Introduction

In Chapter 2 we introduced the class of hypercube systems as an interesting special case of static interpreted systems. We argued that hypercube systems represent MAS in which agents have no information about each other's state. The main result of Section 2 was contained in Theorem 2.16 and Corollary 2.17, in which we proved that hypercubes and hypercube systems are semantically equivalent to special classes of Kripke frames and Kripke models respectively.

We can now take stock of this position and ask the following question. What properties of knowledge do hypercube systems satisfy? Of course we would like to answer this question in a precise way, i.e. by presenting an axiomatisation. Now that we have the counterparts of hypercube systems in the class of Kripke models (i.e. the class of models built on the class $\mathcal{G}$ of frames) we can ask that question in *this* semantic class; indeed this is the topic of this chapter as shown in Figure 3.1. This has the advantage of being a standard axiomatisation problem known in modal logic for which we can use some of techniques presented in Section 1.2.4.

This chapter is organised as follows. In Section 3.2 we will examine the correspondence problem (as defined in Theorem 1.19) for the class of frames $\mathcal{G}$ of Definition 2.14. In Sections 3.3 and 3.4 we will prove completeness for a logic system with respect to this class of frames. In Section 3.5 we will prove decidability. In Section 3.6 we will provide alternative axiomatisations by subsequently proving completeness with respect to three different logics. In Section 3.7 we will use these axiomatisations to understand how hypercubes share part of their knowledge. Finally in Section 3.8 we will report a result that explains the low-level mechanism that produces the results discussed in the rest of the chapter.

Assume a set of agents A. We recall that $\mathcal{G}$ is the class of equivalence frames satisfying the two properties of Definition 2.14 that, independently from equivalence relations, we define as follows:

Definition 3.1 (I). *A frame $F = (W, \{R_i\}_i)$ is an* I frame *(or F has the* identity-intersection *property) if $\bigcap_{i \in A} R_i = id_W$.*

Definition 3.2 (nD). *A frame $F = (W, \{R_i\}_i)$ is* an nD frame *(or F is* n-Directed*) if for all $w_1, \ldots, w_n \in W$ there exists a point $\overline{w} \in W$ such that $w_i R_i \overline{w}$ for $i = 1, \ldots, n$.*

It is immediate to observe the following:

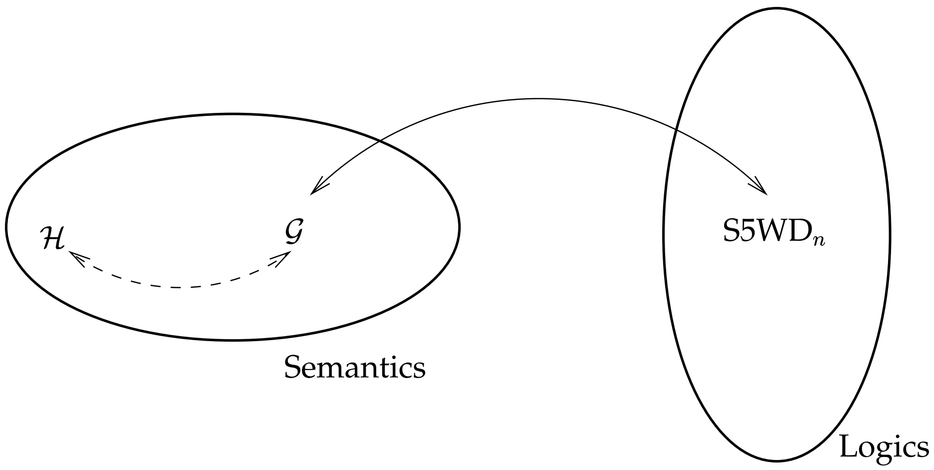

Figure 3.1: The content of the main part of this chapter. The solid arrowed line between $\mathcal{G}$ and the logic S5WD$_n$ represents the completeness result that we will achieve in this chapter. The dashed arrow line between the classes $\mathcal{G}$ and $\mathcal{H}$ represents the equivalence we proved in the previous chapter.

Lemma 3.3. *If a frame is nD, then it is also mD, for any $m \leq n$.*

A few conventions. We sometimes will say that F is a D frame to mean that it is an nD frame, where n is exactly the number of the relations of F. A DI frame is a frame which is both D and I. Now that we have names for the properties of the class of frames $\mathcal{G}$, in the following we will use the more immediate symbol $\mathcal{F}_{EDI}$ to represent $\mathcal{G}$. Since we will only be referring to knowledge, we can also switch back to the usual notation for modal operators, i.e. $\Box_i$ for K_i.

3.2 Correspondences

When facing the problem of axiomatising a class of frames (as in this chapter) it is often useful to try and understand whether the semantic constraints of the frames in study correspond to the validity of certain axiom schemas. So, a promising start would be for us to prove correspondence results of the shape of Theorem 1.19. Should we achieve this, those axioms would be our first bet to prove completeness.

In this case, unfortunately the two semantic properties have no modal correspondences at all.

Lemma 3.4. *If A is not a singleton, then no modal formula corresponds to property I, even in the case of equivalence frames.*

Proof. Suppose the opposite and assume there is a formula ϕ that corresponds to property I, i.e. $F \models \phi$ if and only if F is I. Consider the frame F' in Figure 3.2. The frame F' is an I frame, so $F' \models \phi$. Consider now the frame F and a function $p : F' \to F$ such that p maps points in F according to the names in the Figure. It is easy to see that p is a p-morphism (Definition 1.13) from F' to F. Since p-morphisms preserve validity on frames (as reported in Lemma 1.14), we have that $F \models \phi$. But F is not an I frame and we have a contradiction.

Since F, F' are equivalence frames, the second part follows too. $\Box$

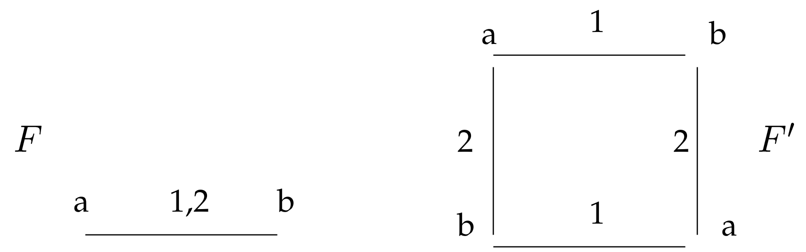

Figure 3.2: Two p-morphic frames used in the proof of Lemma 3.4.

Figure 3.3: Two equivalence (the relations are intended to be reflexive) directed frames such that their disjoint union is not directed as in Lemma 3.5.

A similar result can be proven for property D.

Lemma 3.5. *No modal formula corresponds to n-directedness.*

Proof. Suppose the opposite and assume there is a formula ϕ that corresponds to n-directedness, i.e. $F \models \phi$ if and only if F is D. Consider two disjoint frames, $F = (W, \{R_i\}_i)$ and $F' = (W', \{R'_i\}_i)$, $W \cap W' = \emptyset$, such that both F and F' are n-directed (for example see the two equivalence frames based on two relations of Figure 3.3). Since by assumption $F \models \phi$ and $F' \models \phi$, then $F \cup F' \models \phi$. But, then ϕ is valid on a frame, which, in general, is not n-directed, which is opposite to what we assumed at the beginning. $\square$

As an aside, we note that these results may not hold in some extension of the formal language. For example, consider the operator of distributed knowledge as described in Section 1.3.2.2. Recall that a formula $D_A\phi$ is interpreted by associating the relation $\sim = \bigcap_{i \in A} \sim_i$ to the operator D_A.

Lemma 3.6. *An equivalence frame F is I if and only if $F \models \phi \Leftrightarrow D_A\phi$.*

Proof. Left to right. Let M be a model based on F such that $M \models_w \phi$. Since $\sim = \bigcap_{i \in A} \sim_i = id_W$, then $M \models_w D_A\phi$. Analogously, suppose $M \models_w D_A\phi$. Since $w(\bigcap_{i \in A} \sim_i)w'$ implies $w = w'$, then $M \models_w \phi$.

Right to left. Suppose $F \models \phi \Leftrightarrow D_A\phi$ and that for all i $w_1 \sim_i w_2$. Take a valuation π such that $\pi(p) = \{w_1\}$. Since $(F, \pi) \models_{w_1} p \Leftrightarrow D_A p$ and $(F, \pi) \models_{w_1} p$, we have $(F, \pi) \models_{w_1} D_A p$ and so $F, \pi \models_{w_2} p$. But since $\pi(p) = \{w_1\}$, it must be that $w_1 = w_2$. $\square$

This lemma is a quite surprising result: in hypercube systems the notion of distributed knowledge collapses to the truth of the formula. Note that this relies on the simplification to consider global states that do not represent the environment. Notwithstanding this, the same result would have been achieved had we considered a MAS whose environment is constant, i.e. set of states for the environment composed by a singleton. We now return to the basic language defined in Section 1.2.1 without the operator D_A.

So far, we proved that there is no correspondence to be found neither on D frames nor on I frames. Note that this also applies to the intersection of these two classes, i.e. to DI frames. In fact the proof of Lemma 3.4 can still be used in this case, as the frame F' in Figure 3.2 is actually a DI frame, while the frame F is not (quite simply because it is not an I frame). In other words, we have the following.

Corollary 3.7. *No modal formula corresponds to property D and I, even in the case of equivalence frames.*

This might mean that equivalence DI frames can be axiomatised quite simply by S5$_n$, that would imply that the modal syntax we are using is unable to distinguish between e-quivalence and equivalence DI frames. In turn, this would imply that the properties of knowledge that hypercubes enjoy are nothing more than the ones of ideal agents (discussed on page 1.3.1.1).

We can prove that the above is not the case, as directed frames validate more formulae than just S5$_n$. To see that, consider the following:

Lemma 3.8. *If F is an equivalence D frame, then $F \models \Diamond_i \Box_j \phi \Rightarrow \Box_j \Diamond_i \phi$, where $i, j \in A, A = \{1, \ldots, n\}, i \neq j, n \geq 2$.*[1]

Proof. For a contradiction suppose that $F \not\models \Diamond_i \Box_j \phi \Rightarrow \Box_j \Diamond_i \phi$. Then there exists a point w and a valuation π such that $(F, \pi) \models_w \Diamond_i \Box_j \phi \wedge \neg \Box_j \Diamond_i \phi$. Therefore there must exist two points w_1 and w_2 such that $w \sim_i w_1$ and $w \sim_j w_2$ and $(F, \pi) \models_{w_1} \Box_j \phi$ and $(F, \pi) \models_{w_2} \Box_i \neg \phi$. But by Lemma 3.3 there exists a point $\overline{w}$ such that $\overline{w} \sim_j w_1$ and $\overline{w} \sim_i w_2$. Since $(F, \pi) \models_{w_1} \Box_j \phi$, we have $(F, \pi) \models_{\overline{w}} \phi$, but this contradicts $(F, \pi) \models_{w_2} \Box_i \neg \phi$ that requires $\overline{w}$ to satisfy $\neg \phi$. $\Box$

It is easy to check that the axiom in Lemma 3.8 is not generally valid on the class of equiv-alence frames. In Figure 3.4, the model M_2 does not satisfy in the point w_0 the formula $\Diamond_1 \Box_2 p \Rightarrow \Box_2 \Diamond_1 p$. Indeed we can also notice that also the formula of Lemma 3.6 is also not valid in general on equivalence frames (see the point w_0 in the model M_1 shown in Fig-ure 3.4).

So, agents modelled by semantic structures that have properties D satisfy the axiom of Lemma 3.8. This axiom says that if the agents described by hypercubes have the property that if agent i considers possible that agent j knows ϕ, than agent j knows that agent i considers ϕ to be possible. This is a constraint on the agents' knowledge because it implies that two agents i and j cannot be in a situation in which i considers that j might know a fact and j considers that i might know the negation of the same fact.

We are now in the position in which no correspondence can be found for property D (Lemma 3.5), still we know that D frames satisfy some formulae that are not generally valid on equivalence frames (Lemma 3.8). We need to look at a property weaker than D.

[1] In the proof of Lemma 3.8 we do not use the assumption that the frame is an equivalence frame. Indeed the lemma holds even without this assumption and property D imposes the frame to validate the axiom in Lemma 3.8. In order not to interrupt the flow of the analysis of equivalence DI frames I find it more appropriate to present it in this way. For equivalence frames the cases $i = j$ or $n = 1$ are already theorems of S5$_n$ because in that case the axiom is equivalent via Lemma 1.30 to the schema $\Box_i \phi \Rightarrow \Diamond_i \phi$ which holds because of axiom T.

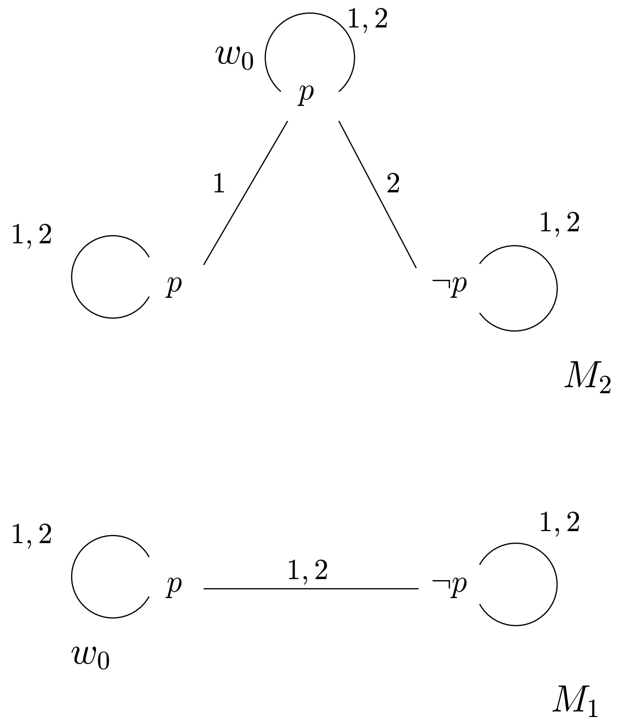

Figure 3.4: Equivalence models not satisfying formulae in Lemma 3.6 and Lemma 3.8.

Definition 3.9 (nWD). *Let P_n be the set of all the permutations of $\{1, \dots, n\}$ without fixed-points, i.e. if $(x_1, \dots, x_n) \in P_n$, then $x_i \neq i$.*

A frame $F = (W, \{R_i\}_i)$ is n-weakly-directed *(nWD) if for all points $w, w_1, \dots, w_n$ in W, such that $wR_iw_i, i = 1, \dots, n$ and for all $(x_1, \dots, x_n)$ in P_n there exists a point $\overline{w}$ such that $w_iR_{x_i}\overline{w}$, for all $i = 1, \dots, n$.*

See Figure 3.5. When n is clear from the context we just refer to nWD just as WD. The property nWD for $n = 2$ (sometimes called "convergence" or "Church-Rosser") is discussed in, among others, [Pop94], [Cat88] and [Bal98]; nWD is a generalisation of it[2]. Note that 2WD resembles the property of convergence for the mono-modal case; nWD is a generalisation of it both with respect of the number of points w_0 departs to and with respect to the labels.

Indeed WD is weaker than D.

Lemma 3.10. *If a frame is directed then it is weakly-directed.*

Property WD is particularly interesting because, differently from property D, this has a well-defined correspondence. In fact, consider the axiom[3]:

$$\bigwedge_{(x_1,\dots,x_n)\in P_n} (\Diamond_1 \Box_{x_1} p_1 \wedge \dots \wedge \Diamond_{n-1} \Box_{x_{n-1}} p_{n-1}) \Rightarrow \Box_n \Diamond_{x_n} (\wedge_{i=1}^{n-1} p_i) \qquad \textbf{WD}$$

We have the correspondence result:

[2]We will discuss later in Section 3.6 that the restriction on P_n can be lifted. For the moment it is easier to introduce it as generalisation of the known properties above.

[3]Note that for $n = 2$ the axiom **WD** expresses the formula in Lemma 3.8.

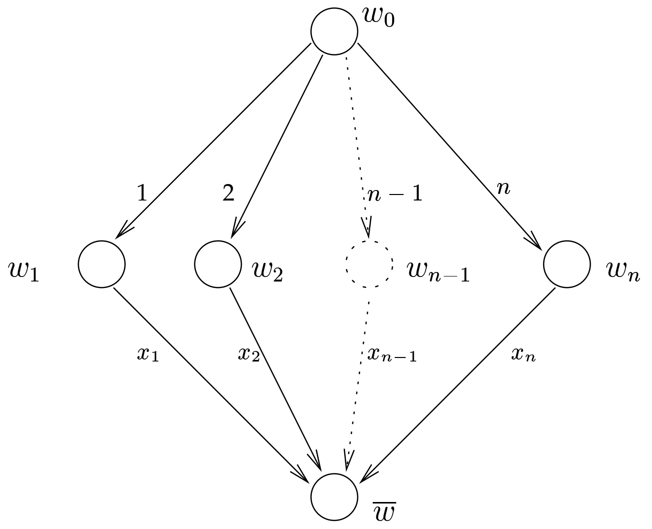

Figure 3.5: The property nWD.

Lemma 3.11. $F \models \mathbf{WD}$ *if and only if* F *is weakly-directed.*

Proof. Suppose $F \models \mathbf{WD}$ and consider $n+1$ points $w, w_1, \ldots, w_n$ of F, such that $wR_iw_i, i = 1, \ldots, n$. Fix a permutation $(x_1, \ldots, x_n) \in P_n$ and consider a valuation π, such that $\pi(p_i) = \{v : w_iR_{x_i}v\}$ for $i = 1, \ldots, n-1$. By construction we have

$$(F, \pi) \models_w \Diamond_1 \Box_{x_1} p_1 \wedge \cdots \wedge \Diamond_{n-1} \Box_{x_{n-1}} p_{n-1}.$$

Then by **WD**, $(F, \pi) \models_w \Box_n \Diamond_{x_n} (\wedge_{i=1}^{n-1} p_i)$, so $(F, \pi) \models_{w_n} \Diamond_{x_n} (\wedge_{i=1}^{n-1} p_i)$. So there is a world $\overline{w}$, such that $(F, \pi) \models_{\overline{w}} \wedge_{i=1}^{n-1} p_i$. But, by construction of the interpretation π, this implies $w_iR_{x_i}\overline{w}$, for $i = 1, \ldots, n$.

For the reverse, consider a permutation $(x_1, \ldots, x_n) \in P_n$, a model (F, π), and a point w, such that $(F, \pi) \models_w \Diamond_1 \Box_{x_1} p_1 \wedge \cdots \wedge \Diamond_{n-1} \Box_{x_{n-1}} p_{n-1}$. We have $(F, \pi) \models_{w_i} \Box_{x_i} p_i$, where $wR_iw_i, i = 1, \ldots, n-1$. We want to prove $(F, \pi) \models_w \Box_n \Diamond_{x_n} (\wedge_{i=1}^{n-1} p_i)$, i.e. that for any point w_n, such that wR_nw_n, $(F, \pi) \models_{w_n} \Diamond_{x_n} (\wedge_{i=1}^{n-1} p_i)$. But, since the frame F is weakly-directed, there exists a point $\overline{w}$, such that $w_iR_{x_i}\overline{w}$, for $i = 1, \ldots, n$. But then $(F, \pi) \models_{\overline{w}} p_i$, for $i = 1, \ldots n-1$, that is $(F, \pi) \models_{\overline{w}} (\wedge_{i=1}^{n-1} p_i)$. So $(F, \pi) \models_{w_n} \Diamond_{x_n} (\wedge_{i=1}^{n-1} p_i)$. □

3.3 Completeness of equivalence D frames

In the previous section we have proved that properties D and I have no modal correspondences. To prove completeness with respect to equivalence DI frames we will proceed as follows.

We first prove completeness of the logic[4] $\mathrm{S5WD}_n = \mathrm{S5}_n + \mathbf{WD}$ with respect to equivalence WD frames (Theorem 3.12). This and Lemma 3.13 will let us prove Theorem 3.14 that expresses completeness of the logic $\mathrm{S5WD}_n$ with respect to equivalence D frames.

[4]This notation was introduced on page 16.

We start by showing that the logic $\mathrm{S5WD}_n$ is complete with respect to equivalence WD frames[5].

Theorem 3.12. *The logic $S5WD_n$ is sound and complete with respect to the class of equivalence WD frames.*

Proof. By Theorem 1.24 the logic $\mathrm{S5}_n$ is sound with respect to equivalence frames. By the second part of Lemma 3.11 we also know that **WD** is valid on the class of WD frames. From these two facts soundness follows.

To prove completeness we use the canonical model technique. We know from Theorem 1.24 that $\mathrm{S5}_n$ is complete with respect to equivalence frames. In particular $\mathrm{S5}_n$ is canonical, i.e. the canonical frame for $\mathrm{S5}_n$ is an equivalence frame. By using arguments employed in the literature (for example [HC96], pages 119-121, it is therefore easy to prove that the frame $F_C^{\mathrm{S5WD}_n} = (W, \{\sim_i\}_i)$ of the canonical model for $\mathrm{S5WD}_n$ is an equivalence frame. We show that the frame $F_C^{\mathrm{S5WD}_n}$ is also WD. Consider $n+1$ maximal $\mathrm{S5WD}_n$-consistent sets, $w, w_1, \dots, w_n$, such that $w \sim_i w_i$, for $i = 1, \dots, n$, and any permutation $(x_1, \dots, x_n) \in P_n$, we want to prove the existence of a point $\overline{w}$ such that $w_i \sim_{x_i} \overline{w}$, for $i = 1, \dots, n$. By definition of the accessibility relations on the frame of the canonical model (Definition 1.21), we only need to prove that the set

$$\Gamma = \bigcup_{i=1}^{n} \{\phi : \Box_{x_i}\phi \in w_i\}$$

is $\mathrm{S5WD}_n$-consistent (since, by the maximal extension lemma (Lemma 1.10) there is a maximal extension $\overline{w}$, which is $\mathrm{S5WD}_n$-consistent and therefore, the frame is WD). Suppose Γ is not $\mathrm{S5WD}_n$-consistent, then there are $\alpha_1^1, \dots, \alpha_{m_1}^1, \dots, \alpha_1^n, \dots, \alpha_{m_n}^n$, with $\Box_{x_1}\alpha_1^1 \in w_1, \dots, \Box_{x_1}\alpha_{m_1}^1 \in w_1, \dots, \Box_{x_n}\alpha_1^n \in w_n, \dots, \Box_{x_n}\alpha_{m_n}^n \in w_n$, such that

$$\vdash_{\mathrm{S5WD}_n} \neg(\alpha_1^1 \wedge \dots \wedge \alpha_{m_1}^1 \wedge \dots \wedge \alpha_1^n \wedge \dots \wedge \alpha_{m_n}^n)$$

Let us now call $\alpha_j = \bigwedge_{i=1}^{m_j} \alpha_i^j$. We have:

$$\vdash_{\mathrm{S5WD}_n} \bigwedge_{k=1}^{n-1} \alpha_k \Rightarrow \neg\alpha_n \tag{3.1}$$

which, by taking contrapositives and by necessitating by $\Box_{x_n}$ becomes:

$$\vdash_{\mathrm{S5WD}_n} \Diamond_{x_n}(\bigwedge_{k=1}^{n-1} \alpha_k) \Rightarrow \neg\Box_{x_n}\alpha_n. \tag{3.2}$$

Observe now that since $\Box_{x_1}\alpha_1^1, \dots, \Box_{x_j}\alpha_{m_1}^1$ are in w_1, then $\Box_{x_1}(\bigwedge_{i=1}^{m_1} \alpha_i^1)$ is in w_1, i.e. $\Box_{x_1}\alpha_1$ is in w_1; in general

$$\Box_{x_j}\alpha_j \in w_j. \tag{3.3}$$

Then by construction we have $\Diamond_i\Box_{x_i}\alpha_i \in w$, for $i = 1, \dots, n$. So, since w is $\mathrm{S5WD}_n$-maximal consistent, the axiom **WD** is valid and so $\Box_n\Diamond_{x_n}(\wedge_{i=1}^{n-1}\alpha_i) \in w$. So, $\Diamond_{x_n}(\wedge_{i=1}^{n-1}\alpha_i) \in w_n$. So,

[5] The same result would follow by using Sahlqvist theorem [Sah75] and computing the resulting first-order condition by using Kracht's theorem [Kra91, Kra93]. Here we rely on more traditional methods.

by maximal consistency and Equation 3.2, $\neg\Box_{x_n}\alpha_n \in w_n$, which by Equation 3.3, implies that w_n is inconsistent, contrary to the assumption. Therefore Γ cannot be inconsistent and in the frame of the canonical model there must exist a point $\overline{w}$ which is a S5WD$_n$-maximal extension of Γ. Therefore the frame $F_C^{\mathrm{S5WD}_n}$ is nWD. $\Box$

We now strengthen the result above by showing that S5WD$_n$ is sound and complete with respect to equivalence and directed frames.

Before we can prove this result we need the following lemma.

Lemma 3.13. *If F is an equivalence, n-weakly-directed and connected frame than it is n-directed.*

Proof. In the following we assume $n \geq 3$. If $n = 2$ the proof is considerably simpler and we do not report it here. So, assume F connected, n-weakly-directed and made by equivalence relations. To prove that F is directed we only need to prove that considering n points, $w_1, \ldots, w_n$ in F, there is a point $\overline{w}$ such that $w_i \sim_i \overline{w}, i = 1, \ldots, n$.

a) Since F is connected, reflexive, symmetric and transitive, then for every pair of points x, y, there is chain that connects them with no repeated links: $x \sim_{z_1} t_1 \cdots \sim_{z_k} t_k \sim_{z_k+1} y$, where $z_i \neq z_{i+1}$.

b) We prove that every chain connecting two points can be replaced by a chain of length 2 in which one of the relations can be chosen arbitrarily. Consider the chain $x \sim_{z_1} t_1, \ldots, t_k \sim_{z_k} y$, connecting x to y; we want to reduce it to a chain of length 2, in which x is connected by $\sim_1$. Consider $x \sim_{z_1} t_1$, and $n-1$ reflexive relations on x. By WD there exists a point v_1 such that $x \sim_1 v_1, t_1 \sim_{y_1} v_1$, where $y_1 \neq 1$. Consider now $x \sim_1 v_1, t_1 \sim_{z_2} t_2$. If $y_1 = z_2$ we can apply transitivity and apply what follows to v_1, t_3. So, consider $y_1 \neq z_2$, then by WD there exists a point v_2, such that $v_1 \sim_1 v_2, v_2 \sim_{y_2} t_2$, where $y_2 \neq 1$; by transitivity $x \sim_1 v_2$. After at most $k+1$ similar steps we have $x \sim_1 v_{k+1}, v_{k+1} \sim_{y_{k+1}} y$, where $y_{k+1} \neq 1$.

c) We now prove that given any two points x, y we can find a point that connects them by two different arbitrary relations. Consider x and y, we want to connect them via $\sim_1$ and $\sim_2$. By point b) there exist a point v such that $x \sim_3 v, v \sim_k y$, in which $k \neq 3$. Suppose $k \neq 2$, and apply WD to this triple by considering $n-2$ relations on v: there exists a point u_1 such that $x \sim_1 u_1, y \sim_2 u_1$. If $k = 2$ we still have the result by constructing u_1 in the same way and applying transitivity to $u_1 \sim_2 v, v \sim_2 y$.

d) Consider now the points $w_1, \ldots, w_n$, by b) and c) we have the existence of $n-1$ points, $u_1, \ldots, u_{n-1}$, such that $w_i \sim_i u_i, u_i \sim_{i+1} w_{i+1}$, for $i = 1, \ldots, n-1$. So, by transitivity, $u_i \sim_{i+1} u_{i+1}$. Consider now $u_1 \sim_2 u_2$ and the reflexive relations on u_1. By WD there exists a point v_1, such that $u_2 \sim_3 v_1$ and $u_1 \sim_j v_1$, with $j = 1, 2, 4, \ldots, n$. Note that, by transitivity, $v_1 \sim_1 w_1, v_1 \sim_2 w_2, v_1 \sim_3 w_3$. By transitivity we also have $v_1 \sim_3 u_3$; consider now $v_1 \sim_3 u_3$, by applying WD we have the existence of a point v_2 such that $u_3 \sim_4 v_2$, and connected to v_1 by all the equivalence relations but $\sim_4$. Then $v_2 \sim_j w_j$, for $j = 1, 2, 3, 4$. Continuing the construction throughout the chain $u_1, \ldots u_{n-1}$, we identify points v_j, such that $v_j \sim_i w_i$, for $i = 1, \ldots, j+2$. The point v_{n-2} is the point $\overline{w}$, that we are interested in. In fact we have $\overline{w} \sim_i w_i, i = 1, \ldots, n$. This proves that the frame F is directed. $\Box$

Lemma 3.13 allows us to prove that S5WD$_n$ is complete with respect to equivalence D frames.

Theorem 3.14. *The logic* S5WD$_n$ *is sound and complete with respect to the class of equivalence directed frames.*

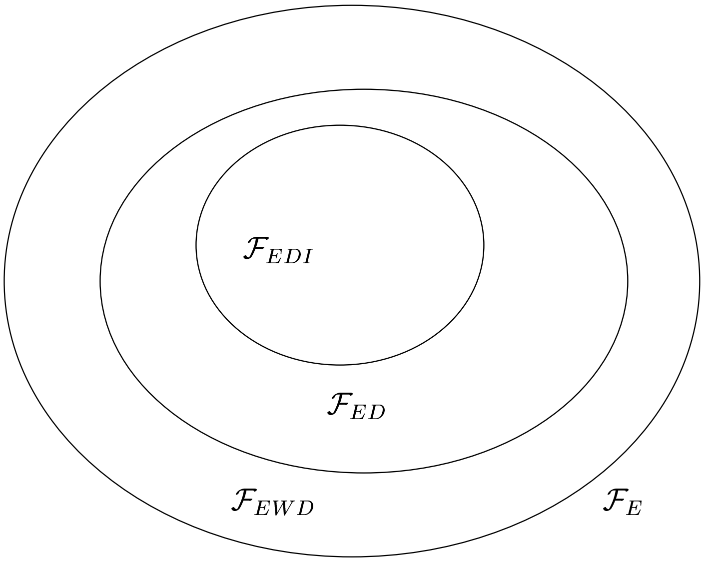

Figure 3.6: A representation of the class of equivalence WD, D and DI frames.

Proof. Soundness follows straightforwardly by considering Lemma 3.10.

For completeness, it is sufficient to show that if a formula ϕ is not a theorem of S5WD$_n$ then it is not valid on a reflexive, symmetric, transitive, and directed frame. So, suppose $\not\vdash_{\mathrm{S5WD}_n} \phi$, then it has to exist a point $\overline{w}$ in the canonical model $M_C^{\mathrm{S5WD}_n}$, such that $M_C^{\mathrm{S5WD}_n} \not\models_{\overline{w}} \phi$. Consider now the model $M_{\overline{w}}$ generated by $\overline{w}$ from $M_C^{\mathrm{S5WD}_n}$ (Definition 1.17). By Lemma 1.18 we have $M_{\overline{w}} \not\models_{\overline{w}} \phi$ and, by Theorem 3.12 the frame of $M_{\overline{w}}$ is reflexive, symmetric, transitive and weakly-directed. But, since $M_{\overline{w}}$ is connected by construction, by Lemma 3.13 the frame is also directed and so ϕ is not valid on the class of reflexive, symmetric, transitive and directed frames. Therefore S5WD$_n$ is complete with respect to this class of frames. $\square$

Theorem 3.14 is our first important result in order to achieve completeness of the class $\mathcal{F}_{EDI}$ of frames.

3.4 Completeness of equivalence DI frames

Section 3.3 ends with an important result: the logic S5WD$_n$ is complete with respect to equivalence directed frames, that we call ED frames from now on. This is somewhat surprising because in the same section we proved that the logic S5WD$_n$ is also sound and complete with respect to the bigger class of equivalence WD frames (or EWD frames). In this section we will push this result even further by proving that the very same logic is sound and complete with respect to equivalence DI (or EDI) frames as well. Indeed this is an interesting example of the modal semantics being able to express much more than what we can axiomatise by using modal operators. Figure 3.6 shows the classes of WD, D and DI frames in the class of equivalence frames. Note that in terms of static interpreted systems the difference between DI and D frames can be seen as whether in the tuples the environment is constant or not (see Figure 2.2 and the observations we drew from it).

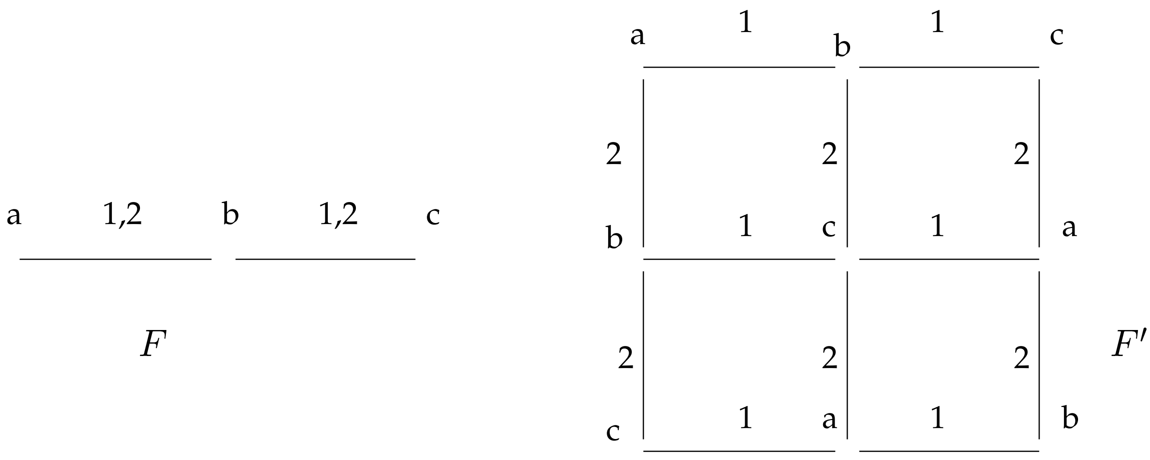

Figure 3.7: An EDI frame mapping an ED frame via a p-morphism (the relations are the reflexive and transitive closure of the ones illustrated).

In order to prove that $S5WD_n$ is sound and complete with respect to equivalence DI frames, we show that the class of equivalence D and equivalence DI frames are semantically equivalent. To do so, in turn, we prove that any D frame can be seen as the target of a p-morphism from a DI frame; the result will then follow in view of the fact that p-morphisms between frames preserve validity and that DI frames are special D frames.

Consider any D frame defined on n equivalence relations on its support set W. Write $\sim$ for the relation $\bigcap_{i=1\dots n} \sim_i$; since each of the $\sim_i$ is an equivalence relation, so is $\sim$. The frame F can then be viewed as the union of equivalence classes of the relation $\sim$, which we call *clusters*. Clusters containing more than a single point are sub-frames in which property I clearly does not hold; in general a cluster may be infinite in size.

If we want to construct a DI frame that maps to a particular D frame by a p-morphism, we can think of replacing every cluster of the D frame with a sub-frame that is DI but that can still be mapped into the cluster. In Figure 3.7 it is shown the relatively simple case of a frame F composed by three points a, b, c connected by all the relations: $\sim_1, \sim_2$, in this case; F clearly is D but not I. The frame F' on the right of the figure is a DI frame; the names of its points represent the targets of the p-morphism from F' onto F. So, for example the top left point of F' is mapped onto a of F; the relations are mapped in the intuitive way. It is an easy exercise to show that F is indeed a p-morphic image of F' and will therefore validate every formula which is valid on F'.

The aim of the following is to define precisely how to build, given any D frame, a new DI frame in which every cluster is "unpacked" into an appropriate structure so that a p-morphism between the two structures can be defined (see Figure 3.8).

In order to achieve the above, we present two set theoretic results. In Lemma 3.15 we show that every infinite set X can be seen as the image of a product X^n under a function p. Intuitively this lemma will be used by taking the set X as one of the clusters of a DI frame F, the function p as the p-morphism and the product X^n (where n is the number of relations on the frame) as the sub-frame that will replace the cluster in the new frame F'. Lemma 3.16 extends the result of Lemma 3.15 to guarantee that even if the clusters differ in size it is always possible to find a single sub-frame that can replace each of them. The two lemmas

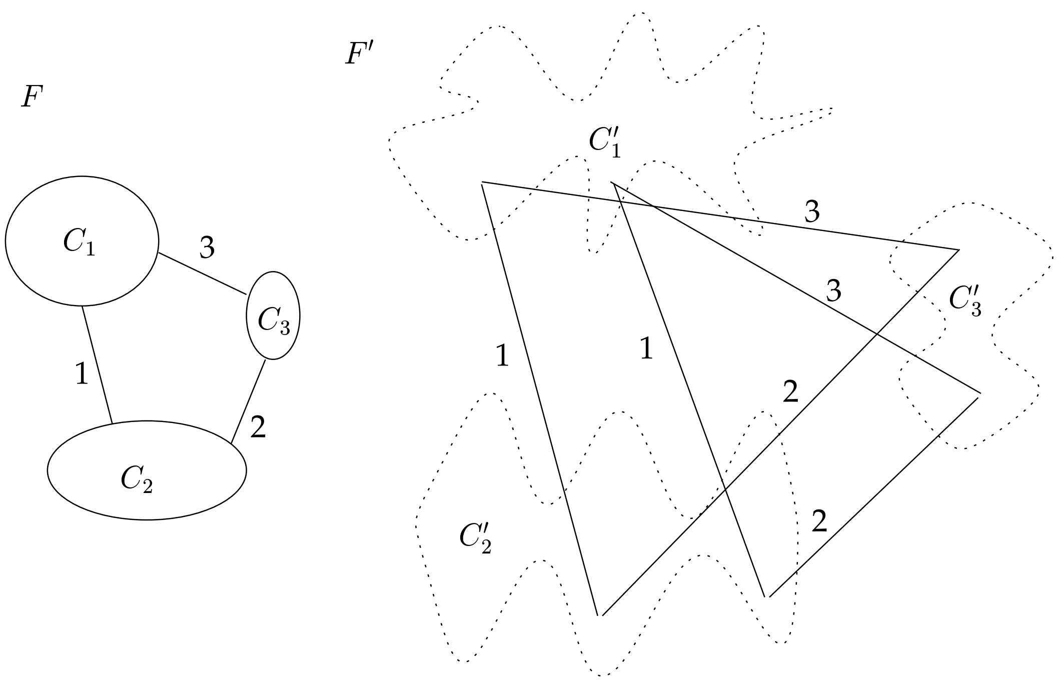

Figure 3.8: A visualisation of the process of cluster explosion of Theorem 3.18, Lemma 3.15 and Lemma 3.16. Given an equivalence frame D frame F that contains clusters $C_1, C_2, \ldots$ of points that do no satisfy property I, a new frame F' that maps F by a p-morphism is defined. In the frame F' the clusters have been properly exploded into sub-frames that enjoy property I while still being D frames. Theorem 3.18 proves constructively how to do so in a way that ensures that the frame F' as a whole is an equivalence DI frame.

will then be used in Theorem 3.18 to show that a p-morphism with these properties can be defined.

Lemma 3.15. *Given any infinite set X, there exists a function $p : X^n \to X$ such that the following holds.*

Let $i \in \{1, \dots, n\}$. For all $u, x_i \in X$, there are $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n, \in X$, such that $p(x_1, \dots, x_n) = u$.

Proof. Consider the set $T = \{\tau_{x,y} \mid x, y \in X\}$ of the transpositions of X, i.e. functions $\tau_{x,y} : X \to X$; where $x, y \in X$, and such that $\tau_{x,y}(z) = y$ if $z = x$; $\tau_{x,y}(z) = x$ if $z = x$; $\tau_{x,y}(z) = z$ otherwise. We have $|X| \leq |T| \leq |X \times X|$. But by set theory ([Lan84] page 701 for example) $|X| = |X \times X|$, and so $|X| = |T|$. So, by induction, we have $|X^{n-1}| = |X| = |T|$. Call f the bijection $f : X^{n-1} \to T$, and define $p(x_1, \dots, x_n) = f(x_1, \dots, x_{n-1})(x_n)$. To prove the lemma holds we consider two cases: $i \neq n$ and $i = n$.

For $i \neq n$, assume any $u \in X$, and any $x_i \in X$. Take any $x_j, j \in \{1 \dots n-1\} \setminus \{i\}$; $f(x_1, \dots, x_{n-1})$ is a transposition of X. So, there exists an $x_n \in X$ such that $f(x_1, \dots, x_{n-1})$ $(x_n) = u$. So $p(x_1, \dots, x_n) = u$.

For $i = n$, assume again any $u \in X$, and any $x_n \in X$. Consider the transposition $\tau_{x_n,u}$; we have $\tau_{x_n,u}(x_n) = u$. But $\tau_{x_n,u} = f(x_1, \dots, x_{n-1})$ for some $x_1, \dots, x_{n-1} \in X$. So $p(x_1, \dots, x_n) = u$. ☐

Lemma 3.15 induces a similar result for sets whose cardinality is smaller than X.

Lemma 3.16. *Given any infinite set X, and a set $C \neq \emptyset$, such that $|C| \leq |X|$, there exists a function $p : X^n \to C$ such that the following holds.*
Let $i \in \{1, \dots, n\}$. For all $x_i \in X, u \in C$, there are $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n, \in X$, such that $p(x_1, \dots, x_n) = u$.

Proof. Consider a set T such that $C \cup T$ and X have the same cardinality, and let g be a bijection from X to $(C \cup T)$. Then there is a function $p' : (C \cup T)^n \to (C \cup T)$, satisfying the property expressed by Lemma 3.15. Define now a function $p'' : (C \cup T) \to C$, such that $p''(x) = x$ if $x \in C$, otherwise $p''(x) = c$, where c is any element in C. Define the function $p : X^n \to C$ by $p(x_1, \dots, x_n) = p''(p'(g(x_1), \dots, g(x_n)))$. The following shows that p has the property required. For, let $i \in \{1, \dots, n\}$ and take any $x_i \in X$ and $u \in C$. Then $g(x_i) \in (C \cup T)$, and so by Lemma 3.15 there exist $c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_n \in C \cup T$, such that $p'(c_1, \dots, c_{i-1}, g(x_i), c_{i+1}, \dots, c_n) = u$. Define $x_j = g^{-1}(c_j)$ for $j \in \{1, \dots, n\} \setminus \{i\}$. We then have $p(x_1, \dots, x_n) = p''(p'(c_1, \dots, c_{i-1}, g(x_i), c_{i+1}, \dots, c_n)) = p''(u) = u$ since $u \in C$. ☐

We rely on the two results above to define a function p that maps tuples $\langle c, x_1, \dots, x_n \rangle$ into c, where c is a cluster and $x_i \in X$, for some appropriate set X. The function p is defined as in Lemma 3.16 but it has an extra component for the cluster.

Corollary 3.17. *Let $\mathcal{C}$ be a set of nonempty subsets of a set W. Then there exists a set X and a function $p : \mathcal{C} \times X^n \to W$ such that*

1. *for all tuples $\langle c, x_1, \dots, x_n \rangle$ we have $p(\langle c, x_1, \dots, x_n \rangle) \in c$, and*

2. *for all $c \in \mathcal{C}$, for all $u \in c$, for all $i = 1 \dots n$, and for all $x_i \in X$, for each $j \in \{1 \dots n\} \setminus \{i\}$ there exists $x_j \in X$, such that $p(\langle c, x_1 \dots x_n \rangle) = u$.*

Proof. Let X be an infinite set with cardinality at least as great as the cardinality of any $c \in \mathcal{C}$. This can be constructed by taking the union of these sets $c \in \mathcal{C}$ or by considering the set of the natural numbers $X = \mathbb{N}$ if all the sets $c \in \mathcal{C}$ are finite. For each $c \in \mathcal{C}$, let $p_c : X^n \to c$ be the function promised by Lemma 3.16. Define $p : \mathcal{C} \times X^n \to W$ by $p(c, x_1, \ldots, x_n) = p_c(x_1, \ldots, x_n)$. It is immediate that this function has the required property. □

Theorem 3.18. *Given any equivalence D frame F, there exists an equivalence DI frame F', and a p-morphism p, such that $p(F') = F$.*

Proof. Let $F = (W, \sim_1, \ldots, \sim_n)$ be a frame with n relations on its support set W. Write $\sim$ for the relation $\bigcap_{i=1\ldots n} \sim_i$. Since each of the $\sim_i$ is an equivalence relation, so is $\sim$. Since the set of worlds W of the frame F is non-empty, it can be viewed as the union of the equivalence classes of the relation $\sim$, which we call *clusters*. Call $\mathcal{C}$ the set of clusters of F. Consider the infinite set X and a function p as described in Corollary 3.17, and define the frame $F' = (W', \sim'_1, \ldots, \sim'_n)$ as follows:

- $W' = \mathcal{C} \times X^n$,
- $\langle c, x_1, \ldots, x_n\rangle \sim'_i \langle d, y_1, \ldots, y_n\rangle$ if $x_i = y_i$ and there exists worlds $u \in c$ and $v \in d$ such that $u \sim_i v$.

We can prove that:

1. The frame F' is an equivalence DI frame.

   Proof. a) F' is clearly an equivalence frame.

   b) We prove F' satisfies property I. Write $\sim'$ for $\bigcap_{i=1\ldots n} \sim'_i$. Suppose $\langle c, x_1, \ldots, x_n\rangle \sim' \langle d, y_1, \ldots, y_n\rangle$. Then for all $i = 1 \ldots n$ we have that $x_i = y_i$, and there exist $u_i \in c$ and $v_i \in d$ such that $u_i \sim_i v_i$. Since c and d are equivalence classes of $\sim$, it follows from the latter that $u_1 \sim v_1$, and consequently that $c = d$. Thus, $\langle c, x_1, \ldots, x_n\rangle = \langle d, y_1, \ldots, y_n\rangle$.

   c) We prove F' satisfies property D. Consider n tuples $\langle c_1, x_1^1, \ldots, x_n^1\rangle, \ldots, \langle c_n, x_1^n, \ldots, x_n^n\rangle$ in W'. For each $i = i \ldots n$ let u_i be a world in cluster c_i. Since F has property D, there exists a world w such that $w \sim_i u_i$ for each $i = 1 \ldots n$. Let c be the cluster containing w. Then, by construction, for each $i = 1 \ldots n$ we have $\langle c, x_1^1, \ldots, x_n^n\rangle \sim'_i \langle c_i, x_1^i, \ldots, x_n^i\rangle$. □

2. The function p is a p-morphism from F' to F.

   Proof. That the function p is surjective follows from property (2) of Corollary 3.17.

   Next, we show that p is a frame p-morphism (i.e. it satisfies properties (2) and (3) of Definition 1.13). Consider two tuples $\langle c, x_1, \ldots, x_n\rangle$, $\langle d, y_1, \ldots, y_n\rangle$ in W' such that $\langle c, x_1, \ldots, x_n\rangle \sim'_i \langle d, y_1, \ldots, y_n\rangle$. Then there exists $u \in c$ and $v \in d$ such that $u \sim_i v$. By property (1) of Corollary 3.17, we have $p(\langle c, x_1, \ldots, x_n\rangle) \sim_i u$ and $p(\langle d, y_1, \ldots, y_n\rangle) \sim_i v$. Since $\sim_i$ is an equivalence relation, it follows that $p(\langle c, x_1, \ldots, x_n\rangle) \sim_i p(\langle d, y_1, \ldots, y_n\rangle)$.

   To show the backward simulation property, consider a tuple $\mathbf{x} = \langle c, x_1, \ldots, x_n\rangle$, and assume $p(\mathbf{x}) \sim_i w$ for some world w of F. Let d be the cluster containing w. By Corollary 3.17(2), there exist y_j for $j \neq i$ such that if $\mathbf{y} = \langle d, y_1, \ldots y_{i-1}, x_i, y_{i+1}, \ldots y_n\rangle$, then $p(\mathbf{y}) = w$. Since $p(\mathbf{x}) \in c$ by Corollary 3.17(1), it is immediate that $\mathbf{x} \sim'_i \mathbf{y}$. □

□

By defining a p-morphism between the two classes of frames we can prove their semantical equivalence.

We call $\mathcal{F}_{ED}$ the class of equivalence D frames; recall that $\mathcal{F}_{EDI}$ denotes the class of equivalence DI frames.

Theorem 3.19. *For any formula ϕ, $\mathcal{F}_{ED} \models \phi$ if and only if $\mathcal{F}_{EDI} \models \phi$.*

Proof. From left to right. Clearly $\mathcal{F}_{EDI} \subseteq \mathcal{F}_{ED}$ so if a formula ϕ is such that $\mathcal{F}_{ED} \models \phi$, then $\mathcal{F}_{EDI} \models \phi$.

From right to left. Consider a formula ϕ, by contradiction suppose $\mathcal{F}_{ED} \not\models \phi$. Then there exists a D frame F such that $F \not\models \phi$. But by Theorem 3.18 there exists a DI frame F' such that F is a p-morphic image of F. Since F is a p-morphic image we have that for any ψ such that $F' \models \psi$ implies $F \models \psi$. So we have that $F' \not\models \phi$, so $\mathcal{F}_{EDI} \not\models \phi$, which is what we needed to prove. □

Corollary 3.20. *The logic S5WD$_n$ is sound and complete with respect to the class of hypercube systems.*

Proof. From Theorem 3.14, the logic S5WD$_n$ is complete with respect to equivalence D frames but by Theorem 3.19 it follows that this logic is complete with respect to equivalence DI frames. But by Theorem 2.17 equivalence DI frames are semantically equivalent to hypercubes, and therefore the result follows. □

Corollary 3.20 expresses the axiomatisation of hypercube systems that we aimed for. It is now clear that the hypercube systems model a special class of ideal agents of knowledge. Still, they not only satisfy all the properties of ideal agents of knowledge discussed in Section 1.3.1.1, but also satisfy the property represented by axiom **WD**.

We will discuss how hypercube systems share knowledge in Section 3.6; before doing that we will prove decidability.

3.5 Decidability

Recall from Section 1.2.4, Definition 1.25 of decidability. As we saw in Theorem 1.26 a sufficient condition for proving decidability is that the logic has the finite model property (Definition 1.27) and is finitely axiomatisable. In our case, the logic S5WD$_n$ is clearly finitely axiomatisable, so all we need to show in order to prove decidability is that it has the fmp.

In order to prove that S5WD$_n$ has the fmp, we use filtrations. The idea of filtrations is the following. If a logic is complete, we know that if a formula ϕ is a non-theorem of L (i.e. if $\neg\phi$ is L-consistent), then ϕ is invalid on some model M for L. The model M might be infinite. Filtrations enable us to produce a model M' from M, such that M' is finite. If we can further prove that M' is also a model for L (and therefore does not validate ϕ), then we have proved that the logic L has the finite model property.

Formally we proceed as follows. Given a formula ϕ, consider the set Φ_ϕ, composed by formulae α such that α is a well-formed sub-formula of ϕ or the negation of a well-formed sub-formula of ϕ. The set Φ_ϕ is obviously finite for any formula ϕ.

Definition 3.21. *Consider a model M. Two points $w, w' \in W$ are equivalent with respect to Φ_ϕ (written as $w \equiv_{\Phi_\phi} w'$ or simply $w \equiv w'$ if it is not ambiguous) if for any $\alpha \in \Phi_\phi$ we have $M \models_w \alpha$ if and only if $M \models_{w'} \alpha$.*

We can now define *filtrations* as follows.

Definition 3.22. *Given a formula ϕ and a model $M = (W, \{R_i\}_i, \pi)$, a* filtration *through Φ_ϕ is a model $M' = (W', \{R'_i\}_i, \pi')$ built as follows:*

- *$W' = W/\equiv_{\Phi_\phi}$, where $\equiv_{\Phi_\phi}$ is the equivalence relation defined by Definition 3.21.*
- *For each $i \in A$ the relation R'_i is* suitable, *i.e. it satisfies the two properties:*
  1. *For all $[w_1], [w_2] \in W'$, if there exists a point $u \in W$ such that $w_1 R_i u$ and $u \equiv w_2$, then $[w_1] R'_i [w_2]$*[6].
  2. *For all $[w_1], [w_2] \in W'$, if $[w_1] R'_i [w_2]$, then for all formulae α such that $\Box_i \alpha \in \Phi_\phi$, we have that if $M \models_{w_1} \Box_i \alpha$, then $M \models_{w_2} \alpha$.*
- *For any $p \in P$, $\pi'(p) = \{[w] \mid w \in \pi(p)\}$.*

Note that M' as defined above is finite because the set Φ_ϕ is.

It can be proved by induction (see for example [HC84] page 139) that if all the relations R'_i are "suitable" then the following holds.

Theorem 3.23. *Given a model M, and any formula ϕ, a filtration M' of M through Φ_ϕ is such that for any point $w \in W$ and and for any formula $\alpha \in \Phi$, $M' \models_{[w]} \alpha$ holds if and only if $M \models_w \alpha$*

We now proceed to the case of interest here: the logic S5WD$_n$.

Consider the canonical model M for S5WD$_n$, we know (see Theorem 3.12 and Lemma 3.13) that M is an equivalence model and that if we consider the model generated by any point of it, this is directed. Consider any formula ϕ and the model M' defined as follows:

Definition 3.24. *Given a model $M = (W, \{\sim_i\}_i, \pi)$ and a formula ϕ define the model $M' = (W', \{\sim'_i\}_i, \pi')$ by:*

- *$W' = W/\equiv_{\Phi_\phi}$, where $\equiv_{\Phi_\phi}$ is the equivalence relation defined by Definition 3.21.*
- *$[w_1] \sim'_i [w_2]$ if for all formulae α such that $\Box_i \alpha \in \Phi_\phi$, we have $M \models_{w_1} \Box_i \alpha$ if and only if $M \models_{w_2} \Box_i \alpha$.*
- *For any $p \in P$, $\pi'(p) = \{[w] \mid w \in \pi(p)\}$.*

Indeed the model M' defined by Definition 3.24 is a filtration as the following shows (stated in [HC84] page 145 for the mono-modal case).

Lemma 3.25. *Given an equivalence model M and a formula ϕ, the model M' as described in Definition 3.24 is a filtration.*

[6] Note that this condition is actually equivalent to $w R_i w'$ implies $[w] R'_i [w']$. Here we follow the definition given in [HC84].

Proof. All we need to prove is that the relations $\sim_i'$ are suitable.

Property 1. Consider worlds $[w_1], [w_2] \in W'$ and a world $u \in W$ such that $w_1 \sim_i u$ and $u \equiv w_2$. We need to prove that $[w_1] \sim_i' [w_2]$, i.e. that for all formulae α such that $\Box_i\alpha \in \Phi_\phi$, we have $M \models_{w_1} \Box_i\alpha$ if and only if $M \models_{w_2} \Box_i\alpha$. We prove it from left to right; the other direction is similar. Note that $M \models_{w_1} \Box_i\alpha$ if and only if $M \models_{w_1} \Box_i\Box_i\alpha$ because M is an equivalence model; but $w_1 \sim_i u$ and so $M \models_u \Box_i\alpha$. But $\Box_i\alpha \in \Phi$ and $w_2 \equiv u$, so $M \models_{w_2} \Box_i\alpha$, which is what we wanted to prove.

Property 2. Consider worlds $[w_1], [w_2] \in W'$ such that $[w_1] \sim_i' [w_2]$. This means that for all $\Box_i\alpha \in \Phi$, we have $M \models_{w_1} \Box_i\alpha$ if and only if $M \models_{w_2} \Box_i\alpha$. Since M is an equivalence model it follows that $M \models_{w_2} \alpha$.

□

We now prove that the filtration defined above produces models for $\mathrm{S5WD}_n$.

Lemma 3.26. *If M is an equivalence directed model, then for all ϕ such that $\vdash_{\mathrm{S5WD}_n} \phi$, the model $M' = (W', \{\sim_i'\}_i, \pi')$ as defined in Definition 3.24 is such that $M' \models \phi$.*

Proof. We prove that $F' = (W', \{\sim_i'\}_i)$ is a frame for $\mathrm{S5WD}_n$, i.e. it is an equivalence directed frame. The relations $\sim_i'$ are clearly equivalence relations. All it remains to show is that F' is directed. To do that, consider any $[w_1], \ldots, [w_n] \in W'$. Since M is directed, there exists $w \in W$ such that $w_i \sim_i \overline{w}$ for $i = 1, \ldots, n$. But each $\sim_i'$ is suitable and so, by a consequence of property 1 of suitability we have that $[w_i] \sim_i' [w]$, for $i = 1, \ldots, n$. Therefore the frame F' is directed. □

We are finally in the position to prove fmp.

Theorem 3.27. *The logic $\mathrm{S5WD}_n$ has the finite model property.*

Proof. Suppose $\not\vdash \phi$. Since by the proof of Theorem 3.12 the logic $\mathrm{S5WD}_n$ is canonical, the canonical model $M = (W, \{\sim_i\}_i, \pi)$ for $\mathrm{S5WD}_n$ is an equivalence model, it is weakly-directed and there is a point $w \in W$, such that $M \models_w \neg\phi$. Consider the model M_w generated (according to Definition 1.17) by w from M; by Lemma 1.18 we have $M_w \models_w \neg\phi$. The model M_w is clearly an equivalence model and, since it is connected, by Lemma 3.13, it is also directed. Consider now the filtration M' of M_w through Φ_ϕ according to Definition 3.24; by Lemma 3.26, M' is an equivalence directed model and it is finite by construction because Φ_ϕ is a finite set. But M' is a filtration, and by Theorem 3.23, $M' \models_{[w]} \neg\phi$, which is what we needed to prove. □

Corollary 3.28. *The logic $\mathrm{S5WD}_n$ is decidable.*

Proof. $\mathrm{S5WD}_n$ has the fmp and it is finitely axiomatisable. The result then follows from Theorem 1.26. □

3.6 The logic $\mathrm{S5WD}_n$

In the previous two sections we have explored hypercubes from a logical point of view. Our results can basically be summarised by Corollary 3.20 and Corollary 3.28. Corollary 3.20 states completeness of the logic $\mathrm{S5WD}_n$ with respect to hypercube systems, while Corollary 3.28 guarantees its decidability. What we did there is a technical analysis of the logic, but we should remember that we were unable to describe in an intuitive way what constraint the axiom **WD** imposes on hypercube systems in terms of knowledge.

In this section we try and explore the logic S5WD$_n$, with this aim in mind. In particular we study an equivalent formulation that can be interpreted more easily in terms of agents of knowledge that share part of the information they hold.

Let us start by analysing the type of constraint imposed by **WD** on the community of agents in the case $n = 2$:

$$\Diamond_1\Box_2 p \Rightarrow \Box_2\Diamond_1 p \qquad \mathbf{2WD}$$

We do not need to make the other conjunct explicit, as this can be obtained by taking the contrapositive of **2WD**.

Axiom **2WD** can be read as "If agent 1 considers possible that agent 2 knows the fact p, then agent 2 knows that agent 1 considers p possible". In other words, axiom **2WD** rules out a situation in which agent 1 considers possible that agent 2 knows p, while agent 2 considers possible that agent 1 knows not p. We can say that axiom **2WD** imposes a sort of homogeneity on the knowledge considered possible by other agents.

It is interesting to note that this constraint is logically equivalent to message passing of possible knowledge between agents. To see this, suppose that every time an agent considers possible that the other agent knows p, it broadcasts this information state, and this message is always safely received by the other agent. We also consider the communication to be always truthful. We have the axiom:

$$(\Diamond_1\Box_2 p \Rightarrow \Box_2\Diamond_1\Box_2 p) \wedge (\Diamond_2\Box_1 q \Rightarrow \Box_1\Diamond_2\Box_1 q) \qquad \mathbf{2WD'}$$

In S5$_2$, the two axioms are equivalent.

Lemma 3.29. $\vdash_{S5_2+\{\mathbf{2WD}\}} \mathbf{2WD'}$ *and* $\vdash_{S5_2+\{\mathbf{2WD'}\}} \mathbf{2WD}$

Proof. First conjunct. Suppose $\Diamond_1\Box_2 p \Rightarrow \Box_2\Diamond_1 p$, and assume $\Diamond_1\Box_2 p$. But then, since by Lemma 1.30 we have $\Box_i p \equiv \Box_i\Box_i p$ for any $i \in A$, we obtain $\Diamond_1\Box_2\Box_2 p$. So, by applying axiom **2WD** with argument $\Box_2 p$, we obtain $\Box_2\Diamond_1\Box_2 p$. Analogously, we can prove $\Diamond_2\Box_1 p \Rightarrow \Box_1\Diamond_2\Box_1 p$, by assuming $\Diamond_2\Box_1 p$, and axiom **2WD**.

Second conjunct. Suppose **2WD'** and $\Diamond_1\Box_2 p$. Then by **2WD'**, we have $\Box_2\Diamond_1\Box_2 p$. Consider now the instance of T $\Box_2 p \Rightarrow p$. By applying twice Lemma 1.29 we have $\Box_2\Diamond_1\Box_2 p \Rightarrow \Box_2\Diamond_1 p$. But then we have $\Diamond_1\Box_2 p \Rightarrow \Box_2\Diamond_1 p$. ☐

So, in the case of $n = 2$, the logic S5WD$_n$ specifies ideal agents of knowledge with an interaction between the two agents' knowledge that can be simulated by truthful communication of possible knowledge between the agents.

Let us now analyse the case $n = 3$. We have:

$$(\Diamond_1\Box_2 p_1 \wedge \Diamond_2\Box_3 p_2 \Rightarrow \Box_3\Diamond_1(p_1 \wedge p_2)) \wedge (\Diamond_1\Box_3 q_1 \wedge \Diamond_2\Box_1 q_2 \Rightarrow \Box_3\Diamond_2(q_1 \wedge q_2)) \qquad \mathbf{3WD}$$

The reading of the first conjunct of axiom **3WD** is "If agent 1 considers possible that agent 2 knows p_1 and agent 2 considers possible that agent 3 knows p_2, then agent 3 knows that agent 1 considers p_1 and p_2 possible".[7]

Considering the first conjunct with the special cases of $p_1 = \top, p_2 = \top, p_2 = \neg p_1$, it can be checked that axiom **3WD** implies the formulae: $\Diamond_1\Box_2 p_1 \Rightarrow \Box_3\Diamond_1 p$, $\Diamond_2\Box_3 p \Rightarrow \Box_3\Diamond_1 p$, and $\Diamond_1\Box_2 p \Rightarrow \Box_2\Diamond_1 p$. More generally, it is easy to see that

[7]We will discuss in Section 3.7 the reading of an axiom equivalent to **WD** for the general case.

$$\vdash_{S5_3+\{\mathbf{3WD}\}} \Diamond_i \Box_j p \Rightarrow \Box_k \Diamond_l p, \text{ where } i,j,k,l \in \{1,2,3\}$$

By Lemma 3.29 we have that these axioms can be rewritten in shape similar to $\mathbf{2WD}'$ and so could be seen as specifying a kind of broadcasting of information.

Notwithstanding this, the intuition is that, differently from the case $n = 2$, the axiom **3WD** is stronger than the simple conjunction of all these axioms. In fact, the semantic condition on the frames that corresponds to $\Diamond_i \Box_j p \Rightarrow \Box_k \Diamond_l p$ is weaker than property $3WD$.

We have the following already known result [Pop94, Cat88][8].

Lemma 3.30. *$F \models \Diamond_i \Box_j p \Rightarrow \Box_k \Diamond_l p$ if and only if F is such that for all w, w_1, w_2, such that wR_iw_1 and wR_kw_2, there exists a point $\overline{w}$ such that $w_1R_j\overline{w}$ and $w_2R_l\overline{w}$.*

Now, while we can prove that the property 3WD implies the property 2WD (the general case is proven later in Lemma 3.35), the opposite does not hold. Given the fact that by Theorem 3.12 all these logics are canonical we cannot have $\vdash_{S5_n} (\Diamond_i \Box_j p \Rightarrow \Box_k \Diamond_l p) \Rightarrow \mathbf{3WD}$, where $i,j,k,l \in \{1,2,3\}$.

In other words, it is unlikely that we can see the constraint imposed by **3WD** on the private knowledge of the agents as the result of a relatively simple broadcasting of possible knowledge among the agents.

If we consider the case for arbitrary n, it is indeed quite hard to have a clear picture of the meaning of the constraint imposed by the axiom **WD** or to see why agent n should play a special role in the group; still this does impose a constraint on the knowledge they hold. Intuitively, given the symmetry of property D, there must be a way of expressing axiom **WD** in a symmetric way. To understand better the implications of the logic S5WD$_n$ we axiomatise it in a slightly different way. We proceed as follows.

In Definition 3.9 we considered P_n to be the set of all the permutations $(x_1, \dots, x_n)$ without fixed-points $x_i = i$. In the following we show that it is possible to extend the results of the previous section and prove correspondence and completeness for the case of arbitrary permutations. It can then be observed that any permutation can be obtained by a sequence of swaps between two elements. By exploiting this observation it will be possible to define a logic which is still equivalent to S5WD$_n$, and therefore complete with respect to the same class of frames, but whose interaction axiom can more easily be understood in terms of knowledge. We proceed as follows: let P_n^* be the set of permutations of $\{1, \dots, n\}$ including those with fixed-points.

Definition 3.31 (nWD*). *A frame $F = (W, \{R_i\}_i)$ is nWD* if for all points $w, w_1, \dots, w_n$ in W, such that $wR_iw_i, i = 1, \dots, n$ and for all $(x_1, \dots, x_n)$ in P_n^*, there exists a point $\overline{w}$ in W such that $w_iR_{x_i}\overline{w}$, for all $i = 1, \dots, n$.*

When n is clear from the context, we just use we refer to nWD* just as WD*.

Definition 3.31 is matched by the axiom:

$$\bigwedge_{(x_1,\dots,x_n) \in P_n^*} (\Diamond_1 \Box_{x_1} p_1 \wedge \dots \wedge \Diamond_{n-1} \Box_{x_{n-1}} p_{n-1}) \Rightarrow \Box_n \Diamond_{x_n} (\wedge_{i=1}^{n-1} p_i) \qquad \mathbf{WD}^*$$

We have the two following results.

[8] This can also be shown to follow from Lemma 3.11.

Lemma 3.32. *$F \models \mathbf{WD}^*$ if and only if F is WD*.*

Proof. Analogous to the proof of Lemma 3.11. □

Theorem 3.33. *The logic S5WD$_n^*$ is sound and complete with respect to the class of equivalence WD* frames.*

Proof. Analogous to the proof of Lemma 3.12. The assumption of the permutations having no fix points can be relaxed with no harm. □

We now investigate how the logics S5WD$_n^*$ and S5WD$_n$ are related to each other. Since it contains more axioms, the logic S5WD$_n^*$ is stronger than S5WD$_n$, so we have:

Lemma 3.34. *If $\vdash_{\mathrm{S5WD}_n} \phi$, then $\vdash_{\mathrm{S5WD}_n^*} \phi$.*

It is interesting to note that the two logics are equivalent. Before showing that, we need the following:

Lemma 3.35. *If a frame is reflexive, transitive and nWD, then it is also mWD, where $m \leq n$.*

Proof. a) Suppose $m < n-1$. Note that $n > 3$. Assume $wR_jw_j, j = 1, \ldots, m$. Consider R_{m+1}, R_{m+2} on w, for any permutation $(x_1, \ldots, x_m) \in P_m$, there exists a permutation in $(y_1, \ldots, y_n) \in P_n$ such that $x_i = y_i, i = 1, \ldots, m$, and $(y_{m+1}, \ldots, y_n) \in P_{\{m+1,\ldots,n\}}$, where $P_{\{m+1,\ldots,n\}}$ is a permutation without fixed-points of the set $\{m+1, \ldots, n\}$. So, by nWD there exists a point $\overline{w}$ such that $w_iR_{y_i}\overline{w}, i = 1, \ldots m$ and $wR_j\overline{w}, j = m+1, \ldots, n$.

b) Suppose $m = n-1$. Consider $wR_jw_j, j = 1, \ldots, m$ and a permutation $(x_1, \ldots, x_m) \in P_m$. By a) there exists a point v_1, such that $w_jR_{x_j}v_1, j = 1, \ldots, m-1$. Consider $wR_{m-1}w_{m-1}$, wR_mw_m, by nWD there exists a point v_2 such that $w_{m-1}R_{x_{m-1}}v_2, w_mR_{x_m}v_2$. By transitivity $v_1R_{x_{m-1}}v_2$. Consider now $v_1R_iv_1$, for each $i \in \{1, \ldots, x_{m-2}\} \cup \{x_m\}$, and $v_1R_{x_{m-1}}v_2$, by nWD there exists a point $\overline{w}$ such that $v_2R_{x_m}\overline{w}$, and v_1 related to $\overline{w}$ by all the relations but R_{x_m}. So, by transitivity $w_jR_{x_j}\overline{w}, j = 1, \ldots, m$. □

Lemma 3.36. *If an equivalence frame is nWD, then it is nWD*.*

Proof. Consider $n+1$ points $w, w_1, \ldots, w_n$ such that $w \sim_i w_i, i = 1, \ldots, n$, and $(x_1, \ldots, x_n) \in P_n^*$, we want to prove that there exists a point $\overline{w}$ such that $w_i \sim_{x_i} \overline{w}, i = 1, \ldots, n$. Without loosing the generality of the problem assume that $(x_1, \ldots, x_n) = (y_1, \ldots, y_i, i+1, \ldots, n)$, where $(y_1, \ldots, y_i) \in P_i$, i.e. only the first i elements are fix-points free. Since the frame is nWD, then by Lemma 3.35, it is also iWD, i.e. there exists a point v_1 such that $w_j \sim_{y_j} v_1$, for $j = 1, \ldots, i$. Consider $w \sim_i w_i, w \sim_{i+1} w_{i+1}$ together with $n-1$ reflexive links on w (note that $y_{i+1} \neq i+1$ by construction), by WD and considerations on the equivalence relations (similar reasoning for the case $n = 3$ is reported in Figure 3.9) there exists a point v_2, such that $w_i \sim_{y_i} v_2, w_{i+1} \sim_{i+1} v_2$. By transitivity $v_1 \sim_{y_i} v_2$. Consider now v_1, v_2, $n-1$ relation on v_1, and $v_1 \sim_{y_i} v_2$; by WD and transitivity we have the existence of a point v_3 such that $v_2 \sim_{i+1} v_3$ and v_3 related to v_2 by all relations but $\sim_{i+1}$. Note that $v_3 \sim_{y_j} w_j, j = 1, \ldots i, v_3 \sim_{i+1} w_{i+1}$. By transitivity we also have $v_3 \sim_{i+1} w$. Consider $v_3 \sim_{i+1} w$ and $n-1$ relations on v_3; by WD there exists a point v_4 such that $w \sim_{i+2} v_4$, and v_4 related to v_3 by all the relations but $\sim_{i+2}$. Note that $v_4 \sim_{i+2} w_{i+2}, v_4 \sim_{i+1} w_{i+1}, v_4 \sim_j w_j, j = 1, \ldots i$. Note that now we can similarly apply WD to $v_4 \sim_{i+2} w$. By continuing this construction we identify the point $\overline{w} = v_{n-i+1}$ such that $\overline{w} \sim_{y_j} w_j, j = 1, \ldots, i, \overline{w} \sim_k w_k, k = i, \ldots, n$. □

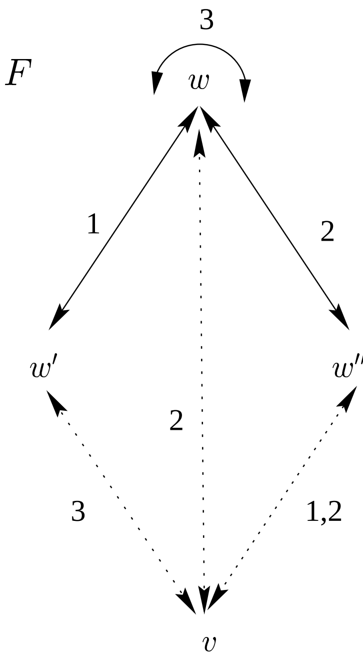

Figure 3.9: Applying property 3WD in the case of equivalence frames as in the proof of Lemma 3.36. By considering a 3-reflexive link on the point w, the point w'' can be proven connected to v not just by relation 1 but also by relation 2.

$$\begin{tikzcd}
\mathrm{S5WD}^*_n \vdash \phi \arrow[rr, dashed] \arrow[d, leftrightarrow, "Theorem\ 3.33"'] & & \mathrm{S5WD}_n \vdash \phi \arrow[d, leftrightarrow, "Theorem\ 3.12"] \\
\mathcal{F}_{EWD^*} \models \phi \arrow[rr, "Lemma\ 3.36"'] & & \mathcal{F}_{EWD} \models \phi
\end{tikzcd}$$

Figure 3.10: Scheme of proof of Theorem 3.37.

The previous lemma allows us to prove:

Theorem 3.37. *If* $\vdash_{\mathrm{S5WD}_n^*} \phi$*, then* $\vdash_{\mathrm{S5WD}_n} \phi$*.*

Proof. See Figure 3.10. Suppose $\vdash_{\mathrm{S5WD}_n^*} \phi$, by Theorem 3.33, we have $\mathcal{F}_{EWD^*} \models \phi$, where $\mathcal{F}_{EWD^*}$ is the class of reflexive, symmetric, transitive, and WD* frames. By Lemma 3.36, $\mathcal{F}_{EWD} \subseteq \mathcal{F}_{EWD^*}$, where $\mathcal{F}_{EWD}$ is the class of reflexive, symmetric, transitive, and WD frames. So, $\mathcal{F}_{EWD} \models \phi$, and by Theorem 3.12 we have $\vdash_{\mathrm{S5WD}_n} \phi$. □

We therefore have the following.

Corollary 3.38. *The logics S5WD$_n$ and S5WD$_n^*$ are equivalent.*

Proof. Immediate from Theorem 3.37 and Lemma 3.34. □

This means that hypercubes can be axiomatised by considering the **WD***-extension of S5$_n$, rather than the one obtained by **WD**. The difference in not ruling out fixed points in the axiom is that some of the conjuncts in the antecedent of the axioms can be replaced by knowledge operators. In fact we have the equivalence (Lemma 1.30) $\vdash_{\mathrm{S5WD}_n} \Diamond_i\Box_i p \Leftrightarrow \Box_i p$ and substitution of equivalents hold.

For example, in the case $n = 4$, the following formulae are derivable from $\mathbf{WD}^*$:

$$\Box_1\phi_1 \wedge \Box_2\phi_2 \wedge \Box_3\phi_3 \Rightarrow \Diamond_4(\phi_1 \wedge \phi_2 \wedge \phi_3) \tag{3.4}$$

and

$$\Box_1\phi_1 \wedge \Diamond_2\Box_3\phi_2 \wedge \Diamond_3\Box_2\phi_3 \Rightarrow \Diamond_4(\phi_1 \wedge \phi_2 \wedge \phi_3) \tag{3.5}$$

Formula 3.4 simply says "agent 4 considers possible the conjunction of the facts known by agents 1, 2 and 3"; in fact, we can substitute for ϕ_1 the conjunction of all the atoms known by agent 1, analogously for agent 2 and 3. This is not a surprising result; it is actually a theorem of S5$_n$, actually even T$_n$ proves it as it follows from reflexivity. So, if we limit our attention to formulae that do not contain nested modal operators, the logics S5WD$_n$ and S5WD$_n^*$ do not add anything new to S5$_n$.

On the contrary, Formula 3.5 is not provable in S5$_n$: it is an instance of $\mathbf{WD}^*$ which follows by taking the permutation $(x_1, x_2, x_3, x_4) = (1, 3, 2, 4)$. It can be read as "The conjunct ϕ_1 and ϕ_2 and ϕ_3 is consistent with the knowledge of agent 4, where ϕ_1 is a fact known by agent 1, ϕ_2 is a fact that agent 2 thinks may be known by agent 3, and ϕ_3 is a fact that agent 3 thinks may be known by agent 2". Intuitively S5WD$_n$-agents do not just consider possible conjuncts of facts known by some agents, but *also* facts that are considered possible to be known by some other agent. The aim of the rest of the section is to clarify this intuition.

In defining the axiom $\mathbf{WD}^*$ we have been very liberal in allowing as many fixed-points in the permutations as we wanted. Can we simplify the axiom $\mathbf{WD}^*$ by fixing the number of fixed points in the permutation without weakening the logic?

Observe that an instance of $\mathbf{WD}^*$ with j fixed points in the permutations translates into j conjuncts composed by a single modal operator. This kind of axioms in turn correspond to the property nWD* with j fixed-points in the permutation of indices of the relations connecting the points to the world $\overline{w}$ in Definition 3.31. Observe now that any of these permutations can be achieved by considering *a sequence* of permutations in which only *two* elements are exchanged (swapped). We formalise these observations as follows.

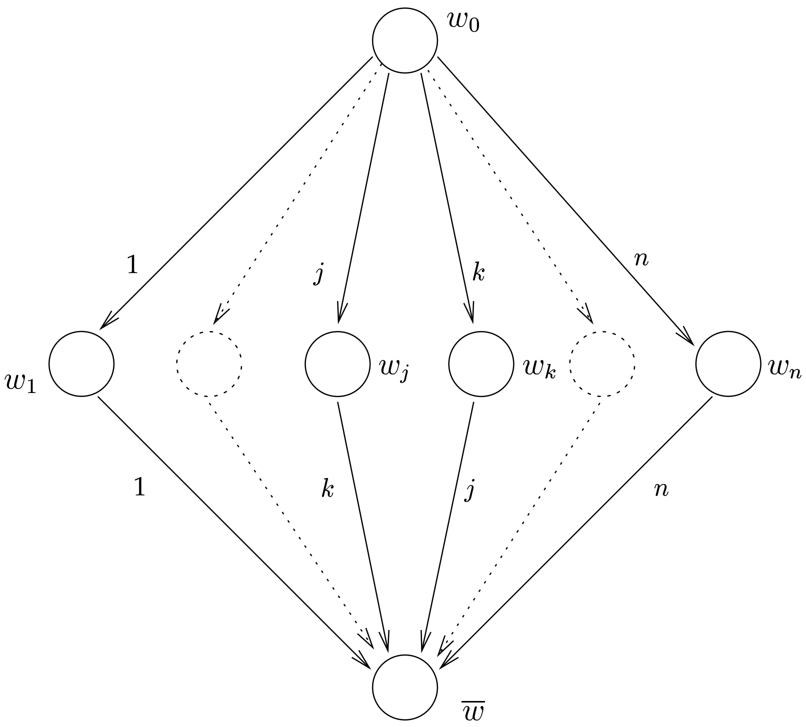

Figure 3.11: The property nWD**.

Definition 3.39 (nWD**). *Let P_n^{**} be the set of permutations $(x_1, \ldots, x_n)$ of $\{1, \ldots, n\}$ in which $x_i = i, i = 1, \ldots, n$, except for $j, k \in \{1, \ldots, n\}$ such that $x_j = k, x_k = j$.*

A frame $F = (W, \{R_i\}_i)$ is nWD** if for all $w, w_1, \ldots, w_n$ in W, such that $wR_iw_i, i = 1, \ldots, n$ and for all $(x_1, \ldots, x_n)$ in P_n^{**} there exists a point $\overline{w}$ such that $w_iR_{x_i}\overline{w}$, for all $i = 1, \ldots, n$.

See Figure 3.11

Consider now the following axiom:

$$\bigwedge_{1\leq j,k,l\leq n} \Bigg(\Bigg(\bigwedge_{\substack{i=1 \\ i\neq j, i\neq k, i\neq l}}^{n} \Box_i p_i \Bigg) \wedge \Diamond_j \Box_k p_j \wedge \Diamond_k \Box_j p_k \Bigg) \Rightarrow \Diamond_l (\bigwedge_{\substack{i=1 \\ i\neq l}}^{n} p_i) \qquad \textbf{WD}^{**}$$

Intuitively, when working with equivalence models, axiom **WD**** corresponds to property nWD**. The proof is analogous to the ones presented in Lemma 3.11 in which the assumption of working on equivalence models is used.

Lemma 3.40. *Consider an equivalence frame F. $F \models$ **WD**** if and only if F is WD**.*

Proof. From left to right. Consider an equivalence frame F and suppose $F \models$ **WD**. Consider $n + 1$ points $w, w_1, \ldots, w_n$ of F, such that $w \sim_i w_i, i = 1, \ldots, n$. Fix a permutation $(1, \ldots, j, \ldots, k, \ldots, n) \in P^{**}$, any element l in $\{1, \ldots, n\}$, with $j \neq k \neq l$ and consider the following valuation π: $\pi(p_i) = \{v : w_i \sim_i v\}$ for $i = 1, \ldots, n$ but $i \neq j, i \neq k, i \neq l$;

$\pi(p_j) = \{v : w_j \sim_k v\}$, $\pi(p_k) = \{v : w_k \sim_j v\}$. By construction we have

$$(F,\pi) \models_w \left(\bigwedge_{\substack{i=1 \\ i\neq j, i\neq k, i\neq l}}^{n} \Diamond_i \Box_i p_i \right) \wedge \Diamond_j \Box_k p_j \wedge \Diamond_k \Box_j p_k.$$

But since (F,π) is an equivalence model (by Lemma 1.30 and soundness of S5$_n$) we have that $(F,\pi) \models_w \Diamond_i \Box_i p \Leftrightarrow \Box_i p$. And so, by applying axiom **WD**** we obtain

$$(F,\pi) \models_w \Diamond_l (\bigwedge_{\substack{i=1 \\ i\neq l}}^{n} p_i).$$

So, again by Lemma 1.30 we have

$$(F,\pi) \models_w \Box_l \Diamond_l (\bigwedge_{\substack{i=1 \\ i\neq l}}^{n} p_i).$$

So, we have

$$(F,\pi) \models_{w_l} \Diamond_l (\bigwedge_{\substack{i=1 \\ i\neq l}}^{n} p_i).$$

So there must exist a world $\overline{w}$, such that $w_l \sim_l \overline{w}$ and

$$(F,\pi) \models_{\overline{w}} (\bigwedge_{\substack{i=1 \\ i\neq l}}^{n} p_i).$$

But, by construction of the interpretation π, this implies $w_i \sim_i \overline{w}$, for $i = 1, \ldots, n; i \neq j, i \neq k$ and $w_j \sim_k \overline{w}$ and $w_k \sim_j \overline{w}$. So the frame satisfies property nWD**.

From right to left. Consider any three distinct indices $j, k, l \in \{1, \ldots, n\}$, associate to this the permutation $(1, \ldots, j, \ldots, k, \ldots n) \in P_n^{**}$ and consider a model (F,π) with a point w, such that

$$(F,\pi) \models_w \left(\bigwedge_{\substack{i=1 \\ i\neq j, i\neq k, i\neq l}}^{n} \Box_i p_i \right) \wedge \Diamond_j \Box_k p_j \wedge \Diamond_k \Box_j p_k.$$

For the same considerations as above, there exist $n-1$ worlds $w_i, i \in \{1, \ldots, n\} \setminus \{l\}$ such that $(F,\pi) \models_{w_i} \Box_i p_i$, where $w \sim_i w_i, i = 1, \ldots, n$, but $i \neq j, i \neq k, i \neq l$ and $(F,\pi) \models_{w_j} \Box_k p_j$ and $(F,\pi) \models_{w_k} \Box_j p_k$, where $w \sim_j w_j, w \sim_k w_k$. We want to prove

$$(F,\pi) \models_w \Diamond_l (\bigwedge_{\substack{i=1 \\ i\neq l}}^{n} p_i).$$

By Lemma 1.30 it is enough to prove

$$(F,\pi) \models_w \Box_l \Diamond_l (\bigwedge_{\substack{i=1 \\ i\neq l}}^{n} p_i).$$

i.e. that for any point w_l, such that $w \sim_l w_l$ we have

$$(F,\pi) \models_{w_l} \Diamond_l(\bigwedge_{\substack{i=1\\i\neq l}}^{n} p_i).$$

Consider any such point w_l, since the frame F is nWD**, by considering the permutation that swaps j with k, there exists a point $\overline{w}$, such that $w_i \sim_i \overline{w}$, for $i = 1,\ldots,n$ with $i \neq j, i \neq k$; $w_j \sim_k \overline{w}$ and $w_k \sim_j \overline{w}$. Note that $w_l \sim_l \overline{w}$. But then $(F,\pi) \models_{\overline{w}} p_i$, for $i = 1,\ldots n, i \neq l$, that is

$$(F,\pi) \models_{\overline{w}} (\bigwedge_{\substack{i=1\\i\neq l}}^{n} p_i).$$

So, since $l \in \{1,\ldots,n\}$ we obtain

$$(F,\pi) \models_{w_l} \Diamond_l(\bigwedge_{\substack{i=1\\i\neq l}}^{n} p_i).$$

□

Completeness also follows.

Theorem 3.41. *The logic* $\mathrm{S5WD}_n^{**}$ *is sound and complete with respect to the class of equivalence WD** frames.*

Proof. Only a sketch of the proof is given as it is very similar to the one presented for Theorem 3.12. By reasoning by contradiction and employing the same construction on the canonical model M given on that proof we can rearrange Equation 3.1 into

$$\vdash_{\mathrm{S5WD}_n^{**}} \bigwedge_{\substack{i=1\\i\neq l}}^{n} \alpha_i \Rightarrow \neg\alpha_l.$$

By necessitate by $\Box_l$ (rather than by $\Box_{x_n}$ as in Theorem 3.12) we have

$$\vdash_{\mathrm{S5WD}_n^{**}} \Diamond_l(\bigwedge_{\substack{i=1\\i\neq l}}^{n} \alpha_i) \Rightarrow \neg\Box_l\alpha_l. \tag{3.6}$$

As in the proof of Theorem 3.12 but for the case of P_n^{**} we have now that $\Box_i\alpha_i \in w_i, i = 1,\ldots,n$, except for j,k for which we have $\Box_j\alpha_k \in w_k, \Box_k\alpha_j \in w_j$. By using the same construction, since the canonical model is an equivalence model, we have that

$$M \models_w \left(\bigwedge_{\substack{i=1\\i\neq j, i\neq k, i\neq l}}^{n} \Box_i\alpha_i\right) \wedge \Diamond_j\Box_k\alpha_j \wedge \Diamond_k\Box_j\alpha_k.$$

So by applying axiom **WD**** we have that

$$M \models_w \Diamond_l(\bigwedge_{\substack{i=1\\i\neq l}}^{n} \alpha_i),$$

which together with Equation 3.6 gives a contradiction on the set w_l exactly as in Theorem 3.12. □

Actually the logics $\mathrm{S5WD}_n^{**}$, $\mathrm{S5WD}_n^{*}$, and $\mathrm{S5WD}_n$ can be proven to be equivalent. To see it, first observe that **WD**** is a special case of **WD*** and so:

Lemma 3.42. *If* $\vdash_{\mathrm{S5WD}_n^{**}} \phi$*, then* $\vdash_{\mathrm{S5WD}_n^{*}} \phi$

We can also prove:

Lemma 3.43. *If an equivalence frame is* n*WD****, then it is* n*WD***.*

Proof. Consider $n+1$ points $w, w_1, \ldots, w_n$ such that $wR_iw_i, i = 1, \ldots, n$, and $(x_1, \ldots, x_n) \in P_n^*$, we want to prove that there exists a point $\overline{w}$ such that $w_iR_{x_i}\overline{w}, i = 1, \ldots, n$. Suppose $1 \neq x_1$ (if not go to the next step) and apply nWD** to $wR_iw_i, i = 1, \ldots, n$ by taking the swap that assigns the element 1 to x_1. We obtain a point v_1 such that $w_1R_{x_1}v_1$, and $w_iR_iv_1, i = 2, \ldots n$, short of the point that has been swapped with 1. We can now apply nWD** keeping x_1 fixed and swapping 2 with x_2, if necessary. By systematically applying nWD**, after n times at maximum, we obtain $\overline{w}$ such that $w_iR_{x_i}\overline{w}, i = 1, \ldots, n$. □

Lemma 3.43 allows us to prove:

Theorem 3.44. *If* $\vdash_{\mathrm{S5WD}_n^{*}} \phi$*, then* $\vdash_{\mathrm{S5WD}_n^{**}} \phi$*.*

Proof. Analogous to the proof of Theorem 3.37. □

We have therefore proved that:

Corollary 3.45. *The logics* $\mathrm{S5WD}_n^{*}$*, and* $\mathrm{S5WD}_n^{**}$ *are equivalent.*

Proof. It follows from Theorem 3.44 and Lemma 3.42. □

Corollary 3.46. *The logics* $\mathrm{S5WD}_n$*,* $\mathrm{S5WD}_n^{*}$*, and* $\mathrm{S5WD}_n^{**}$ *are equivalent.*

Proof. It follows from Corollary 3.38 and Corollary 3.45. □

We have now come to the conclusion that formally the logic $\mathrm{S5WD}_n^{**}$ is just as good as $\mathrm{S5WD}_n$ to represent MAS modelled by hypercube systems. From a mathematical point of view it looks as if the logic $\mathrm{S5WD}_n$ should be preferable as it seems to convey better the information about the underlying semantics. Notwithstanding this, the equivalence result we have achieved is still useful. Indeed we will use it in the next section to give a more concrete meaning of how MAS modelled by hypercube systems share part of their knowledge.

3.7 Knowledge sharing in hypercube systems

In this chapter we have proven soundness, completeness and decidability for the class of hypercube systems that we explored semantically in Chapter 2. In section 3.6 we observed that our first result, the logic $\mathrm{S5}_n$ enriched by the axiom **WD**, seems to suggest an asymmetric property of the agents of the group, with agent n having a special role.

This is counter-intuitive since the semantic properties of hypercube systems are symmetric and all the relations play the same role in defining properties D and I. This observation lead us to the study of Section 3.6 in which we proved that the logic $S5WD_n$ is equivalent to the logic $S5WD_n^{**}$ (Corollary 3.46). To achieve this result we had to prove an intermediate equivalence with the logic $S5WD_n^{*}$ (Corollary 3.38).

The advantage of enriching $S5_n$ by using the axiom $\mathbf{WD}^{**}$ rather than $\mathbf{WD}$ or $\mathbf{WD}^{*}$ is that its meaning is more immediate in terms of the knowledge of the agents. In fact, axiom $\mathbf{WD}^{**}$ specifies a class of knowledge agents with the following property:

Observation 3.47 (Knowledge sharing in hypercubes). *Let $A = \{1, \ldots, n\}$ be a community of agents and consider any three agents, j, k, l. Then agent l thinks that the conjunction of*

- *anything that j thinks may be known by k,*
- *anything that k thinks may be known by j,*
- *anything known by any other agent,*

is possible.

MAS modelled by hypercube systems share knowledge among themselves following Observation 3.47. They are ideal agents of knowledge (as described on page 24, but they also consider possible facts known by some agents, or regarded by some agent to be possibly known by some other agent. Observation 3.47 specifies how private knowledge is shared in such community of agents.

By means of Observation 3.47 we can now see that that hypercubes are homogeneous; i.e. there is no distinction in capabilities of the agents, differently from the examples presented in Section 1.4.2. Still, the axiom $\mathbf{WD}^{**}$ does indicate an *interaction* among the agents. What kind of interaction does $\mathbf{WD}^{**}$ model? We answer this question in the next section.

3.8 Homogeneous broadcasting systems

3.8.1 Introduction

In the last section we presented a relatively intuitive explanation of the extent to which agents modelled by hypercubes share part of their knowledge. This is quite valuable because that formulation was also proved to be sound and complete with respect to hypercubes.

Still, even that axiomatisation does not give us many hints about what kind of process hypercubes model. In Section 3.6 we did try to explore this question, but our analysis was successful only for the case of 2 agents. To produce a more meaningful result we have to abandon the proof-theoretical level and go back to the semantics. More precisely, we would really like to have a low-level description language able to model the processes of the agents. Given the fact that hypercubes share part of their knowledge, *communication* seems to be a key aspect that we would like to model. In doing so, we would also like to be able to re-introduce a mechanism able to represent the temporal evolution of hypercubes, so that, if appropriate, we could analyse whether hypercube systems model a class of agents with one of the characteristics (perfect recall, no learning, etc.) that we discussed in Section 1.3.2.3.2.

We present an answer to this question in the rest of this chapter. Ultimately, we will present a result showing that hypercube systems represent a special class of synchronous agents with perfect recall that exchange information only by broadcasting and that start their configuration in a state of complete ignorance of each other's local states.

The section is organised as follows. In Section 3.8.2 we present the low-level semantic formalism of *environments*, that we will use to model hypercubes. In Section 3.8.3 we will introduce additional constraints on it that will let us characterise hypercubes as broadcasting agents.

Note. The material presented in this section is joint work with Ron van der Meyden, University of Technology (Sydney) and the results are sketched here only because of the relevance they have with the rest of this chapter. The interested reader is referred to [LMR99] for a more detailed exposition.

3.8.2 Environments

Although interpreted systems (Section 1.3.1.1) are certainly a very expressive formalism for expressing MAS systems, one could argue that they are perhaps a bit too general to represent certain classes of distributed processes. One example for which they seem to be inadequate is the case of *protocols*. Intuitively, in practical applications we are not interested in the set of *all* possible runs of a system, but only in the ones that follow some constraints. Indeed in certain cases we would like to be able to express *explicitly* these constraints regulating the transitions from a state to another.

This is the reason why Fagin and colleagues introduced the formalism of *contexts* (see [FHMV97, FHMV95]), in which the notion of protocol is explicitly present. In the following we use a variant of contexts, called *environments*, presented in [Mey96][9]. The interested reader can find more details and motivations in [Mey96]. The aim of the following is to show that executing a protocol in an environment determines a Kripke frame that describes states of knowledge of the agents. In Section 3.8.3 we will discuss under what assumptions these evolutions generate exactly a class of agents which is modelled by the same logic as hypercubes.

Differently from the rest of this thesis, we work with a set $A = \{0, 1, \ldots, n\}$ of agents. Agent 0 is intuitively the environment of interpreted systems and it will play the special role of modelling the architecture of the systems.

Definition 3.48 (Actions). *For any $i \in A$ the non-empty set ACT_i is the set of* actions *for agent i. A* joint action *is a tuple $(a_0, \ldots, a_n)$, where $a_i \in ACT_i$ is an action for agent i. The set ACT is the set of joint actions.*

Actions of agent 0 correspond to nondeterministic behaviour of the context in which the agents are situated. The fundamental notion of this model is the definition of *environment*.

Definition 3.49 (Environment). *An* interpreted environment *is a tuple of the form $E = (S, I, P_0, \tau, O, \pi)$ where the components are as follows:*

- *S is a set of* states of the environment*. Intuitively, states of the environment may encode such information as messages in transit, failure of components, etc.*
- *I is a subset of S, representing the possible* initial states *of the environment.*

[9] For convenience, we use a variant of it already encoding perfect recall.

- $P_0 : S \to 2^{ACT_0}$ *is a function, called the* protocol of the environment, *mapping states to subsets of the set* ACT_0 *of actions that can be performed by the environment. Intuitively,* $P_0(s)$ *represents the set of actions that may be performed by the environment when the system is in the state* s.
- τ *is a function mapping joint actions* $a \in ACT$ *to state transition functions* $\tau(a) : S \to S$. *Intuitively, when the joint action* a *is performed in the state* s*, the resulting state of the environment is* $\tau(a)(s)$.
- O *is a function from* S *to* $\mathcal{O}^n$ *for some set* $\mathcal{O}$ *of observations. For each* $i, i = 1, \ldots, n$*, the function* O_i *mapping* $s \in S$ *to the* i*th component of* $O(s)$*, is called the* observation function of agent i. *Intuitively,* $O_i(s)$ *represents the* observation *performed by agent* i *in the state* s.
- $\pi : P \to 2^S$ *is a valuation for the atoms.*

The definition above defines transitions over states. Given an environment, sequences of states related by transition functions define a *trace*.

Definition 3.50 (Traces). *A* trace *of an environment* E *is a* finite *sequence* $s_0 \ldots s_m$ *of states such that* $s_0 \in I$ *and for all* k *such that* $k = 0 \ldots m-1$ *there exists a joint action* $a = (a_0, \ldots, a_n)$ *such that* $s_{k+1} = \tau(a)(s_k)$ *and* $a_0 \in P_0(s_k)$. *Given a trace* $r = s_0 \ldots s_m$*, fin*$(r) = s_m$ *is the final state of the trace.*

The intuition is that a trace r represents a finite history of the system.

Note that in the transitions above, agent 0 follows its own protocol P_0. This is not the case for the other agents that in principle can perform any possible action. In practice we would like to specify what protocol these agents follow. In the context of this work we will assume the agents follow a *perfect recall protocol* which is defined as follows:

Definition 3.51 (Perfect Recall). *Given an environment* E *and a trace* $r = s_0 \ldots s_m$ *on it, the* perfect recall local state *of an agent* $i, i \in \{1, \ldots, n\}$ *in a trace* r *is defined as the sequence* $\{r\}_i = O_i(s_0) \ldots O_i(s_m)$ *of observations made by agent* i *in the trace* r.

A perfect recall protocol *for agent* $i, i \in \{1, \ldots, n\}$ *is a function* P_i *mapping each sequence of observations in* $\mathcal{O}^*$ *to a non-empty subset of* ACT_i. *A* joint *perfect recall protocol is a tuple* $P = (P_1, \ldots, P_n)$*, where for every agent* $i, i \in \{1, \ldots, n\}$ P_i *is a perfect recall protocol for agent* i.

Protocols specify the actions that are allowed for the agents. More precisely, we say that given a trace r on an environment E and an agent $i, i \in \{1, \ldots, n\}$, an action $a_i \in ACT_i$ is *enabled* with respect to a joint protocol $P = (P_1, \ldots P_n)$ if $a_i \in P_i(\{r\}_i)$. We also say that an action a_0 of the environment is enabled at r if $a_0 \in P_0(\mathit{fin}(r))$. A joint action $a = (a_1, \ldots, a_n)$ is enabled at r with respect to a protocol P if each of its components a_i is enabled at r.

If all the agents follow their protocol by executing enabled actions we obtain a *consistent* trace. More precisely, given an environment E and a joint protocol P, a trace $r = s_0 \ldots s_m$ on E is *enabled* if for each $k < m$, there exists a joint action a enabled at $s_0 \ldots s_k$ with respect to P, such that $\tau(a)(s_k) = s_{k+1}$.

All the enabled traces define the intended evolutions of the environment according to the joint protocol. Similarly to what we saw in the case of interpreted systems it is possible to ascribe knowledge to agents following a perfect recall protocol. Since agents perform observations it is meaningful to assume two states to be indistinguishable for an agent if the series of observation she has performed in the two states are the same. So, once again we can define a Kripke model from a low-level description.

Definition 3.52 (Perfect recall frame derived from a protocol and environment). *Let E be an environment and let P be a joint protocol. The* perfect recall frame *derived from E and P is the structure $F_{E,P} = (W, \sim_1, \ldots, \sim_n)$, where*

- *W is the set of all traces of the environment E consistent with the protocol P,*
- *For every $i, i \in \{1, \ldots, n\}$, the relation $\sim_i \subseteq W \times W$ is defined by $r \sim_i r'$ if $\{r\}_i = \{r'\}_i$.*

It is also possible to derive a Kripke model by considering the same valuation π of the environment E in question.

From the way the relations $\sim_i$ are defined, it is clear that every agent has perfect recall and it is common knowledge that the environment they are operating in is E and that the joint protocol is P.

By taking other accessibility relations one can encode different phenomena (for example one could define two states to be indistinguishable if their latest observation is equal). The assumption of perfect recall is widely used in computer science because it amounts to assuming that the agents use all the information they acquired in an ideal way.

3.8.3 Homogeneous broadcasting agents

Above we have briefly described the machinery of interpreted environments and observed these generate a Kripke frame in a way which is similar to interpreted systems. But environments offer more expressive potentialities than interpreted systems, because of the expressiveness they offer.

From the definitions of the previous section it is possible to define a subclass of environments, called *homogeneous broadcasting environment* in [LMR99] that enjoy special properties. The details are not fully reported here and we refer the reader to that paper.

Briefly, a homogeneous broadcasting environment is an environment with the following additional constraints.

- Actions consist of pairs of internal and external actions,
- The states of the environment consist of tuples incorporating the external actions being performed on that state together with the private states of the agents;
- In initial states no external action is present and agents are ignorant of each other's state and of the state of the environment;
- Agents observe their own private states and the external actions performed by the other agents;
- Every agent updates its private state depending on its own private state and on the external actions performed by the agents;
- The protocol of the environment depends only on its own internal state and on the last performed external action; the agents run a perfect recall protocol.

Under these assumptions, it was proven in [Mey98] that the class of perfect recall frames generated by homogeneous broadcasting agents has the following interesting property.

Theorem 3.53 ([Mey98]). *The logic* $\mathrm{S5WD}_n$ *is sound and complete with respect to the class of perfect recall frames generated by homogeneous broadcasting agents.*

Theorem 3.53 provides a characterisation of hypercubes on a low-level semantics like the one of environments. The broadcasting activity that they perform is the act responsible for the sharing of knowledge that they exhibit (Observation 3.47).

We have now completed our analysis of hypercube systems. We defined them as a special class of interpreted systems, we found a semantic correspondence in the class of Kripke frames, we provided some complete and decidable axiomatisations for them and in this section we reported a low-level description of their communication ability.

Chapter 4

A spectrum of degrees of knowledge sharing

4.1 Introduction

In Chapter 3 we explored the axiomatisation of hypercube systems and proved that the logic S5WD$_n$ is sound and complete with respect to that semantic class. The logic S5WD$_n$ is axiomatised by extending S5$_n$ with the axiom **WD**.

For the case $n = 2$ the axiom **WD** becomes the formula:

$$\Diamond_1 \Box_2 p \Rightarrow \Box_2 \Diamond_1 p \qquad \textbf{2WD}$$

which can be read as "If agent 1 considers possible that agent 2 knows p then agent 2 must know that agent 1 considers possible that p is the case". In Section 3.8 we showed that this logic can be seen as modelling a particular class of MAS that exchange information by broadcasting.

Axiom **2WD** is an interaction axiom that models a particular class of agents of knowledge; but surely there must be other interesting classes of interactions between agents that can be modelled by extensions of S5$_n$. For example, on page 38 we presented two other examples of agents sharing knowledge.

In the first example we described a collective map making scenario from [dMAE+97] in which an agent j is told any knowledge of any other agent i and therefore knows everything that is known by any other agent. Assuming all the agents being ideal, this scenario can be represented by S5$_n$ enriched by the interaction axiom:

$$\Box_i p \Rightarrow \Box_j p; \text{ for all } i \in A.$$

The second scenario of page 38 concerns a MAS whose agents have computation capabilities that can be ordered. If the agents are executing the same program on the same data, then it is reasonable to model the MAS by enriching the logic S5$_n$ by:

$$\Box_i p \Rightarrow \Box_j p;\ i \prec j, \text{ for all } i, j \in A.$$

The relation $\prec$ expresses the linear order in the computational power at disposal to the agents. In this as in the previous case some information is being shared among the agents of the group.

It is easy to imagine other meaningful axioms that express *interactions* between the agents in the system; clearly there is a *spectrum* of possible degrees of knowledge sharing. At one end of the spectrum is $S5_n$, with no sharing at all. At the other end, there is $S5_n$ together with

$$\Box_i p \Leftrightarrow \Box_j p; \text{ for all } i, j \in A,$$

saying that the agents have precisely the same knowledge (total sharing). The three examples mentioned above exist somewhere in the (partially ordered) spectrum between these two extremes.

Our aim in this chapter is to explore the spectrum systematically. We restrict our attention to the case of two agents (i.e. to extensions of $S5_2$), and explore axioms of the forms

$$\begin{array}{l} \boxdot p \Rightarrow \boxdot p \\ \boxdot p \Rightarrow \boxdot \boxdot p \\ \boxdot \boxdot p \Rightarrow \boxdot p \\ \boxdot \boxdot p \Rightarrow \boxdot \boxdot p \end{array}$$

where each occurrence of $\boxdot$ is in the set $\{\Diamond_1, \Box_1, \Diamond_2, \Box_2\}$.

Technically we will prove correspondence properties and completeness for extensions of $S5_2$ with axioms of these forms. Naturally, this will not give the complete picture: there may be interesting axioms of other forms than those listed above. However, analysis of the literature certainly suggests that most axioms studied for this purpose are of one of these forms. They are sufficient for expressing how knowledge and facts considered possible are related to each other up to a level of nesting of two, which is about the maximum that human intuition can grasp. Note also that the examples above, including the case of bi-dimensional hypercubes, are included in the axiom patterns.

Although our analysis is limited both from considering the case of two agents and from considering only interaction axioms of the shape above, we will see that some non trivial technical problems are present here. Indeed, we will leave two completeness problems as open.

The rest of this chapter is organised as follows. In Section 4.2 we analyse and discuss interaction axioms of the form $\boxdot p \Rightarrow \boxdot p$. We will then extend these results in Section 4.3 where we discuss the case of the consequent being composed by two modal operators. In Section 4.4 we will analyse the interaction axioms resulting from two nested modalities both in the antecedent and in the consequent. Finally in Section 4.5 we present the lattice generated by the logics and discuss our results.

This chapter is devoted to extensions of $S5_2$ and so in the following we will use the provability symbol $\vdash$ for $\vdash_{S5_2}$. Given this context we will always be working in the class $\mathcal{F}_E$ of equivalence frames $F = (W, \sim_1, \sim_2)$ built on two equivalence relations on W. Many of these results can be translated for the system K_2; given that the chapter is devoted exclusively to knowledge agents we will not discuss this.

4.2 Interaction axioms of the form $\boxdot p \Rightarrow \boxdot p$

In this section we study extensions of $S5_2$ with respect to interaction axioms that can be expressed as:

Interaction Axioms	*Completeness*	*Lemmas of reference*	*Notes*
$\Box_1 p \Rightarrow \Box_1 p$	—	–	–
$\Box_1 p \Rightarrow \Diamond_1 p$	—	–	$\vdash p \Rightarrow \Diamond_1 p$
$\Box_1 p \Rightarrow \Box_2 p$	$\sim_2 \subseteq \sim_1$	4.1 and 4.2	–
$\Box_1 p \Rightarrow \Diamond_2 p$	—	–	$\vdash p \Rightarrow \Diamond_2 p$
$\Diamond_1 p \Rightarrow \Box_1 p$	$\sim_1 = id_W$	4.3 and 4.4	–
$\Diamond_1 p \Rightarrow \Diamond_1 p$	—	–	–
$\Diamond_1 p \Rightarrow \Box_2 p$	$\sim_1 = \sim_2 = id_W$	4.5 and 4.6	–
$\Diamond_1 p \Rightarrow \Diamond_2 p$	$\sim_1 \subseteq \sim_2$	4.7 and 4.8	–
$\Box_2 p \Rightarrow \Box_1 p$	$\sim_1 \subseteq \sim_2$	4.7 and 4.8	–
$\Box_2 p \Rightarrow \Diamond_1 p$	—	–	$\vdash p \Rightarrow \Diamond_1 p$
$\Box_2 p \Rightarrow \Box_2 p$	—	–	–
$\Box_2 p \Rightarrow \Diamond_2 p$	—	–	$\vdash p \Rightarrow \Diamond_2 p$
$\Diamond_2 p \Rightarrow \Box_1 p$	$\sim_1 = \sim_2 = id_W$	4.5 and 4.6	–
$\Diamond_2 p \Rightarrow \Diamond_1 p$	$\sim_2 \subseteq \sim_1$	4.1 and 4.2	–
$\Diamond_2 p \Rightarrow \Box_2 p$	$\sim_2 = id_W$	4.5 and 4.10	–
$\Diamond_2 p \Rightarrow \Diamond_2 p$	—	–	–

Figure 4.1: An exhaustive list of interaction axioms generated by Equation 4.1.

$$\boxdot\phi \Rightarrow \boxdot\phi, \text{ where each occurrence of } \boxdot \text{ is in the set } \{\Box_1, \Box_2, \Diamond_1, \Diamond_2\}. \tag{4.1}$$

The axiom schema 4.1 expresses a relation between knowledge or epistemic possibility of a fact. It expresses a first basic type of interaction between mental attitudes.

In Section 4.2.1 we prove the correspondence and completeness results for the logics obtained by adding to $S5_2$ axioms derivable from Formula 4.1. In Section 4.2.2 we will discuss these results.

4.2.1 Correspondence and completeness

Figure 4.1 shows all the interaction axioms that are expressible as Equation 4.1 together with the results proven in this section. If equivalence frames *already* provide a sound and complete semantic class for the logic no relation is given in the corresponding column.

By going through the table we see that half of the axioms are already theorems of the logic $S5_2$; the proofs for these are trivial. We now analyse the ones which are not provable in the logic $S5_2$.

4.2.1.1 $\Box_1 p \Rightarrow \Box_2 p$

Lemma 4.1. *$F \models \Box_1 p \Rightarrow \Box_2 p$ if and only if F is such that $\sim_2 \subseteq \sim_1$.*

Proof. From right to left; consider any model M such that $\sim_2 \subseteq \sim_1$ and a point w such that $M \models_w \Box_1 p$. So, for every point w' such that $w \sim_1 w'$ we have $M \models_{w'} p$. But $[w]_{\sim_2} \subseteq [w]_{\sim_1}$ and so we have $M \models_w \Box_2 p$.

For the converse, suppose $w \sim_2 w'$ on a frame F, such that $F \models \Box_1 p \Rightarrow \Box_2 p$; it remains to prove that $w \sim_1 w'$. Consider a valuation $\pi(p) = \{w'\}$. So $(F, \pi) \models_w \Diamond_2 p$, but then $(F, \pi) \models_w \Diamond_1 p$, and so, since w' is the only point in which p is satisfied we have $w \sim_1 w'$. $\Box$

Lemma 4.2. *The logic $S5_2 + \{\Box_1 p \Rightarrow \Box_2 p\}$ is sound and complete with respect to equivalence frames such that $\sim_2 \subseteq \sim_1$.*

Proof. Soundness was proven in the first part of Lemma 4.1.

Consider the canonical model $M = (W, \sim_1, \sim_2, \pi)$ for the logic $S5_2 + \{\Box_1 p \Rightarrow \Box_2 p\}$. We know by Theorem 1.23 that $S5_2$ is canonical, i.e. the frame underlying M is an equivalence frame. We prove that the extension $S5_2 + \{\Box_1 p \Rightarrow \Box_2 p\}$ is also canonical.

Suppose $w \sim_2 w'$, with $w, w' \in W$; it remains to show that $w \sim_1 w'$. For this, by Definition 1.21, it suffices to prove that there is a consistent set

$$\{\alpha_1, \ldots, \alpha_m\} \cup \{\beta_i \mid \Box_1 \beta_i \in w\}$$

for if that is the case by the maximal extension lemma (Lemma 1.10) there exists a point in the canonical model M that contains those formulae. By contradiction assume this is not the case; then we can choose some $\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_n$ such that $\vdash \neg(\alpha_1, \ldots, \alpha_m \wedge \beta_1, \ldots, \beta_n)$. Call $\alpha = \wedge_{i=1}^m \alpha_i$ and $\beta = \wedge_{i=1}^n \beta_i$. So $\vdash \neg\alpha \vee \neg\beta$, i.e. $\vdash \beta \Rightarrow \neg\alpha$. But $\Box_1 \beta_i \in w$, for $i = 1, \ldots, n$ and so $\Box_1 \beta \in w$; for similar reasons we have $\alpha \in w'$. Since $\vdash \Box_1 \phi \Rightarrow \Box_2 \phi$, we have $\Box_2 \beta \in w$. But then by axiom T we have $\beta \in w'$ and so it has to be $\neg\alpha \in w'$. But then it would be $\alpha \notin w'$ which is absurd.

So the set $\{\alpha_1, \ldots, \alpha_m\} \cup \{\beta_i \mid \Box_1 \beta_i \in w\}$ has to be consistent and there is on the canonical model a point w' such that $w \sim_1 w'$. By canonicity the logic $S5_2 + \{\Box_1 p \Rightarrow \Box_2 p\}$ is then complete with respect to this class of frames. $\Box$

4.2.1.2 $\Diamond_1 p \Rightarrow \Box_1 p$

We now investigate the formula $\Diamond_1 p \Rightarrow \Box_1 p$. Since in $S5_n$ we already have $\vdash p \Rightarrow \Diamond_1 p$, the above entails validity of the formula $p \Rightarrow \Box_1 p$. In the mono-modal case the formula $p \Rightarrow \Box p$ is sometimes known as Triv, because it causes the collapse of modal formulae onto their propositional calculus transforms, obtained simply by removing any modal operator from a formula (see Section 1.2.3 and [HC96], page 66 for details). In the multi-modal case this behaviour will be implied only for the modal fragment that the formula refers to. We can prove the following.

Lemma 4.3. *$F \models \Diamond_1 p \Rightarrow \Box_1 p$ if and only if $F = (W, \sim_1, \sim_2)$ is such that $\sim_1 = id_W$.*

Proof. From left to right. Consider a frame F and suppose there exist two points $w, w' \in W$ such that $w \sim_1 w'$. Consider a valuation π such that $\pi(p) = \{w'\}$. We have $(F, \pi) \models_w \Diamond_1 p$. Then $(F, \pi) \models_w \Box_1 p$, and since F is reflexive this implies that $(F, \pi) \models_w p$, which is absurd unless $w = w'$.

From right to left. Consider any equivalence model $M = (W, \sim_1, \sim_2, \pi)$ such that $M \models_w \Diamond_1 p$. Then there exists a point $w' \in W$ such that $w \sim_1 w'$ and $M \models_{w'} p$. But since $\sim_1 = id_W$, then it must be that $w = w'$ and so $M \models_w \Box_1 p$. $\Box$

Completeness also follows.

Lemma 4.4. *The logic $S5_2+\{\Diamond_1 p \Rightarrow \Box_1 p\}$ is sound and complete with respect to equivalence frames such that $\sim_1 = id_W$.*

Proof. Soundness was proven in the second part of Lemma4.3.

We prove that the logic $S5_2 + \{\Diamond_1 p \Rightarrow \Box_1 p\}$ is canonical. Consider the canonical model M and suppose, by contradiction, that $\sim_1 \neq id_W$ on the canonical frame. So there exist two points $w, w' \in W$ such that there is at least a formula $\alpha \in \mathcal{L}$ such that $\alpha \in w, \alpha \notin w'$ and $w \sim_1 w'$. So we have $M \models_w \Diamond_1 \alpha$, and then by $\vdash \Diamond_1 p \Rightarrow \Box_1 p$ we have $M \models_w \Box_1 \alpha$. But this is absurd because $\sim_1$ is reflexive and we have $\alpha \notin w$. So, $\sim_1 = id_W$ and therefore the logic is canonical. $\square$

Note that in the logic $S5_2 + \{\Diamond_1 p \Rightarrow \Box_1 p\}$ any modal operators for agent 1 can be effectively removed from a formula of the language. In fact it can be easily proved that $\vdash_{S5_2+\{\Diamond_1 p \Rightarrow \Box_1 p\}} p \Leftrightarrow \Diamond_1 p$ and $\vdash_{S5_2+\{\Diamond_1 p \Rightarrow \Box_1 p\}} \Diamond_1 p \Leftrightarrow \Box_1 p$. We will discuss this point in more detail in Section 4.2.2.

4.2.1.3 $\Diamond_1 p \Rightarrow \Box_2 p$

Observe that the converse of this formula expresses the same property with swapped indexes. Since in $S5_2$ we already have reflexivity, this formula is powerful enough to make the whole system collapse into Triv as we prove in the following:

Lemma 4.5. *$F \models \Diamond_1 p \Rightarrow \Box_2 p$ if and only if $F = (W, \sim_1, \sim_2)$ is such that $\sim_1 = \sim_2 = id_W$.*

Proof. From left to right. We prove that it cannot be that $\sim_1 \neq id_W$; the proof for the other relation is equivalent by using the contrapositive of the axiom. Suppose there exist two points $w, w' \in W$ on a frame F such that $w \sim_1 w'$ and consider a valuation π such that $\pi(p) = \{w'\}$. We have $(F, \pi) \models_w \Diamond_1 p$. Then $(F, \pi) \models_w \Box_2 p$, and since F is reflexive this implies that $(F, \pi) \models_w p$, which is absurd, unless $w = w'$.

From right to left. Consider any equivalence model $M = (W, \sim_1, \sim_2, \pi)$ such that $M \models_w \Diamond_1 p$. Then there exists a point $w' \in W$ such that $w \sim_1 w'$ and $M \models_{w'} p$. But since $\sim_1 = \sim_2 = id_W$, then it must be that $w = w'$ and so $M \models_w \Box_2 p$. $\square$

The logic $S5_2 + \{\Diamond_1 p \Rightarrow \Box_2 p\}$ is also complete with respect to the class of frames above.

Lemma 4.6. *The logic $S5_2+\{\Diamond_1 p \Rightarrow \Box_2 p\}$ is sound and complete with respect to equivalence frames such that $\sim_1 = \sim_2 = id_W$.*

Proof. Soundness was proven in the second part of Lemma 4.5.

We prove that the logic $S5_2 + \{\Diamond_1 p \Rightarrow \Box_2 p\}$ is canonical. Consider the canonical model M and suppose, by contradiction, that $\sim_1 \neq id_W$ on the canonical frame. So there exist two points $w, w' \in W$ such that there is at least a formula $\alpha \in \mathcal{L}$ such that $\alpha \in w, \alpha \notin w'$ and $w \sim_1 w'$. So we have $M_w \models \Diamond_1 \alpha$, and then by $\vdash \Diamond_1 p \Rightarrow \Box_2 p$ we have $M_w \models \Box_2 \alpha$. But this is absurd because $\sim_1$ is reflexive and we have $\alpha \notin w$. So, $\sim_1 = id_W$. In a similar way, we can prove that $\sim_2$ is also the identity on W is also the identity on W. The logic is then canonical and complete with respect to the frames above. $\square$

Indeed, the logic $S5_2 + \{\Diamond_1 p \Rightarrow \Box_2 p\}$ is equivalent to Triv_2. In fact, it can be proven straightforwardly that $\vdash_{S5_2+\{\Diamond_1 p \Rightarrow \Box_2 p\}} p \Leftrightarrow \Diamond_1 p$, $\vdash_{S5_2+\{\Diamond_1 p \Rightarrow \Box_2 p\}} \Diamond_1 p \Leftrightarrow \Box_1 p$, $\vdash_{S5_2+\{\Diamond_1 p \Rightarrow \Box_2 p\}} \Box_1 p \Leftrightarrow \Diamond_2 p$ and $\vdash_{S5_2+\{\Diamond_1 p \Rightarrow \Box_2 p\}} \Diamond_2 p \Leftrightarrow \Box_2 p$.

4.2.1.4 Remaining axioms

We now examine the interaction axiom $\Diamond_1 p \Rightarrow \Diamond_2 p$. In Lemma 4.1 and Lemma 4.2 we examined the axiom for the two indexes swapped. So without repeating the proofs we can simply state the following:

Lemma 4.7. *$F \models \Box_2 p \Rightarrow \Box_1 p$ if and only if F is such that $\sim_1 \subseteq \sim_2$.*

Lemma 4.8. *The logic $S5_2 + \{\Box_2 p \Rightarrow \Box_1 p\}$ is sound and complete with respect to equivalence frames such that $\sim_1 \subseteq \sim_2$.*

The next formula we analyse is $\Diamond_2 p \Rightarrow \Box_2 p$. We examined this axiom with respect to agent 1 in Lemma 4.3 and Lemma 4.4. So, without repeating the proofs we report the two results:

Lemma 4.9. *$F \models \Diamond_2 p \Rightarrow \Box_2 p$ if and only if $F = (W, \sim_1, \sim_2)$ is such that $\sim_2 = id_W$.*

Lemma 4.10. *The logic $S5_2 + \{\Diamond_2 p \Rightarrow \Box_2 p\}$ is sound and complete with respect to equivalence frames such that $\sim_2 = id_W$.*

4.2.2 Discussion

In Section 4.2.1 we showed that out of 16 possible interaction axioms of the form of Equation 4.1 only 5 of them lead to a different proper extension of $S5_2$. For these 5 logics we proved correspondence and completeness. In particular since all the logics were proven to be canonical we have the more general result.

Theorem 4.11. *All the logics $S5_2 + \{\phi\}$, where ϕ is the conjunction of formulae expressible as Equation 4.1 are complete with respect to the intersection of the respective classes of frames.*

The logics are ordered as in Figure 4.2.

Proof. It follows from all the canonicity results proved in Section 4.2.1. Proving the relation between the logics is straightforward. $\Box$

Figure 4.2 shows the relations between all the logics discussed in Section 4.2.1 and the logic $S5_2 + \{\Box_1 p \Leftrightarrow \Box_2 p\}$ that can be obtained by taking the union of $S5_2 + \{\Diamond_1 p \Rightarrow \Diamond_2 p\}$ and $S5_2 + \{\Diamond_2 p \Rightarrow \Diamond_1 p\}$. The lines in the figure represent set inclusion between logics, i.e. the logics are ordered in terms of how many formulae they contain. For example it is straightforward to prove that if $\vdash_{S5_2+\{\Diamond_1 p \Rightarrow \Box_1 p\}} \phi$ then $\vdash_{S5_2+\{\Diamond_1 p \Rightarrow \Box_2 p\}} \phi$. The pictured relations between the logics are reflexive and transitive.

In Figure 4.3 the relation between the different classes of frames that we examined in this section is shown.

Although possibly never presented systematically, the results of Theorem 4.11 and of Section 4.2.1 are quite well known. The most important logic is probably the one that forces the knowledge of an agent to be a subset of the knowledge of another. In Section 1.4.1 we have discussed two agent scenarios in which this can be proven useful. By combining two axioms of this kind for both agents we obtain the logic $S5_2 + \{\Box_1 p \Leftrightarrow \Box_2 p\}$, in which both agents have exactly the same knowledge base. Note the central position of this logic in Figure 4.2 reflecting the symmetry of the group of agents in this case.

Stronger logics can be defined by assuming that the modal component for one of the agents collapses onto the propositional calculus. When this happens we are in a situation in

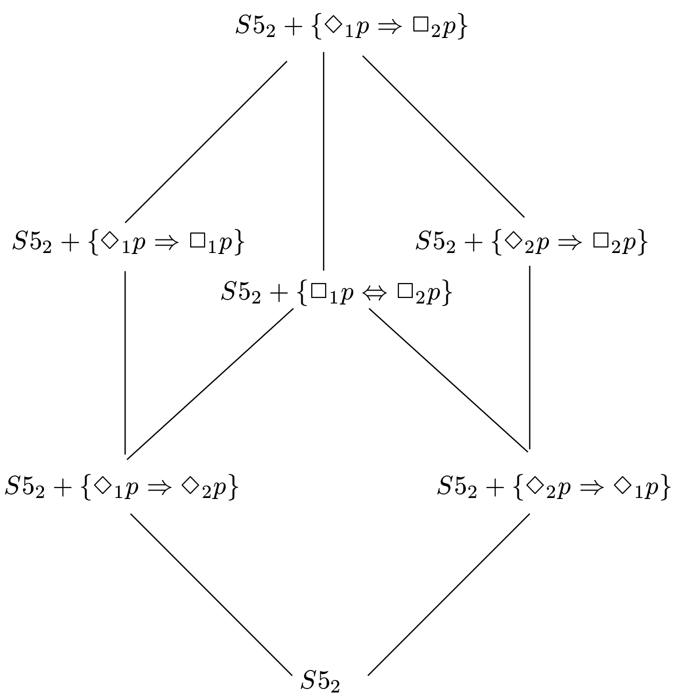

Figure 4.2: The proper extensions of $S5_2$ that can be obtained by adding axioms of the shape of Formula 4.1.

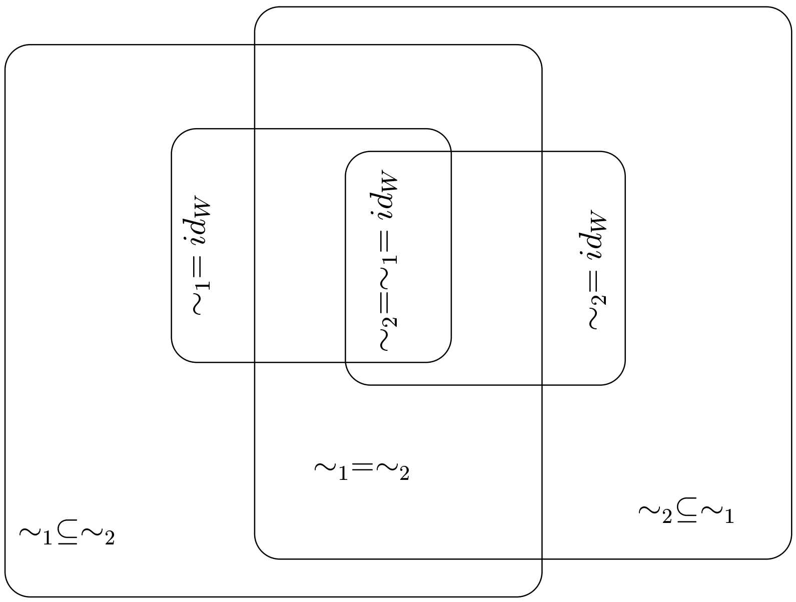

Figure 4.3: The classes of frames corresponding to the proper extensions of $S5_2$ obtained by adding axioms of the shape of Formula 4.1.

which "being possible according to one agent" is equivalent to "being known" and this in turn is equivalent to "being true". It is clear that this is indeed a very strong constraint which limits the expressivity of our language. Still these logics can be proven to be consistent.

The strongest consistent logic is Triv_2 that can be defined from $\mathrm{S5}_2$ by adding the axiom $\Diamond_1 p \Rightarrow \Box_2 p$ to $\mathrm{S5}_2$ or equivalently by adding both $\Diamond_1 p \Rightarrow \Box_1 p$ and $\Diamond_2 p \Rightarrow \Box_2 p$. In this logic the two agents have equal knowledge that is equivalent to the truth on the world of evaluation.

4.3 Interaction axioms of the form $\boxdot p \Rightarrow \boxdot \boxdot p$

In section 4.2.1 we have studied a relatively simple class of interactions. We noticed that there are only 5 interaction axioms that we can add to $\mathrm{S5}_2$ without collapsing to the uninteresting system logic system Triv, in which knowledge of a fact collapses to the truth of the fact.

In this section we study logics that can be defined by adding to $\mathrm{S5}_2$ more complex interaction axioms of the shape:

$$\boxdot \phi \Rightarrow \boxdot \boxdot \phi, \text{ where each occurrence of } \boxdot \text{ is in the set } \{\Box_1, \Box_2, \Diamond_1, \Diamond_2\}. \tag{4.2}$$

Equation 4.2 expresses $4 \times 4 \times 4 = 64$ different formulae.

4.3.1 Correspondence and completeness

We need to examine the correspondence and completeness problem for 32 different axioms (those with an antecedent referring to agent 1); results for the remaining 32 can be inferred from these by consideration of symmetry. We further sub-divide these axioms in two classes. The first class of axioms contains $\Diamond_1 p$ as antecedent, the second $\Box_1 p$.

A quite surprising result that we will find in the following sections is that most of the logics obtained from $\mathrm{S5}_2$ by adding axiom schemas of the shape of Equation 4.2 are actually equivalent either to $\mathrm{S5}_2$ itself or to one of the logics explored in Section 4.2.1.

Proving that these logics are equivalent is sometimes not obvious; still it would be somewhat tedious to report all of them in the main track. What we will do is to discuss of them in this chapter and leave the proofs for the other equivalences in the Appendix A. The same applies for the correspondences.

4.3.1.1 Interaction axioms of the form $\Diamond_1 p \Rightarrow \boxdot \boxdot p$

We begin by analysing all the formulae that follow from Equation 4.2 by using $\Diamond_1 p$ as antecedent of the formula. We have the table of Figure 4.4 in which the results of this section are summarised.

Theorem 4.12. *For all the axiom schemas ϕ in Figure 4.4 we have that $F \models \phi$ if and only if F satisfies the corresponding property in the figure.*

Proof. In this chapter we only proof the correspondence for the axiom $\Diamond_1 p \Rightarrow \Diamond_2 \Box_1 p$ (see Lemma 4.14). All the other proofs are standard and can be found in the Appendix in Section A.1.1. □

The second result is that we have completeness for all the logics of Figure 4.4.

Interaction Axioms	*Completeness*	*Lemmas of Reference*	*Notes*
$\Diamond_1 p \Rightarrow \Diamond_1 \Box_1 p$	$\sim_1 = id_W$	4.3 and 4.4	$\vdash \Box_1 p \Leftrightarrow \Diamond_1 \Box_1 p$
$\Diamond_1 p \Rightarrow \Diamond_1 \Box_2 p$	$\sim_2 = id_W$	A.1 and A.2	–
$\Diamond_1 p \Rightarrow \Diamond_1 \Diamond_1 p$	—	–	$\vdash \Diamond_1 p \Leftrightarrow \Diamond_1 \Diamond_1 p$
$\Diamond_1 p \Rightarrow \Diamond_1 \Diamond_2 p$	—	–	$\vdash p \Rightarrow \Diamond_2 p$
$\Diamond_1 p \Rightarrow \Box_1 \Box_1 p$	$\sim_1 = id_W$	4.3 and 4.4	$\vdash \Box_1 p \Leftrightarrow \Box_1 \Box_1 p$
$\Diamond_1 p \Rightarrow \Box_1 \Box_2 p$	$\sim_1 = \sim_2 = id_W$	A.3 and A.4	–
$\Diamond_1 p \Rightarrow \Box_1 \Diamond_1 p$	—	–	$\vdash \Diamond_1 p \Leftrightarrow \Box_1 \Diamond_1 p$
$\Diamond_1 p \Rightarrow \Box_1 \Diamond_2 p$	$\sim_1 \subseteq \sim_2$	A.5 and A.6	–
$\Diamond_1 p \Rightarrow \Diamond_2 \Box_1 p$	$\sim_1 = id_W$	4.14 and 4.15	–
$\Diamond_1 p \Rightarrow \Diamond_2 \Box_2 p$	$\sim_1 = \sim_2 = id_W$	4.5 and 4.6	$\vdash \Box_2 p \Leftrightarrow \Diamond_2 \Box_2 p$
$\Diamond_1 p \Rightarrow \Diamond_2 \Diamond_1 p$	—	–	$\vdash p \Rightarrow \Diamond_2 p$
$\Diamond_1 p \Rightarrow \Diamond_2 \Diamond_2 p$	$\sim_1 \subseteq \sim_2$	4.7 and 4.8	$\vdash \Diamond_2 p \Leftrightarrow \Diamond_2 \Diamond_2 p$
$\Diamond_1 p \Rightarrow \Box_2 \Box_1 p$	$\sim_2 = \sim_1 = id_W$	A.7 and A.8	–
$\Diamond_1 p \Rightarrow \Box_2 \Box_2 p$	$\sim_1 = \sim_2 = id_W$	4.5 and 4.6	$\vdash \Box_2 p \Leftrightarrow \Box_2 \Box_2 p$
$\Diamond_1 p \Rightarrow \Box_2 \Diamond_1 p$	$\sim_2 \subseteq \sim_1$	A.9 and A.10	–
$\Diamond_1 p \Rightarrow \Box_2 \Diamond_2 p$	$\sim_1 \subseteq \sim_2$	4.7 and 4.8	$\vdash \Diamond_2 p \Leftrightarrow \Box_2 \Diamond_2 p$

Figure 4.4: An exhaustive list of interaction axioms generated by Equation 4.2 in the case the antecedent is equal to $\Diamond_1 p$.

Theorem 4.13. *All the logics in Figure 4.4 are sound and complete with respect to the class of equivalence frames satisfying the corresponding property.*

Proof. Soundness can be checked straightforwardly.

For completeness, consider any logic $S5_2 + \{\phi\}$, where ϕ is an axiom in Figure 4.4. We have two cases.

- $\vdash_{S5_2} \phi$. In this case, we obviously have that $S5_2 + \{\phi\}$ is equivalent to $S5_2$ and so the completeness of the logic $S5_2 + \{\phi\}$ with respect to equivalence frames follows.

  Showing that axiom ϕ can be proven in $S5_2$ is immediate in most cases by using the notes in Figure 4.4.

- $\nvdash_{S5_2} \phi$. In this case, although the logic $S5_2 + \{\phi\}$ is a proper extension of $S5_2$, it can be proven equivalent to a logic $S5_2 + \{\psi\}$ for some axiom ψ[1] examined in Section 4.2.1. The equivalence between $S5_2 + \{\phi\}$ and $S5_2 + \{\psi\}$, i.e. that $\vdash_{S5_2+\{\phi\}} \alpha$ if and only if $\vdash_{S5_2+\{\psi\}} \alpha$, follows once we have $\vdash_{S5_2+\{\phi\}} \psi$ and $\vdash_{S5_2+\{\psi\}} \phi$; in fact in this case any proof of α in one logic can be repeated in the other. All the proofs for these equivalences between axioms can be found in the Appendix (in Section A.1.1) and they follow the numbering given in Figure 4.4. As an interesting example the proof for $\vdash_{S5_2+\{\Diamond_1 p \Rightarrow \Diamond_2 \Box_1 p\}} \Diamond_1 p \Rightarrow \Box_1 p$ and $\vdash_{S5_2+\{\Diamond_1 p \Rightarrow \Box_1 p\}} \Diamond_1 p \Rightarrow \Diamond_2 \Box_1 p$ is given in Lemma 4.15 in this chapter.

  Now, since $S5_2 + \{\psi\}$ was proven complete with respect to equivalence frames satisfying property P_ψ (the results are in Figure 4.1), the completeness of $S5_2 + \{\phi\}$ with respect to equivalence P_ψ frames also follows.

$\square$

[1] The axiom ψ can be found in the corresponding lemmas of reference.

4.3.1.1.1 $\Diamond_1 p \Rightarrow \Diamond_2 \Box_1 p$. As a sample of the proofs of Section A.1.1 of the Appendix, the correspondence proof and the axiom-dependent part for the completeness proof of Theorem 4.13 for axiom $\Diamond_1 p \Rightarrow \Diamond_2 \Box_1 p$ are reported here.

Lemma 4.14. *$F \models \Diamond_1 p \Rightarrow \Diamond_2 \Box_1 p$ if and only if $F = (W, \sim_1, \sim_2)$ is such that $\sim_1 = id_W$.*

Proof. From left to right. Suppose there exist two points $w, w' \in W$ such that $w \sim_1 w'$. Consider a valuation π such that $\pi(p) = \{w'\}$. We have $(F, \pi) \models_w \Diamond_1 p$. Then $(F, \pi) \models_{w'} \Diamond_2 \Box_1 p$, and since p is true only at w', which is related to w by relation $\sim_1$, then it must be that $[w]_{\sim_1} = \{w\}$.

From right to left. Consider any equivalence model $M = (W, \sim_1, \sim_2, \pi)$ such that $M \models_w \Diamond_1 p$. Then there exists a point $w' \in W$ such that $w \sim_1 w'$ and $M \models_{w'} p$. But since $\sim_1 = id_W$, we have $w = w'$. So $M \models_w \Box_1 p$ and so $M \models_w \Diamond_2 \Box_1 p$. $\Box$

For completeness all it remains to show from Theorem 4.13 is that the logic is equivalent to a logic sound and complete with respect to equivalence frames such that $\sim_1 = id_W$. It is enough to prove the following lemma.

Lemma 4.15. *$\vdash_{S5_2 + \{\Diamond_1 p \Rightarrow \Diamond_2 \Box_1 p\}} \Diamond_1 p \Rightarrow \Box_1 p$ and $\vdash_{S5_2 + \{\Diamond_1 p \Rightarrow \Box_1 p\}} \Diamond_1 p \Rightarrow \Diamond_2 \Box_1 p$.*

Proof. First part. Suppose $\Diamond_1 p \Rightarrow \Diamond_2 \Box_1 p$; so $\Box_2 \Diamond_1 p \Rightarrow \Box_1 p$. Substitute the term $(p \Rightarrow \Box_1 p)$ for p uniformly in the axiom above; we obtain $\Box_2 \Diamond_1 (p \Rightarrow \Box_1 p) \Rightarrow \Box_1 (p \Rightarrow \Box_1 p)$. We prove that the antecedent of this formula is a theorem of $S5_2$. In fact we have $\neg \Box_1 p \vee \Box_1 p$ so by Lemma 1.30 we have $\Diamond_1 \neg p \vee \Diamond_1 \Box_1 p$. Now since, as it can easily be verified, diamond distributes over logical or, we have $\Diamond_1 (\neg p \vee \Box_1 p)$, which by necessitating by $\Box_2$ leads to $\Box_2 \Diamond_1 (\neg p \vee \Box_1 p)$. So, it follows that $\Box_1 (p \Rightarrow \Box_1 p)$, which gives $p \Rightarrow \Box_1 p$. Now, from this formula it follows from Lemma 1.29 that $\Diamond_1 p \Rightarrow \Diamond_1 \Box_1 p$, which is equivalent to $\Diamond_1 p \Rightarrow \Box_1 p$.

Second part. Suppose $\Diamond_1 p \Rightarrow \Box_1 p$. By the axiom T we then obtain $\Diamond_1 p \Rightarrow \Diamond_2 \Box_1 p$.

$\Box$

4.3.1.2 Interaction axioms of the form $\Box_1 p \Rightarrow \boxdot \boxdot p$

We now move to the study of instances of Equation 4.2 where the antecedent is of the form $\Box_1 p$. Given the strength of the antecedent we do not have many results to prove in this subsection. In fact, most of the axioms are already theorems of $S5_2$.

As in the previous section, here we only prove the results for one interesting axiom and refer the reader interested in others to the proofs given in the Appendix.

We have the following two theorems.

Theorem 4.16. *For all the axiom schemas ϕ in Figure 4.5 we have that $F \models \phi$ if and only if F satisfies the corresponding property in the figure.*

Proof. In Lemma 4.18 we prove the result for the axiom $\Box_1 p \Rightarrow \Diamond_1 \Box_2 p$. All the others can be found in Section A.1.2. $\Box$

Theorem 4.17. *All the logics in Figure 4.5 are sound and complete with respect to the class of equivalence frames satisfying the corresponding property.*

Proof. In Lemma 4.19 we prove canonicity of the logic $S5_2 + \{\Box_1 p \Rightarrow \Diamond_1 \Box_2 p\}$. The rest of the theorem can be proven similarly to Theorem 4.13. All the relevant proofs of equivalence between axioms are in Section A.1.2. $\Box$

Interaction Axioms	*Completeness*	*Lemmas of Ref.*	*Notes*
$\Box_1 p \Rightarrow \Diamond_1 \Box_1 p$	—	–	$\vdash \Diamond_1 \Box_1 p \Leftrightarrow \Box_1 p$
$\Box_1 p \Rightarrow \Diamond_1 \Box_2 p$	$\forall w \exists w' \in [w]_{\sim_1} : [w']_{\sim_2} \subseteq [w]_{\sim_1}$	4.18 and 4.19	–
$\Box_1 p \Rightarrow \Diamond_1 \Diamond_1 p$	—	–	$\vdash \Diamond_1 p \Leftrightarrow \Diamond_1 \Diamond_1 p$
$\Box_1 p \Rightarrow \Diamond_1 \Diamond_2 p$	—	–	$\vdash p \Rightarrow \Diamond_2 p$
$\Box_1 p \Rightarrow \Box_1 \Box_1 p$	—	–	$\vdash \Box_1 p \Leftrightarrow \Box_1 \Box_1 p$
$\Box_1 p \Rightarrow \Box_1 \Box_2 p$	$\sim_2 \subseteq \sim_1$	A.11 and A.12	–
$\Box_1 p \Rightarrow \Box_1 \Diamond_1 p$	—	–	$\vdash \Diamond_1 p \Leftrightarrow \Box_1 \Diamond_1 p$
$\Box_1 p \Rightarrow \Box_1 \Diamond_2 p$	—	–	$\vdash p \Rightarrow \Diamond_2 p$
$\Box_1 p \Rightarrow \Diamond_2 \Box_1 p$	—	–	$\vdash p \Rightarrow \Diamond_2 p$
$\Box_1 p \Rightarrow \Diamond_2 \Box_2 p$	$\sim_2 \subseteq \sim_1$	4.1 and 4.2	$\vdash \Diamond_2 p \Leftrightarrow \Diamond_2 \Box_2 p$
$\Box_1 p \Rightarrow \Diamond_2 \Diamond_1 p$	—	–	$\vdash p \Rightarrow \Diamond_2 p$
$\Box_1 p \Rightarrow \Diamond_2 \Diamond_2 p$	—	–	$\vdash \Diamond_2 p \Leftrightarrow \Diamond_2 \Diamond_2 p$
$\Box_1 p \Rightarrow \Box_2 \Box_1 p$	$\sim_2 \subseteq \sim_1$	A.13 and A.14	–
$\Box_1 p \Rightarrow \Box_2 \Box_2 p$	$\sim_2 \subseteq \sim_1$	4.1 and 4.2	$\vdash \Box_2 p \Leftrightarrow \Box_2 \Box_2 p$
$\Box_1 p \Rightarrow \Box_2 \Diamond_1 p$	$\sim_2 \subseteq \sim_1$	A.15 and A.16	–
$\Box_1 p \Rightarrow \Box_2 \Diamond_2 p$	—	–	$\vdash \Diamond_2 p \Leftrightarrow \Box_2 \Diamond_2 p$

Figure 4.5: An exhaustive list of interaction axioms generated by Equation 4.2 in the case the antecedent is equal to $\Box_1 p$.

4.3.1.2.1 $\Box_1 p \Rightarrow \Diamond_1 \Box_2 p$. This axiom generates a class of frames that we have not seen before. We start by proving correspondence.

Lemma 4.18. *$F \models \Box_1 p \Rightarrow \Diamond_1 \Box_2 p$ if and only if F is such that $\forall w \exists w' \in [w]_{\sim_1} : [w']_{\sim_2} \subseteq [w]_{\sim_1}$.*

Proof. From right to left; consider any model M and a point w in it such that $M \models_w \Box_1 p$. So, for every point w' such that $w \sim_1 w'$ we have $M \models_{w'} p$. But, by assumption, there exists a point $w' \in [w]_{\sim_1}$ such that $[w']_{\sim_2} \subseteq [w]_{\sim_1}$. So, p holds at any point of the equivalence class $[w']_{\sim_2}$, and so $M \models_{w'} \Box_2 p$. Therefore $M \models_w \Diamond_1 \Box_2 p$.

For the converse, suppose the relational property above does not hold. Then there exists a frame F and a point w in F such that for any $w' \in [w]_{\sim_1}$ we have $[w']_{\sim_2} \not\subseteq [w]_{\sim_1}$, i.e. we have the existence of a point $w'' \in [w']_{\sim_2}$ such that $w'' \notin [w]_{\sim_1}$. Consider a valuation π such that $\pi(p) = \{w' \mid w \sim_1 w'\}$. We have $(F, \pi) \models_w \Box_1 p$ and $(F, \pi) \not\models_{w''} p$. So $(F, \pi) \not\models_{w'} \Box_2 p$. So we have $(F, \pi) \not\models_w \Diamond_1 \Box_2 p$ which is absurd. $\Box$

Completeness with respect to the same class of frames also follows.

Lemma 4.19. *The logic $S5_2 + \{\Box_1 p \Rightarrow \Diamond_1 \Box_2 p\}$ is sound and complete with respect to equivalence frames satisfying the property $\forall w \exists w' \in [w]_{\sim_1} : [w']_{\sim_2} \subseteq [w]_{\sim_1}$.*

Proof. Soundness was proven in first part of Lemma 4.18.

For completeness we prove that the logic $S5_2 + \{\Box_1 p \Rightarrow \Diamond_1 \Box_2 p\}$ is canonical. In order to do that, suppose, by contradiction, that the frame of the canonical model does not satisfy the relational property above. Then, it must be that there exists a point w such that:

$$\forall w' \in [w]_{\sim_1} \exists w'' : w' \sim_2 w'' \text{ and } w \not\sim_1 w''.$$

Call $w'_1, \ldots, w'_n, \ldots$ the points in $[w]_{\sim_1}$, and w''_i the point in $[w'_i]_{\sim_2}$ such that $w \not\sim_1 w''_i$; $i = 1, \ldots, n, \ldots$. Recall from Definition 1.21 that $w \sim_1 w'$ is defined as $\forall \alpha \in \mathcal{L}\,(\Box_i \alpha \in w$ implies $\alpha \in w')$; $w \not\sim_j w'$ is defined as $\exists \alpha \in \mathcal{L}\,(\Box_j \alpha \in w$ and $\neg\alpha \in w')$. So we can find some formulae $\alpha_i \in \mathcal{L}; i = 1, \ldots, n, \ldots$ such that $\Box_1 \alpha_i \in w, \alpha_i \in w'_i, \neg\alpha_i \in w''_i; i = 1, \ldots, n, \ldots$ Call $\alpha = \wedge_{i=1}^{n} \alpha_i$; we have $\Box_1 \alpha_i \in w; i = 1, \ldots, n, \ldots$. So $\Box_1 \alpha \in w$. But $\neg\alpha \in w''_i, i = 1, \ldots, n, \ldots$. So $\Diamond_2 \neg\alpha \in w'_i$ for every i in $\{1, \ldots, n, \ldots\}$. So $\Box_1 \Diamond_2 \neg\alpha \in w$, i.e. $\neg\Diamond_1 \Box_2 \alpha \in w$. But $\Box_1 \alpha \in w$ and $\vdash \Box_1 \alpha \Rightarrow \Diamond_1 \Box_2 \alpha$, so w would be inconsistent. Therefore the canonical frame must satisfy the property above and the logic is complete with respect to equivalence frames satisfying the property $\forall w \exists w' \in [w]_{\sim_1} : [w']_{\sim_2} \subseteq [w]_{\sim_1}$. $\Box$

4.3.1.3 Interaction axioms of the form $\Diamond_2 p \Rightarrow \boxdot \boxdot p$ and $\Box_2 p \Rightarrow \boxdot \boxdot p$

We can now take stock of the position we have achieved in Section 4.3.1.2 and Section 4.3.1.1 and present the completeness results for Equation 4.2 in the case the first modal operator is indexed by 2. The proofs are completely equivalent to the ones we have seen before, so we just present the final tables in Figure 4.6 and Figure 4.7 that will be useful later on.

In Figure 4.6 and Figure 4.7 some of the lemmas cited do not refer to the axiom in analysis but to one with the indexes uniformly swapped. These have been analysed in the previous section and we do not repeat here the proofs.

Corollary 4.20. *For all the axiom schemas ϕ in Figure 4.6 and Figure 4.7 we have that $F \models \phi$ if and only if F satisfies the corresponding property in the figure.*

Corollary 4.21. *All the logics in Figure 4.6 and Figure 4.7 are sound and complete with respect to the class of equivalence frames satisfying the corresponding property.*

4.3.2 Discussion

In this section we have analysed the class of interaction axioms expressed by Equation 4.2. This represents the interactions between the state of knowledge of one agents and the knowledge of an agent referred to the knowledge of the other agent. We have proved correspondence properties and completeness for all the logics above.

Given the fact that all the logics were proven to be canonical we have the general result:

Corollary 4.22. *All the logics $S5_2 + \{\phi\}$, where ϕ is the conjunction of formulae expressible as Equation 4.2 are complete with respect to the intersection of the corresponding classes of frames shown in Figure 4.4, Figure 4.5, Figure 4.6, Figure 4.7.*

Proof. It follows from the canonicity proofs that compose Theorem 4.13, Theorem 4.17 and Corollary 4.21. $\Box$

Among all these axioms, the most intuitive ones in terms of knowledge are probably $\Box_1 p \Rightarrow \Box_2 \Box_1 p$ and its "dual" $\Box_2 p \Rightarrow \Box_1 \Box_2 p$, representing scenarios in which agent 1 knows that agent 2 knows something every time this happens to be the case. It is interesting to see that this is equivalent to agent 1 knowing everything known by agent 2. The reader can explore other interesting equivalences from the table above and refer to the appropriate cited lemmas for the proofs.

Interaction Axioms	*Completeness*	*Lemmas of Reference*	*Notes*
$\Diamond_2 p \Rightarrow \Diamond_1 \Box_1 p$	$\sim_1 = \sim_2 id_W$	4.5 and 4.6	$\vdash \Box_1 p \Leftrightarrow \Diamond_1 \Box_1 p$
$\Diamond_2 p \Rightarrow \Diamond_1 \Box_2 p$	$\sim_2 = id_W$	4.14 and 4.15	–
$\Diamond_2 p \Rightarrow \Diamond_1 \Diamond_1 p$	$\sim_2 \subseteq \sim_1$	4.1 and 4.2	$\vdash \Diamond_1 p \Leftrightarrow \Diamond_1 \Diamond_1 p$
$\Diamond_2 p \Rightarrow \Diamond_1 \Diamond_2 p$	—	–	$\vdash p \Rightarrow \Diamond_1 p$
$\Diamond_2 p \Rightarrow \Box_1 \Box_1 p$	$\sim_1 = \sim_2 = id_W$	4.5 and 4.6	$\vdash \Box_1 p \Leftrightarrow \Box_1 \Box_1 p$
$\Diamond_2 p \Rightarrow \Box_1 \Box_2 p$	$\sim_1 = \sim_2 = id_W$	A.7 and A.8	See $\Diamond_1 p \Rightarrow \Box_2 \Box_1 p$
$\Diamond_2 p \Rightarrow \Box_1 \Diamond_1 p$	$\sim_2 \subseteq \sim_1$	4.1 and 4.2	$\vdash \Diamond_1 p \Leftrightarrow \Box_1 \Diamond_1 p$
$\Diamond_2 p \Rightarrow \Box_1 \Diamond_2 p$	$\sim_1 \subseteq \sim_2$	A.9 and A.10	See $\Diamond_1 p \Rightarrow \Box_2 \Diamond_1 p$
$\Diamond_2 p \Rightarrow \Diamond_2 \Box_1 p$	$\sim_1 = id_W$	A.1 and A.2	–
$\Diamond_2 p \Rightarrow \Diamond_2 \Box_2 p$	$\sim_2 = id_W$	4.9 and 4.10	$\vdash \Box_2 p \Leftrightarrow \Diamond_2 \Box_2 p$
$\Diamond_2 p \Rightarrow \Diamond_2 \Diamond_1 p$	—	–	$\vdash p \Rightarrow \Diamond_1 p$
$\Diamond_2 p \Rightarrow \Diamond_2 \Diamond_2 p$	—	–	$\vdash \Diamond_2 p \Leftrightarrow \Diamond_2 \Diamond_2 p$
$\Diamond_2 p \Rightarrow \Box_2 \Box_1 p$	$\sim_2 = \sim_1 = id_W$	A.3 and A.4	See $\Diamond_1 p \Rightarrow \Box_1 \Box_2 p$
$\Diamond_2 p \Rightarrow \Box_2 \Box_2 p$	$\sim_2 = id_W$	4.9 and 4.10	$\vdash \Box_2 p \Leftrightarrow \Box_2 \Box_2 p$
$\Diamond_2 p \Rightarrow \Box_2 \Diamond_1 p$	$\sim_2 \subseteq \sim_1$	A.5 and A.6	See $\Diamond_1 p \Rightarrow \Box_1 \Diamond_2 p$
$\Diamond_2 p \Rightarrow \Box_2 \Diamond_2 p$	—	–	$\vdash \Diamond_2 p \Leftrightarrow \Box_2 \Diamond_2 p$

Figure 4.6: An exhaustive list of interaction axioms generated by Equation 4.1 in the case the antecedent is equal to $\Diamond_2 p$.

Interaction Axioms	*Completeness*	*Lemmas of Ref.*	*Notes*
$\Box_2 p \Rightarrow \Diamond_1 \Box_1 p$	$\sim_1 \subseteq \sim_2$	–	$\vdash \Diamond_1 \Box_1 p \Leftrightarrow \Box_1 p$
$\Box_2 p \Rightarrow \Diamond_1 \Box_2 p$	—	–	$\vdash p \Rightarrow \Diamond_1 p$
$\Box_2 p \Rightarrow \Diamond_1 \Diamond_1 p$	—	–	$\vdash \Diamond_1 p \Leftrightarrow \Diamond_1 \Diamond_1 p$
$\Box_2 p \Rightarrow \Diamond_1 \Diamond_2 p$	—	–	$\vdash p \Rightarrow \Diamond_1 p$
$\Box_2 p \Rightarrow \Box_1 \Box_1 p$	$\sim_1 \subseteq \sim_2$	4.7 and 4.8	$\vdash \Box_1 p \Leftrightarrow \Box_1 \Box_1 p$
$\Box_2 p \Rightarrow \Box_1 \Box_2 p$	$\sim_1 \subseteq \sim_2$	A.13 and A.14	See $\Box_1 p \Rightarrow \Box_2 \Box_1 p$
$\Box_2 p \Rightarrow \Box_1 \Diamond_1 p$	—	–	$\vdash \Diamond_1 p \Leftrightarrow \Box_1 \Diamond_1 p$
$\Box_2 p \Rightarrow \Box_1 \Diamond_2 p$	$\sim_1 \subseteq \sim_2$	A.15 and A.16	See $\Box_1 p \Rightarrow \Box_2 \Diamond_1 p$
$\Box_2 p \Rightarrow \Diamond_2 \Box_1 p$	$\forall w \exists w' \in [w]_{\sim_2} : [w']_{\sim_1} \subseteq [w]_{\sim_2}$	4.18 and 4.19	See $\Box_1 p \Rightarrow \Diamond_1 \Box_2 p$
$\Box_2 p \Rightarrow \Diamond_2 \Box_2 p$	—	–	$\vdash \Box_2 p \Leftrightarrow \Diamond_2 \Box_2 p$
$\Box_2 p \Rightarrow \Diamond_2 \Diamond_1 p$	—	–	$\vdash p \Rightarrow \Diamond_1 p$
$\Box_2 p \Rightarrow \Diamond_2 \Diamond_2 p$	—	–	$\vdash \Diamond_2 p \Leftrightarrow \Diamond_2 \Diamond_2 p$
$\Box_2 p \Rightarrow \Box_2 \Box_1 p$	$\sim_1 \subseteq \sim_2$	A.11 and A.12	See $\Box_1 p \Rightarrow \Box_1 \Box_2 p$
$\Box_2 p \Rightarrow \Box_2 \Box_2 p$	—	–	$\vdash \Box_2 p \Leftrightarrow \Box_2 \Box_2 p$
$\Box_2 p \Rightarrow \Box_2 \Diamond_1 p$	—	–	$\vdash p \Rightarrow \Diamond_1 p$
$\Box_2 p \Rightarrow \Box_2 \Diamond_2 p$	—	–	$\vdash \Diamond_2 p \Leftrightarrow \Box_2 \Diamond_2 p$

Figure 4.7: An exhaustive list of interaction axioms generated by Equation 4.2 in the case the antecedent is equal to $\Box_2 p$.

A more subtle, independent axiom expressed by Equation 4.2 is the formula

$$\Box_1 p \Rightarrow \Diamond_1 \Box_2 p,$$

which reads "If agent 1 knows p, then he considers possible that agent 2 also knows p". The above is an axiom that regulates a natural kind of "prudence" assumption of agent 1 in terms of what knowledge agent 2 may have. This is meaningful in MAS in which agents have similar characteristics. In these scenarios when an agent knows a fact, it may be appropriate to assume that the other agent, by acquiring the same information from the environment and by following her same reasoning, could have reached the same conclusion. Note that very often humans act as if they followed this axiom.

For the moment we do not add the two new logics to the figure of the previous section; we will discuss the general picture of the logics after proving the results for two modal operators from either side of the implication. Note that axioms of the shape $\boxdot \boxdot p \Rightarrow \boxdot p$ are simply the contrapositive of the ones we analysed here, so they do not need being studied.

4.4 Interaction axioms of the form $\boxdot \boxdot p \Rightarrow \boxdot \boxdot p$

We now discuss the last class of interaction axioms we will see in this chapter, i.e. extensions of $\mathrm{S5}_2$ with interaction axioms expressible as:

$$\boxdot \boxdot \phi \Rightarrow \boxdot \boxdot \phi \text{ where each occurrence of } \boxdot \text{ is in the set } \{\Diamond_1, \Box_1, \Diamond_2, \Box_2\}.$$

4.4.1 Correspondence and completeness

Equation 4.4 generates 256 different axioms. 64 of them have both operators indexed by agent 1 in the antecedent, 64 by agent 2. Both of these, by Lemma 1.30 are equivalent to one of the axioms we examined in Section 4.3. The remaining 128 can in turn be divided between the ones with a 1-2 sequence and those with a 2-1. In this chapter we analyse only the former; results for the latter can be obtained by swapping the indexes uniformly in an axiom and referring to the results we prove in this section.

Among the remaining 64 axioms there are two non-trivial axioms for which a completeness proof is not given here. I will leave these as open problems, conjecture that these logics are not canonical. Still, a frame-correspondence for the axioms is presented.

Given the number of axioms in analysis, and differently from what we have done so far, in this section we will mostly focus on completeness results (given their importance) without analysing the correspondence problem.

4.4.1.1 Interaction axioms of the form $\Box_1 \Box_2 p \Rightarrow \boxdot \boxdot p$

We begin the study of this class of interactions by examining the case of the strongest antecedent in Equation 4.4, i.e. a sequence of two modal boxes. The results of all these interaction axioms are summarised in Figure 4.8. It can be noted that all but one axiom are already theorems of $\mathrm{S5}_2$; this is because of the strength of the antecedent.

Theorem 4.23. *All the logics in Figure 4.8 are sound and complete with respect to the class of equivalence frames with the exception of the logic* $\mathrm{S5}_2 + \{\Box_1 \Box_2 p \Rightarrow \Box_2 \Box_1 p\}$*, which is sound and complete with respect to equivalence 2WD frames.*

Interaction Axioms	*Completeness*	*Lemmas of Ref.*	*Notes*
$\Box_1\Box_2 p \Rightarrow \Diamond_1\Box_1 p$	—	–	$\vdash \Diamond_1\Box_1 p \Leftrightarrow \Box_1 p$
$\Box_1\Box_2 p \Rightarrow \Diamond_1\Box_2 p$	—	–	$\vdash \Box_1 p \Rightarrow \Diamond_1 p$
$\Box_1\Box_2 p \Rightarrow \Diamond_1\Diamond_1 p$	—	–	$\vdash \Diamond_1 p \Leftrightarrow \Diamond_1\Diamond_1 p$
$\Box_1\Box_2 p \Rightarrow \Diamond_1\Diamond_2 p$	—	–	$\vdash p \Rightarrow \Diamond_2 p$
$\Box_1\Box_2 p \Rightarrow \Box_1\Box_1 p$	—	–	$\vdash \Box_1 p \Leftrightarrow \Box_1\Box_1 p$
$\Box_1\Box_2 p \Rightarrow \Box_1\Box_2 p$	—	–	–
$\Box_1\Box_2 p \Rightarrow \Box_1\Diamond_1 p$	—	–	$\vdash \Diamond_1 p \Leftrightarrow \Box_1\Diamond_1 p$
$\Box_1\Box_2 p \Rightarrow \Box_1\Diamond_2 p$	—	–	$\vdash p \Rightarrow \Diamond_2 p$
$\Box_1\Box_2 p \Rightarrow \Diamond_2\Box_1 p$	—	–	$\vdash p \Rightarrow \Diamond_2 p$
$\Box_1\Box_2 p \Rightarrow \Diamond_2\Box_2 p$	—	–	$\vdash \Diamond_2 p \Leftrightarrow \Diamond_2\Box_2 p$
$\Box_1\Box_2 p \Rightarrow \Diamond_2\Diamond_1 p$	—	–	$\vdash p \Rightarrow \Diamond_2 p$
$\Box_1\Box_2 p \Rightarrow \Diamond_2\Diamond_2 p$	—	–	$\vdash \Diamond_2 p \Leftrightarrow \Diamond_2\Diamond_2 p$
$\Box_1\Box_2 p \Rightarrow \Box_2\Box_1 p$	2WD	4.24	–
$\Box_1\Box_2 p \Rightarrow \Box_2\Box_2 p$	—	–	$\vdash \Box_2 p \Leftrightarrow \Box_2\Box_2 p$
$\Box_1\Box_2 p \Rightarrow \Box_2\Diamond_1 p$	—	–	–
$\Box_1\Box_2 p \Rightarrow \Box_2\Diamond_2 p$	—	–	$\vdash \Diamond_2 p \Leftrightarrow \Box_2\Diamond_2 p$

Figure 4.8: An exhaustive list of interaction axioms generated by Equation 4.4 in the case the antecedent is equal to $\Box_1\Box_2 p$.

Proof. It can easily be checked that all the axioms of Figure 4.8 but $\Box_1\Box_2 p \Rightarrow \Box_2\Box_1 p$ are already theorems of $S5_2$. Therefore any extension of $S5_2$ with those axioms is still sound and complete with respect to equivalence frames.

The completeness proof for the logic $S5_2 + \{\Box_1\Box_2 p \Rightarrow \Box_2\Box_1 p\}$ follows from Lemma 4.24, the considerations presented in the proof of Theorem 4.13 and Lemma A.20. $\Box$

4.4.1.1.1 $\Box_1\Box_2 p \Rightarrow \Box_2\Box_1 p$

Lemma 4.24. $\vdash_{S5_2+\{\Box_1\Box_2 p\Rightarrow\Box_2\Box_1 p\}} \Diamond_1\Box_2 p \Rightarrow \Box_2\Diamond_1 p$ *and* $\vdash_{S5_2+\{\Diamond_1\Box_2 p\Rightarrow\Box_2\Diamond_1 p\}} \Box_1\Box_2 p \Rightarrow \Box_2\Box_1 p$.

Proof. First part. Suppose $\Box_1\Box_2 p \Rightarrow \Box_2\Box_1 p$. First we prove that this axiom implies $\Box_2\Box_1 p \Rightarrow \Box_1\Box_2 p$. In fact we have $\Diamond_1\Diamond_2 p \Rightarrow \Diamond_2\Diamond_1\Diamond_2 p$; but by using Lemma 1.29 on the contrapositive of the assumption we have $\Diamond_2\Diamond_1\Diamond_2 p \Rightarrow \Diamond_2\Diamond_2\Diamond_1 p$; so we have $\Diamond_1\Diamond_2 p \Rightarrow \Diamond_2\Diamond_1 p$. So the axiom is symmetric; we now use this intermediate result to prove that the axiom implies $\Diamond_1\Box_2 p \Rightarrow \Box_2\Diamond_1 p$.
By Lemma 1.30 we have $\Diamond_1\Box_2 p \Rightarrow \Diamond_1\Diamond_2\Box_2 p$. But by what we proved above we have $\Diamond_1\Diamond_2\Box_2 p \Rightarrow \Diamond_2\Diamond_1\Box_2 p$. So by Lemma 1.30 and transitivity we obtain $\Diamond_1\Box_2 p \Rightarrow \Box_2\Diamond_2\Diamond_1\Box_2 p$. But by assumption and Lemma 1.29 we have $\Box_2\Diamond_2\Diamond_1\Box_2 p \Rightarrow \Box_2\Diamond_1\Diamond_2\Box_2 p$. So $\Diamond_1\Box_2 p \Rightarrow \Box_2\Diamond_1\Box_2 p$. So by applying axiom T we obtain the result $\Diamond_1\Box_2 p \Rightarrow \Box_2\Diamond_1 p$.

Second part. Suppose $\Diamond_1\Box_2 p \Rightarrow \Box_2\Diamond_1 p$. We have $\Diamond_1\Diamond_2 p \Rightarrow \Diamond_1\Diamond_2\Diamond_1 p$ and since $\Diamond_1 p \Rightarrow \Box_1\Diamond_1 p$ we also have $\Diamond_1\Diamond_2 p \Rightarrow \Diamond_1\Diamond_2\Box_1\Diamond_1 p$. From the previous formula by applying our working hypothesis in the instance $\Diamond_1 p$ for p, by taking the contrapositive and necessitating by $\Box_1$ we obtain $\Diamond_1\Diamond_2 p \Rightarrow \Diamond_1\Box_1\Diamond_2\Diamond_1 p$. This in turn, thanks to the equivalence $\Diamond_1\Box_1 p \Leftrightarrow \Box_1 p$ brings us to $\Diamond_1\Diamond_2 p \Rightarrow \Box_1\Diamond_2\Diamond_1 p$. Now by applying T we finally obtain $\Diamond_1\Diamond_2 p \Rightarrow \Diamond_2\Diamond_1 p$, which by contraposition and substitution leads to the formula $\Box_2\Box_1 p \Rightarrow \Box_1\Box_2 p$. Now by a

Interaction Axioms	*Completeness*	*Lemmas of Ref.*	*Notes*
$\Diamond_1\Box_2 p \Rightarrow \Diamond_1\Box_1 p$	$\sim_1\subseteq\sim_2$	A.6	$\vdash \Diamond_1\Box_1 p \Leftrightarrow \Box_1 p$
$\Diamond_1\Box_2 p \Rightarrow \Diamond_1\Box_2 p$	—	–	–
$\Diamond_1\Box_2 p \Rightarrow \Diamond_1\Diamond_1 p$	—	–	$\vdash \Diamond_1 p \Leftrightarrow \Diamond_1\Diamond_1 p$
$\Diamond_1\Box_2 p \Rightarrow \Diamond_1\Diamond_2 p$	—	–	$\vdash p \Rightarrow \Diamond_2 p$
$\Diamond_1\Box_2 p \Rightarrow \Box_1\Box_1 p$	$\sim_1\subseteq\sim_2$	A.6	$\vdash \Box_1 p \Leftrightarrow \Box_1\Box_1 p$
$\Diamond_1\Box_2 p \Rightarrow \Box_1\Box_2 p$	$\sim_1\subseteq\sim_2$	A.17	–
$\Diamond_1\Box_2 p \Rightarrow \Box_1\Diamond_1 p$	—	–	$\vdash \Diamond_1 p \Leftrightarrow \Box_1\Diamond_1 p$
$\Diamond_1\Box_2 p \Rightarrow \Box_1\Diamond_2 p$	$\sim_1\subseteq\sim_2$	A.18	–
$\Diamond_1\Box_2 p \Rightarrow \Diamond_2\Box_1 p$	$\sim_1\subseteq\sim_2$	4.26	–
$\Diamond_1\Box_2 p \Rightarrow \Diamond_2\Box_2 p$	$\sim_1\subseteq\sim_2$	A.10	$\Box_2 p \Rightarrow \Diamond_2\Box_2 p$
$\Diamond_1\Box_2 p \Rightarrow \Diamond_2\Diamond_1 p$	—	–	$\vdash \Box_2 p \Rightarrow p$
$\Diamond_1\Box_2 p \Rightarrow \Diamond_2\Diamond_2 p$	$\sim_1\subseteq\sim_2$	A.16	$\Diamond_2 p \Rightarrow \Diamond_2\Diamond_2 p$
$\Diamond_1\Box_2 p \Rightarrow \Box_2\Box_1 p$	$\sim_1\subseteq\sim_2$	A.19	–
$\Diamond_1\Box_2 p \Rightarrow \Box_2\Box_2 p$	$\sim_1\subseteq\sim_2$	A.10	$\Box_2 p \Rightarrow \Box_2\Box_2 p$
$\Diamond_1\Box_2 p \Rightarrow \Box_2\Diamond_1 p$	2WD	A.20	—
$\Diamond_1\Box_2 p \Rightarrow \Box_2\Diamond_2 p$	$\sim_1\subseteq\sim_2$	A.16	$\Box_2\Diamond_2 p \Rightarrow \Diamond_2 p$

Figure 4.9: An exhaustive list of interaction axioms generated by Equation 4.4 in the case the antecedent is equal to $\Diamond_1\Box_2 p$.

proof analogous to the one presented in the first part of this lemma we can get $\Box_1\Box_2 p \Rightarrow \Box_2\Box_1 p$, which is what we needed to prove.

$\Box$

4.4.1.2 Interaction axioms of the form $\Diamond_1\Box_2 p \Rightarrow \boxdot\,\boxdot\, p$

We analyse here interaction axioms of the form of Equation 4.4, whose antecedent is made by one diamond indexed by 1 followed by one box indexed by 2. The results of all these interaction axioms are summarised in Figure 4.9. Note that for some axioms the reference lemmas refer to axioms with the indexes swapped. For example for the last axiom $\Diamond_1\Box_2 p \Rightarrow \Box_2\Diamond_2 p$, which by the observation in the note is equivalent to $\Diamond_1\Box_2 p \Rightarrow \Diamond_2 p$. The reference lemma is relative to the axiom $\Box_1 p \Rightarrow \Box_2\Diamond_1 p$, which is the contrapositive of that formula with indexes swapped. It should be clear that by swapping the indexes in that proof for that axiom we obtain a proof for $\Diamond_1\Box_2 p \Rightarrow \Diamond_2 p$.

Once again we will go through only one result and we will leave all the others in the Appendix in Section A.2.1.

Theorem 4.25. *All the logics in Figure 4.9 are sound and complete with respect to the class of equivalence frames satisfying the corresponding property.*

Proof. Equivalent to the proof for Theorem 4.13. All the relevant proofs (except the one for axiom $\Diamond_1\Box_2 p \Rightarrow \Diamond_2\Box_1 p$, which is presented below) are in the Appendix in Section A.2.1. $\Box$

4.4.1.2.1 $\Diamond_1\Box_2 p \Rightarrow \Diamond_2\Box_1 p$

Lemma 4.26. $\vdash_{S5_2+\{\Diamond_1\Box_2 p\Rightarrow\Diamond_2\Box_1 p\}} \Box_2 p \Rightarrow \Box_1 p$ *and* $\vdash_{S5_2+\{\Box_2 p\Rightarrow\Box_1 p\}} \Diamond_1\Box_2 p \Rightarrow \Diamond_2\Box_1 p$.

Interaction Axioms	*Completeness*	*Lemmas of Ref.*	*Notes*
$\Diamond_1\Diamond_2 p \Rightarrow \Diamond_1\Box_1 p$	$\sim_1=\sim_2= id_W$	A.4	$\vdash \Diamond_1\Box_1 p \Leftrightarrow \Box_1 p$
$\Diamond_1\Diamond_2 p \Rightarrow \Diamond_1\Box_2 p$	$\sim_2= id_W$	A.21	–
$\Diamond_1\Diamond_2 p \Rightarrow \Diamond_1\Diamond_1 p$	$\sim_2\subseteq\sim_1$	A.12	$\vdash \Diamond_1 p \Leftrightarrow \Diamond_1\Diamond_1 p$
$\Diamond_1\Diamond_2 p \Rightarrow \Diamond_1\Diamond_2 p$	—	–	–
$\Diamond_1\Diamond_2 p \Rightarrow \Box_1\Box_1 p$	$\sim_1=\sim_2= id_W$	A.4	$\vdash \Box_1 p \Leftrightarrow \Box_1\Box_1 p$
$\Diamond_1\Diamond_2 p \Rightarrow \Box_1\Box_2 p$	$\sim_1=\sim_2= id_W$	A.22	–
$\Diamond_1\Diamond_2 p \Rightarrow \Box_1\Diamond_1 p$	$\sim_2\subseteq\sim_1$	A.12	$\vdash \Diamond_1 p \Leftrightarrow \Box_1\Diamond_1 p$
$\Diamond_1\Diamond_2 p \Rightarrow \Box_1\Diamond_2 p$	$\sim_1\subseteq\sim_2$	A.17	See $\Diamond_1\Box_2 p \Rightarrow \Box_1\Box_2 p$
$\Diamond_1\Diamond_2 p \Rightarrow \Diamond_2\Box_1 p$	$\sim_1= id_W$	A.23	–
$\Diamond_1\Diamond_2 p \Rightarrow \Diamond_2\Box_2 p$	$\sim_1=\sim_2= id_W$	A.8	$\Diamond_2 p \Leftrightarrow \Diamond_2\Box_2 p$
$\Diamond_1\Diamond_2 p \Rightarrow \Diamond_2\Diamond_1 p$	2WD	A.24	See $\Box_1\Box_2 p \Rightarrow \Box_2\Box_1 p$
$\Diamond_1\Diamond_2 p \Rightarrow \Diamond_2\Diamond_2 p$	$\sim_1\subseteq\sim_2$	A.14	$\Diamond_2 p \Leftrightarrow \Diamond_2\Diamond_2 p$
$\Diamond_1\Diamond_2 p \Rightarrow \Box_2\Box_1 p$	$\sim_1=\sim_2= id_W$	A.25	–
$\Diamond_1\Diamond_2 p \Rightarrow \Box_2\Box_2 p$	$\sim_1=\sim_2= id_W$	A.8	$\Box_2 p \Leftrightarrow \Box_2\Box_2 p$
$\Diamond_1\Diamond_2 p \Rightarrow \Box_2\Diamond_1 p$	$\sim_2\subseteq\sim_1$	4.28	–
$\Diamond_1\Diamond_2 p \Rightarrow \Box_2\Diamond_2 p$	$\sim_1\subseteq\sim_2$	A.14	$\Diamond_2 p \Rightarrow \Box_2\Diamond_2 p$

Figure 4.10: An exhaustive list of interaction axioms generated by Equation 4.4 in the case the antecedent is equal to $\Diamond_1\Diamond_2 p$.

Proof. First part. Suppose $\Diamond_1\Box_2 p \Rightarrow \Diamond_2\Box_1 p$. By axiom T we have $\Diamond_1 p \Rightarrow \Diamond_1\Diamond_2 p$, which, by Lemma 1.30 is equivalent to $\Diamond_1 p \Rightarrow \Diamond_1\Box_2\Diamond_2 p$. Now, by applying our hypothesis with $\Diamond_2 p$ in place of p, we have $\Diamond_1\Box_2\Diamond_2 p \Rightarrow \Diamond_2\Box_1\Diamond_2 p$. So we have $\Diamond_1 p \Rightarrow \Diamond_2\Box_1\Diamond_2 p$. But by an instance of T and Lemma 1.30 we have $\Diamond_2\Box_1\Diamond_2 p \Rightarrow \Diamond_2 p$, which with the above gives us $\Diamond_1 p \Rightarrow \Diamond_2 p$.

Second part. Suppose $\Box_2 p \Rightarrow \Box_1 p$. By Lemma 1.29 we have $\Diamond_1\Box_2 p \Rightarrow \Diamond_1\Box_1 p$, which is equivalent to $\Diamond_1\Box_2 p \Rightarrow \Box_1 p$. But then by an instance of axiom T we obtain $\Diamond_1\Box_2 p \Rightarrow \Diamond_2\Box_1 p$. $\Box$

4.4.1.3 Interaction axioms of the form $\Diamond_1\Diamond_2 p \Rightarrow \boxdot\boxdot p$

We now analyse interaction axioms of the form of Equation 4.4, whose antecedent is composed by two diamonds. The results of this section are summarised in Figure 4.10. In the table the reference lemmas for the axioms whose consequent can be reduced to an operator indexed by 2 are the ones with labels swapped. For example, the axiom $\Diamond_1\Diamond_2 p \Rightarrow \Diamond_2\Diamond_2 p$ gives completeness with respect to equivalence frames such that $\sim_1\subseteq\sim_2$; the reference lemma shown in this case is Lemma A.14. This refers to the axiom $\Box_1 p \Rightarrow \Box_2\Box_1 p$, which gives completeness with respect to equivalence frames such that $\sim_2\subseteq\sim_1$. Again, any proof given for the latter can be repeated for the former simply by swapping the indexes.

Once again we only discuss one of the axioms here and refer the reader interested in a particular axiom to Section A.2.2 in the Appendix.

We have completeness for all the axioms in the figure.

Theorem 4.27. *All the logics in Figure 4.10 are sound and complete with respect to the class of equivalence frames satisfying the corresponding property.*

Interaction Axioms	*Completeness*	*Lemmas of Ref.*	*Notes*
$\Box_1\Diamond_2 p \Rightarrow \Diamond_1\Box_1 p$	$\sim_2 = id_W$	A.1 and A.2	$\vdash \Diamond_1\Box_1 p \Leftrightarrow \Box_1 p$
$\Box_1\Diamond_2 p \Rightarrow \Diamond_1\Box_2 p$	? $\forall w \exists w' \in [w]_{\sim_1} : [w']_{\sim_2} = \{w'\}$	4.32	Only conj.
$\Box_1\Diamond_2 p \Rightarrow \Diamond_1\Diamond_1 p$	$\forall w \exists w' \in [w]_{\sim_1} : [w']_{\sim_2} \subseteq [w]_{\sim_1}$	4.18 and 4.19	$\vdash \Diamond_1 p \Leftrightarrow \Diamond_1\Diamond_1 p$
$\Box_1\Diamond_2 p \Rightarrow \Diamond_1\Diamond_2 p$	—	–	$\vdash \Box_1 p \Rightarrow \Diamond_1 p$
$\Box_1\Diamond_2 p \Rightarrow \Box_1\Box_1 p$	$\sim_2 = id_W$	A.1 and A.2	$\vdash \Box_1 p \Leftrightarrow \Box_1\Box_1 p$
$\Box_1\Diamond_2 p \Rightarrow \Box_1\Box_2 p$	$\sim_2 = id_W$	A.21	–
$\Box_1\Diamond_2 p \Rightarrow \Box_1\Diamond_1 p$	$\forall w \exists w' \in [w]_{\sim_1} : [w']_{\sim_2} \subseteq [w]_{\sim_1}$	4.18 and 4.19	$\vdash \Diamond_1 p \Leftrightarrow \Box_1\Diamond_1 p$
$\Box_1\Diamond_2 p \Rightarrow \Box_1\Diamond_2 p$	—	–	–
$\Box_1\Diamond_2 p \Rightarrow \Diamond_2\Box_1 p$	?Either $\sim_1 = id_W$ or $\sim_1 = id_W$ [2]	4.35	Only conj.
$\Box_1\Diamond_2 p \Rightarrow \Diamond_2\Box_2 p$	$\sim_2 = id_W$	4.14 and 4.15	$\Box_2 p \Leftrightarrow \Diamond_2\Box_2 p$
$\Box_1\Diamond_2 p \Rightarrow \Diamond_2\Diamond_1 p$	—	–	–
$\Box_1\Diamond_2 p \Rightarrow \Diamond_2\Diamond_2 p$	—	–	$\vdash \Diamond_2 p \Leftrightarrow \Diamond_2\Diamond_2 p$
$\Box_1\Diamond_2 p \Rightarrow \Box_2\Box_1 p$	$\sim_2 = id_W$	A.23	–
$\Box_1\Diamond_2 p \Rightarrow \Box_2\Box_2 p$	$\sim_2 = id_W$	4.14 and 4.15	$\Box_2 p \Leftrightarrow \Box_2\Box_2 p$
$\Box_1\Diamond_2 p \Rightarrow \Box_2\Diamond_1 p$	$\sim_2 \subseteq \sim_1$	4.26	–
$\Box_1\Diamond_2 p \Rightarrow \Box_2\Diamond_2 p$	—	–	$\Diamond_2 p \Leftrightarrow \Box_2\Diamond_2 p$

Figure 4.11: An exhaustive list of interaction axioms generated by Equation 4.4 in the case the antecedent is $\Box_1\Diamond_2 p$.

Proof. Equivalent to the proof for Theorem 4.13. All the relevant proofs (except the one for axiom $\Diamond_1\Diamond_2 p \Rightarrow \Box_2\Diamond_1 p$, which is presented below) are in the Appendix in Section A.2.2. $\Box$

4.4.1.3.1 $\Diamond_1\Diamond_2 p \Rightarrow \Box_2\Diamond_1 p$

Lemma 4.28. *$\vdash_{S5_2+\{\Diamond_1\Diamond_2 p \Rightarrow \Box_2\Diamond_1 p\}} \Box_1 p \Rightarrow \Box_2 p$ and $\vdash_{S5_2+\{\Box_1 p \Rightarrow \Box_2 p\}} \Diamond_1\Diamond_2 p \Rightarrow \Box_2\Diamond_1 p$.*

Proof. First part. Assume $\Diamond_1\Diamond_2 p \Rightarrow \Box_2\Diamond_1 p$. So, we have $\Diamond_2 p \Rightarrow \Diamond_2\Diamond_1 p$. But by Lemma 1.30 we have $\Diamond_2 p \Rightarrow \Diamond_2\Box_1\Diamond_1 p$. So by using the contrapositive and substituting $\Diamond_1 p$ for p, we obtain $\Diamond_2 p \Rightarrow \Box_1\Box_2\Diamond_1 p$. So by applying axiom T twice we obtain $\Diamond_2 p \Rightarrow \Diamond_1 p$.

Second part. Assume $\Diamond_2 p \Rightarrow \Diamond_1 p$. By substituting $\Box_1 p$ for p in the above we obtain $\Diamond_2\Box_1 p \Rightarrow \Diamond_1\Box_1 p$, so by Lemma 1.30 $\Diamond_2\Box_1 p \Rightarrow \Box_1 p$. But from $\Box_1 p \Rightarrow \Box_2 p$ we also have (by Lemma 1.29 and Lemma 1.30) $\Box_1 p \Rightarrow \Box_1\Box_2 p$. So by transitivity we have $\Diamond_2\Box_1 p \Rightarrow \Box_1\Box_2 p$, which gives the result.

$\Box$

4.4.1.4 Interactions of the form $\Box_1\Diamond_2 p \Rightarrow \boxdot\boxdot\, p$

Finally, we examine the interactions generated by an antecedent composed by the term $\Box_1\Diamond_2 p$ in Equation 4.4. These axiom schemas are intrinsically harder than the ones studied so far. All the axioms in this class except two can be rewritten into one of shapes we have already examined by taking the contrapositive of them. The results for these axioms are therefore to be found in the preceding sections.

[2] On connected sub-frames.

Theorem 4.29. *All the logics of Figure 4.11 except those referring to the axioms $\Box_1\Diamond_2 p \Rightarrow \Diamond_1\Box_2 p$ and $\Box_1\Diamond_2 p \Rightarrow \Diamond_2\Box_1 p$ in Figure 4.11 are sound and complete with respect to the class of equivalence frames satisfying the corresponding property.*

We now analyse in more details the two axioms that we left out from the theorem above. These axioms cannot intrinsically be rewritten in terms of the axioms analysed so far and remind us of the famous McKinsey axiom (for the mono-modal case)

$$\Box\Diamond p \Rightarrow \Diamond\Box p \qquad \textbf{McK}$$

which so much attention has received in the past. It is known [Ben75, Gol75] that the McKinsey axiom in itself (i.e. when analysed on the class of arbitrary Kripke frames) does not correspond to any first-order condition on the Kripke frames. Still, when added to the logic S4, it generates a condition sometimes called *finality*:

$$\forall w \exists w'(wRw' \wedge \forall w''(w'Rw'' \Rightarrow w' = w'')) \qquad Fin$$

On the one hand, it is well-known [Lem77] that S4 + **McK** is sound and complete with respect to reflexive transitive frames satisfying the condition Fin[3]. On the other hand it has been proved in [Gol91] that the logic K + **McK** (for long time one of *the* open problems of modal logic) is not canonical, although still complete[4].

What follows suggests that in a bi-modal context the extensions of $S5_2$ with multi-modal versions of **McK** seem to behave similarly to K + **McK** rather than S4 + **McK**. In fact, at the end of this section we conjecture that some of these systems are not canonical.

In order to analyse these two interaction axioms we will find it convenient to use the concept of dead ends[5].

Definition 4.30. *A point $w \in W$ is called an i-dead-end if for all $w' \in W$ we have $w \sim_i w'$ implies $w = w'$.*

We have the lemma:

Lemma 4.31. *Given a frame $F = (W, \sim_1, \sim_2)$ and a point w on it, w is an i-dead-end if any only if for any valuation π, we have $(F, \pi) \models_w p \Rightarrow \Box_i p$.*

Proof. From left to right. Suppose $M \models_w p$ where w is an i-dead end. Then $[w]_{\sim_i} = \{w\}$. So $M \models_w \Box_i p$.

From right to left. Suppose w is not an i-dead-end and so there is a point $w' \neq w$ such that $w \sim_i w'$. Consider then a valuation $\pi(p) = \{w\}$. It is easy to check that we then have $(F, \pi) \models_w p$ and $(F, \pi) \not\models_w \Box_i p$. $\Box$

[3]Note that the axiom **McK** has little meaning in the context of the logic S5, because in that case the modalities collapse producing the logic Triv.

[4]It is also been proven that K + **McK** is not compact [Wan92].

[5]"Dead ends" are sometimes referred to as points with no links out of them, i.e. points w such that there is no w' with wR_iw' for the given relation R_i. In the context of reflexive frames, as in this section, no point can be a dead-end. Our definition is the "intuitive" notion of dead-end when using reflexive frames.

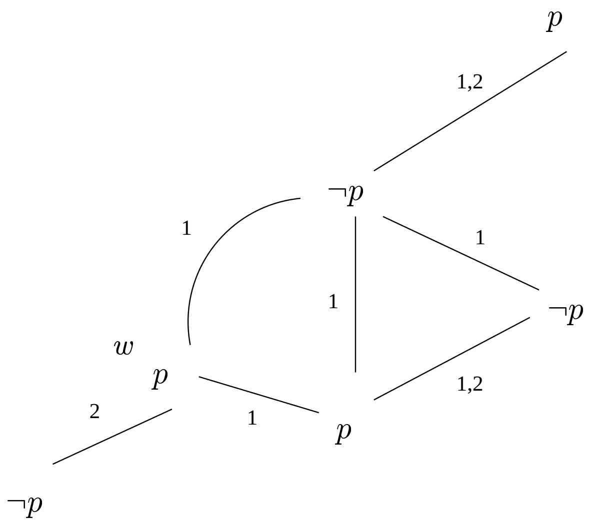

Figure 4.12: A model not satisfying the property of Lemma 4.32.

4.4.1.4.1 $\Box_1\Diamond_2 p \Rightarrow \Diamond_1\Box_2 p$

Lemma 4.32. *$F \models \Box_1\Diamond_2 p \Rightarrow \Diamond_1\Box_2 p$ if and only if F is such that every point w is related by relation 1 to a 2-dead-end; i.e. for all $w \in W$ there exists a $w' \in W$, $w \sim_1 w'$ such that $[w']_{\sim_2} = \{w'\}$.*

Proof. From right to left; consider any model M such that every point sees via 1 a 2-dead-end. Suppose $M \models_w \Box_1\Diamond_2 p$; so for every point w' such that $w \sim_1 w'$ we have that there must be a w'' such that $w' \sim_2 w''$ and $M \models_{w''} p$. But by assumption one of the w' is a 2-dead-end, so we have the existence of a point $\overline{w} \in [w]_{\sim_1}$ such that (by Lemma 4.31) $M \models_{\overline{w}} \Box_2 p$. Then $M \models_w \Diamond_1\Box_2 p$.

For the converse, consider any equivalence frame F, such that $F \models \Box_1\Diamond_2 p \Rightarrow \Diamond_1\Box_2 p$ and suppose by contradiction that the property above does not hold. Consider the set $X = [w]_{\sim_1}$, the equivalence relation $\sim \; = \; \sim_1 \cap \sim_2$ and the quotient set $X/_\sim$. Consider now the set Y constructed by taking one and only one representative w for each class $[w]_\sim$ in $X/_\sim$. Consider a valuation $\pi(p) = Y$ and consider the model $M = (W, \sim_1, \sim_2, \pi)$. By construction we have $M \models_w \Box_1\Diamond_2 p$ (see also Figure 4.12). Then by our assumption we also have $M \models_w \Diamond_1\Box_2 p$. So there must be a point w' such that $w \sim_1 w'$ such that $M \models_{w'} \Box_2 p$. But since w' by assumption is not a 2-dead-end, the equivalence class $[w']_{\sim_2}$ must contain more than w' itself and by construction p is true only at one point in that class and false for every $y \notin X$. So we have $M \not\models_{w'} \Box_2 p$ for every $w' \in [w]_{\sim_1}$ and so $M \not\models_w \Diamond_1\Box_2 p$, which is absurd. So for every point $w \in W$ there must be a 2-dead-end accessible from it. □

Lemma 4.32 gives us hope that the logic $S5_2 + \{\Box_1\Diamond_2 p \Rightarrow \Diamond_1\Box_2 p\}$ is complete with respect to the class of frames above. Unfortunately a completeness proof for that systems does not seem to be so straightforward. In the following we can offer a sufficient condition for canonicity and this problem remains open.

Observation 4.33. *If $\vdash_{S5_2+\{\Box_1\Diamond_2 p \Rightarrow \Diamond_1\Box_2 p\}} \Diamond_1 \wedge_{j=1}^m (\alpha_j \Rightarrow \Box_2\alpha_j)$, then the logic $S5_2 + \{\Box_1\Diamond_2 p \Rightarrow \Diamond_1\Box_2 p\}$ is sound and complete with respect to equivalence frames such that every point is related by relation 1 to a 2-dead-end; i.e. for all $w \in W$ there exists a $w' \in W$, $w \sim_1 w'$ such that $[w']_{\sim_2} = \{w'\}$.*

Proof. Soundness was proven in the first part of Lemma 4.32.

For completeness, we prove that, under the assumption above, the logic $S5_2 + \{\Box_1\Diamond_2 p \Rightarrow \Diamond_1\Box_2 p\}$ is canonical, i.e. that the underlying frame of the canonical model $M = (W, \sim_1, \sim_2, \pi)$ has the property that for all $w \in W$ there exists a $w' \in W$, $w \sim_1 w'$ such that $[w']_{\sim_2} = \{w'\}$. By Definition 1.21 and Lemma 4.31 we have that a characterising condition for 2-dead-ends on the canonical model is that the formula $(\alpha \Rightarrow \Box_2\alpha)$ is in the world for any $\alpha \in \mathcal{L}$. So given any point $w \in W$ to prove canonicity, all we need to prove is that the set

$$\{\alpha \Rightarrow \Box_2\alpha\} \cup \{\beta \mid \Box_1\beta \in w\},$$

is consistent. For that to be the case by Lemma 1.10 we have a point in the canonical model that satisfies the property above.

We prove it by contradiction. Suppose the set above is not consistent; then it means that we can find some $\beta_1, \ldots, \beta_n$ with $\Box_1\beta_1 \in w, \ldots, \Box_1\beta_n \in w$ and some $\alpha_1, \ldots, \alpha_m$ such that

$$\vdash \neg(\beta_1 \wedge \cdots \wedge \beta_n \wedge (\alpha_1 \Rightarrow \Box_2\alpha_1) \wedge \cdots \wedge (\alpha_m \Rightarrow \Box_2\alpha_m))$$

Call now $\beta = \wedge_{i=1}^n \beta_i$; by operating some re-writing we have:

$$\vdash \beta \Rightarrow \neg \wedge_{j=1}^m (\alpha_j \Rightarrow \Box_2\alpha_j)$$

which by applying Lemma 1.29 gives:

$$\vdash \Box_1\beta \Rightarrow \Box_1\neg \wedge_{j=1}^m (\alpha_j \Rightarrow \Box_2\alpha_j)$$

Now, it can be observed that, since $\Box_1\beta_1 \in w, \ldots, \Box_1\beta_n \in w$, we have $(\Box_1\beta_1 \wedge \ldots \Box_1\beta_n) \in w$. So $\Box_1\beta \in w$. Therefore it must be that $\Box_1\neg\wedge_{j=1}^m(\alpha_j \Rightarrow \Box_2\alpha_j) \in w$, i.e. $\neg\Diamond_1\wedge_{j=1}^m(\alpha_j \Rightarrow \Box_2\alpha_j) \in w$. But this is impossible because by assumption we actually have $\vdash \Diamond_1 \wedge_{j=1}^m (\alpha_j \Rightarrow \Box_2\alpha_j)$, and so w would be inconsistent if it contained that formula. $\Box$

Note that we do have

$$\vdash_{S5_2+\{\Box_1\Diamond_2 p \Rightarrow \Diamond_1\Box_2 p\}} \Diamond_1(\alpha \Rightarrow \Box_2\alpha)$$

as that follows from

$$\vdash_{S5_2+\{\Box_1\Diamond_2 p \Rightarrow \Diamond_1\Box_2 p\}} \Box_1\alpha \Rightarrow \Diamond_1\Box_2\alpha$$

which is clearly true. Unfortunately after some work on this issue it seems to me that the hypothesis of Observation 4.33 does not hold. Actually I conjecture a result stronger than the above.

Conjecture 4.34. *The logic* $S5_2 + \{\Box_1\Diamond_2 p \Rightarrow \Diamond_2\Box_1 p\}$ *is not canonical.*

4.4.1.4.2 $\Box_1\Diamond_2 p \Rightarrow \Diamond_2\Box_1 p$

The situation is similar to the one above. We do have a correspondence result but we lack a completeness proof.

Lemma 4.35. $F \models \Box_1\Diamond_2 p \Rightarrow \Diamond_2\Box_1 p$ *if and only* F *is such that if in every connected sub-frame either* $\sim_1 = id_W$ *or* $\sim_2 = id_W$.

Proof. From left to right. This part of the proof is structured as follows:

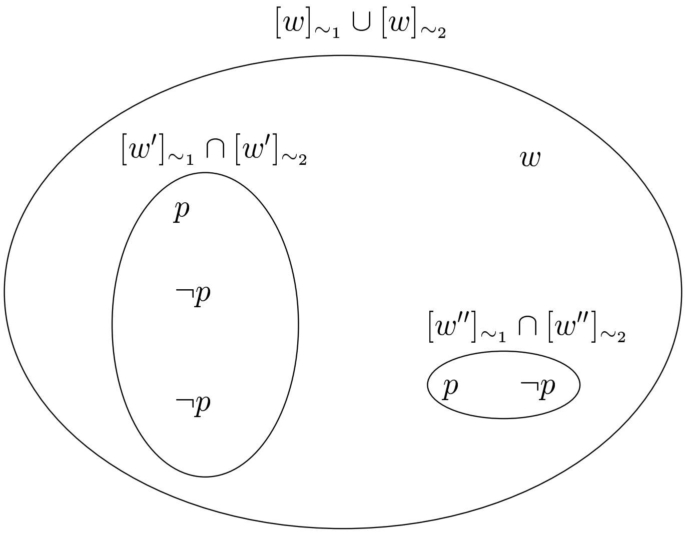

Figure 4.13: A model not satisfying the property of Lemma 4.35. Note that $[w]_{\sim_1} \neq \{w\}$ and $[w]_{\sim_2} \neq \{w\}$.

1. We prove that $F \models \Box_1\Diamond_2 p \Rightarrow \Diamond_2\Box_1 p$ implies that any point $w \in W$ either sees via 1 a 2-dead-end, or the point w sees via 2 a 1-dead-end.

2. We prove that if on a frame F such that $F \models \Box_1\Diamond_2 p \Rightarrow \Diamond_2\Box_1 p$ and there is point w which is an i-dead-end, then $\sim_i = id_W$ on the whole connected sub-frame generated by w; where $i \in \{1, 2\}$.

3. The two facts above together prove that if $F \models \Box_1\Diamond_2 p \Rightarrow \Diamond_2\Box_1 p$, then in every connected sub-frame either $\sim_1 = id_W$ or $\sim_2 = id_W$.

1) By contradiction, consider any connected equivalence frame F, in which a $w \in W$ does not see via i any j-dead end, i.e. $\forall w' \in [w]_{\sim_i}, [w']_{\sim_j} \neq \{w'\}, i \neq j, i, j \in \{1, 2\}$; we prove that $F \not\models \Box_1\Diamond_2 p \Rightarrow \Diamond_2\Box_1 p$. To see this, consider the set $X = [w]_{\sim_1} \cup [w]_{\sim_2} \setminus \{w\}$, the equivalence relation $\sim\, = \,\sim_1 \cap \sim_2$ and the quotient set $X/_\sim$. Consider now the set Y defined by taking one representative y for every equivalence class $[y]_\sim \in X/_\sim$: the set Y is such that $\forall y_1, y_2 \in Y$ we have $[y_1]_\sim \cap [y_2]_\sim = \emptyset$ and $\bigcup_{y \in Y} [y]_\sim = X$. Consider now the model $M = (F, \pi)$, by taking the valuation $\pi(p) = Y$. By construction, in the model M for any $x \in X$, there is a point accessible from x via $\sim_2$ which satisfies p, and since by hypothesis w is neither a 1-dead-end nor a 2-dead-end (as otherwise it would see itself as dead-end) we have $M \models_w \Box_1\Diamond_2 p$. So by the validity of the axiom we also have $M \models_w \Diamond_2\Box_1 p$, i.e. there must be a $w' \in [w]_{\sim_2}$, such that $M \models_{w'} \Box_1 p$, but this is impossible because by hypothesis $[w']_{\sim_1} \neq \{w'\}$, and by construction p is true at just one point in $[w']_{\sim_1} \cap [w']_{\sim_2}$, and false at every point not in X. See Figure 4.13.

2) Consider now a connected frame F such that $F \models \Box_1\Diamond_2 p \Rightarrow \Diamond_2\Box_1 p$ and suppose for example that w is a 1-dead-end, we want to prove that $\sim_1 = id_W$ on the connected sub-frame generated by w[6]. If w is also a 2-dead-end, then $\sim_1 = \sim_2 = id_W$ on the generated frame which gives us the result. If not, suppose that $\sim_1 \neq id_W$; so there must be two points $w', w'' \in$

[6]If w is a 2-dead-end then the argument is symmetric.

$W; w' \neq w''$, such that $w' \sim_1 w''$. So, since the frame is connected, without loss of generality assume $w \sim_2 w'$. Consider now valuation $\pi(p) = \{x \mid x \in [w]_{\sim_2}, x \neq w'\} \cup \{w''\}$ and the model $M = (F, \pi)$ built on F from π. So, we have $M \models_w \Box_2\Diamond_1 p$, and so, by validity of the axiom, we also have $M \models_w \Diamond_1\Box_2 p$. So we must have $M \models_w \Box_2 p$, which is a contradiction because $M \models_{w'} \neg p$.

So we have that if the axiom is valid, then in every connected component one of the two relations is the identity.

From right to left. Consider any equivalence model M whose underlying frame satisfies the property above and suppose that $M \models_w \Box_1\Diamond_2 p$.

Suppose $\sim_1 = id_W$ and $M \models_w \Box_1\Diamond_2 p$, so there is a $w' \in [w]_{\sim_2}$, such that $M \models_{w'} p$. But since $\sim_1 = id_W$ on the connected part, we also have $M \models_{w'} \Box_1 p$. So $M \models_{w'} \Diamond_2\Box_1 p$. Suppose now $\sim_2 = id_W$ and $M \models_w \Box_1\Diamond_2 p$. So for every $w' \in [w]_{\sim_1}$ we have $M \models_{w'} p$. But then we also have $M \models_w \Diamond_2\Box_1 p$. ☐

Considerations similar to the ones presented for the axiom of Section 4.4.1.4.1 lead to the following.

Conjecture 4.36. *The logic* $\mathrm{S5}_2 + \{\Box_1\Diamond_2 p \Rightarrow \Diamond_2\Box_1 p\}$ *is not canonical.*

4.4.2 Discussion

In Section 4.4.1 we proved that with the exception of two logics that we discussed above, all the interaction axioms of the shape of Equation 4.4 when added to the logic $\mathrm{S5}_2$ produce complete extensions of $\mathrm{S5}_2$. Once again it is quite interesting to note that, considering the relatively large number of axioms that we have examined, we did not find many new independent logics. Indeed the only independent extension that we had not met in the previous sections of this chapter is the logic $\mathrm{S5}_2 + \mathbf{2WD}$ that we had discussed in its more general form (i.e. $\mathrm{S5WD}_n$) in Chapter 3.

Corollary 4.37. *All the logics $S5_2 + \{\phi\}$, where ϕ is the conjunction of formulae expressible as Equation 4.4 but not the axioms of Section 4.4.1.4.1 and Section 4.4.1.4.2, are complete with respect to the intersection of the corresponding classes of frames given in Figure 4.8, Figure 4.9, Figure 4.10, Figure 4.11.*

The logics are ordered as in Figure 4.14.

Proof. It follows from the canonicity proofs of Theorem 4.23, Theorem 4.25, Theorem 4.27, Theorem 4.29. Proving the relation between the logics is straightforward. ☐

If we consider all the logics examined so far in this chapter except the two McKinsey style logics we have the general picture of Figure 4.14. In the figure, the logics are ordered strength-wise. So, the strongest logic is of course Triv_2 (represented as $\mathrm{S5}_2 + \{\Diamond_1 p \Rightarrow \Box_2 p\}$), the weakest simply $\mathrm{S5}_2$. In between we have a few logic systems, the weakest of which are the logic $\mathrm{S5WD}_n$ and the ones generated by the two axioms that we examined in Section 4.3. Note that these three logics are independent. Stronger extensions include logics in which the knowledge of an agent is included in the knowledge of the other and combination of these. In the figure we can note that logics which are positioned on either one of the two wings of the lattice represent scenarios in which the agents have different capabilities, typically one having introspection over the other's knowledge. Note that the logic

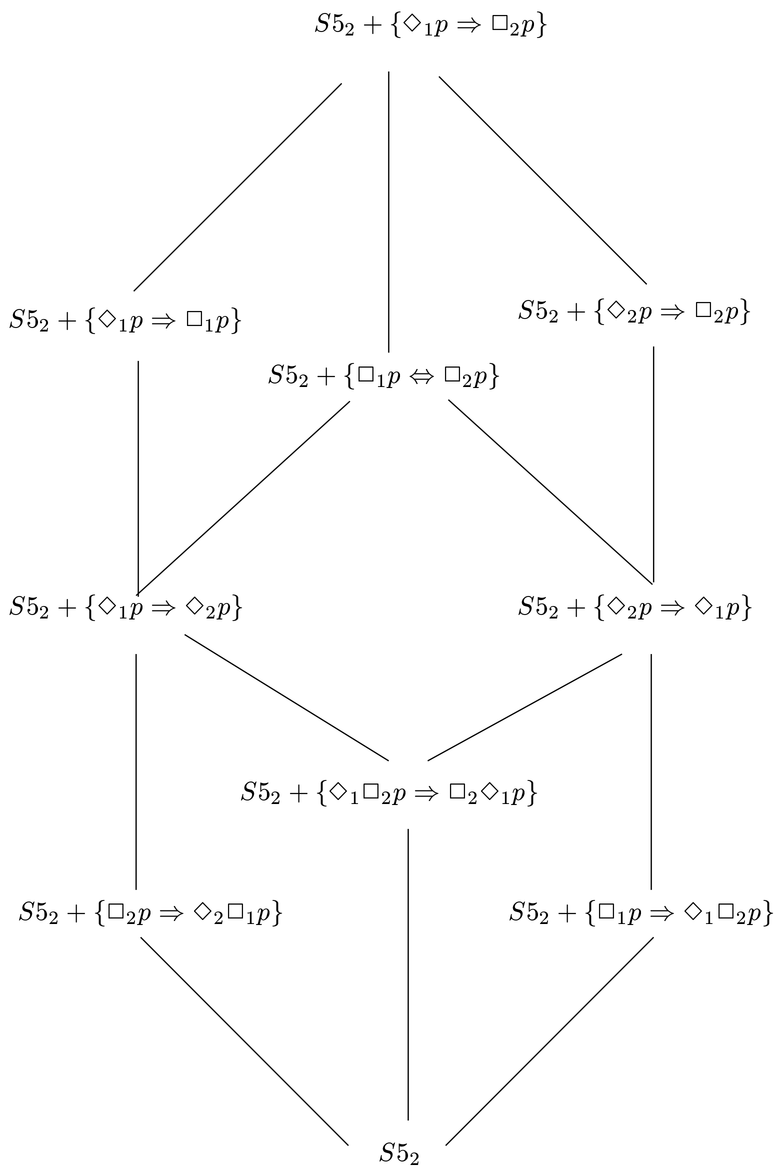

Figure 4.14: The independent extensions of $S5_2$ that can be obtained by adding the axioms studied in this chapter. The logics for which results are only conjectured are not included in the figure.

$S5_2 + \{\Diamond_1\Box_2 p \Rightarrow \Box_2\Diamond_1 p\}$ is placed in the centre of the diagram reflecting the homogeneity of hypercube agents.

The equivalences we have proved in this section are also quite interesting. For example we have seen that in the case of 2 agents axiom **2WD**, sometimes known as Catach's axiom, is equivalent to $\Box_1\Box_2 p \Rightarrow \Box_2\Box_1 p$, whose meaning in terms of knowledge is more intuitive. This axiom basically says that the knowledge of facts that are known to be known by one agent is effectively shared. In fact, note that in Lemma 4.24 we also proved that the axiom $\Box_2\Box_1 p \Rightarrow \Box_1\Box_2 p$ also follows from $\mathbf{2WD}$.

4.5 Conclusions

In Section 1.4.1 we motivated the need for formal models of agents that share part of their knowledge. Hypercubes form one of these classes; in fact, as we have seen in Chapter 3, agents modelled by hypercubes share part of their knowledge following the logic $\mathrm{S5WD}_n$.

In Chapter 3, although we were able to see the axiom $\mathbf{WD}$ in a more intuitive form, we were still left with the question of how the logic $\mathrm{S5WD}_n$ was related to the logic $\mathrm{S5}_n$ and to its neighbours. We conducted this analysis in this chapter for the case $n = 2$ were we examined all the interactions axioms that can be written as an implication expressing the fact that knowledge and facts considered possible are related to each other up to a level of nesting of two.

A *spectrum* of degrees of knowledge sharing has emerged. For the case $n = 2$, the logic $\mathrm{S5WD}_n$ was shown to be the weakest symmetric extension of $\mathrm{S5}_2$, but other logics, meaningful in epistemic settings have emerged.

Although not analysed in this chapter the finite model property for all the canonical logics should follow and with it decidability. Unfortunately, results for the two McKinsey-style axioms remain conjectured at this stage. These seem to be quite non-trivial to obtain but if proven they would provide very interesting examples of non-canonical extensions of $\mathrm{S5}_2$, something that to my knowledge has not been shown before.

Chapter 5

Refining and checking knowledge structures

5.1 Introduction

In Chapter 2 and Chapter 3 we have analysed hypercube systems at various conceptual levels. We have seen that hypercubes are modelled by the logic $\mathrm{S5WD}_n$ and represent agents with no a-priori information about each other's local states and whose evolution is linked to information broadcasting and no private message-passing.

In particular in Section 3.8 we discussed how to adapt the formal model of contexts to account for the evolution of hypercubes. We used internal and external actions explicitly in order to model the change that broadcasting actions impose on the local states of the agents. In that model Kripke structures (although "generated" by the formalism) are not present explicitly in the definition of change in the local states. In this chapter we discuss whether Kripke semantics can be proven useful to model the update of knowledge states.

More specifically, this chapter can be seen as addressing the following question. We already know that Kripke models are good tools for representing static private and common knowledge. Suppose we have a model M representing a snapshot of the knowledge of a MAS and further suppose that a broadcasting act is performed. The broadcasted message will affect the status of knowledge of the agents. Is it possible to refine the model M so that the new model takes into account the changes in the knowledge of the agents following the receipt of the message? This chapter proposes a method for performing this operation. Actually this chapter tries to go beyond that as it is also closely related to the more general question of which approach (proof-theoretical or semantical) is best suited to reason about knowledge in a MAS.

From the discussion of page 22 it follows that the standard approach to using the formal machinery provided by the logic $\mathrm{S5}_n$ is to describe a situation as a set of formulae Γ and to attempt to show that the situation satisfies a property ϕ by establishing $\Gamma \vdash_{\mathrm{S5}_n} \phi$ or $\Gamma \models_{\mathcal{F}_E} \phi$. Establishing $\Gamma \vdash \phi$ involves finding a proof in $\mathrm{S5}_n$ of ϕ from Γ, while establishing $\Gamma \models_{\mathcal{F}_E} \phi$ involves reasoning about all (usually infinitely many) equivalence Kripke models satisfying Γ to show that they also satisfy ϕ. The completeness of $\mathrm{S5}_n$ shows that these two notions are equivalent. However, the complexity of the validity problem for the logic $S5_n$ was proved to be PSPACE-complete and that goes up to EXPTIME-completeness for the case of $S5_n^{E,C}$ (see [HM92a] for details).

In order to overcome the intractability of this approach, Halpern and Vardi have proposed to use *model checking* as an alternative to theorem proving [HV91]. In the model checking approach, the situation to be modelled is codified as a single Kripke model M rather than as a set of formulae Γ. The task of verifying that a property ϕ holds boils down to checking that M satisfies ϕ, written $M \vDash \phi$. This task is computationally much easier than the theorem proving task, being linear in the size of M and the size of ϕ (see [HV91] for details).

Halpern and Vardi informally illustrate their approach by modelling the muddy children puzzle. In that puzzle, there are n children and n atomic propositions $p_1, p_2, \ldots, p_n$ representing whether each of the children has mud on their faces or not. Various announcements are made, first by the father of the children and then by the children themselves. The children thus acquire information about what other children know, and after some time the muddy ones among them are able to conclude that they are indeed muddy. We will discuss the problem in greater detail in Section 5.2.1.

Halpern and Vardi propose the following way of arriving at the model M to be checked. They start with the most general model for the set of atomic propositions at hand. In order to deal with the announcements made, they successively *refine* the model with formulae expressing the announcements made. This refinement process consists of removing some links from the Kripke model. At any time during this process, they can check whether child i knows p_i (for example), by checking whether the current model satisfies $\Box_i p_i$. This method is illustrated in the paper [HV91] and the book [FHMV95], but a precise definition of the refinement operation is not given.

This chapter is organised as follows. In Section 5.2, we visit the idea of model refinement and show the intrinsic problems with the idea of refining Kripke models. In Section 5.3 we introduce a structure derived from a Kripke model, which we call a *Kripke tree*, and define the refinement operation on Kripke trees. We illustrate this notion using the muddy children example in Section 5.4. In Section 5.5 we prove the properties of refinement of Kripke trees and we then conclude in Section 5.6 with some conclusions and general remarks. First in Section 5.1.1 we fix the syntax and semantics used in this chapter.

5.1.1 Syntax and semantics

In this chapter we will make use of slightly different syntax and semantics from the ones introduced in Section 1.2.1.

The first change from Section 1.2.1 is that we use a finite set of variables for the propositional atoms P. Given the more applied nature of this chapter this is not a serious technical limitation[1] but it will help us with the machinery. The set of agents is unchanged from Section 1.2.1 and it is any finite set $A = \{1, \ldots n\}$.

Here we will use the extended syntax introduced in Section 1.3.1.1. In particular we will have operators for "Everybody knows" and common knowledge. Formally, it suffices to define formulae as:

$$\phi \ ::= \ p \mid \neg\phi \mid \phi_1 \wedge \phi_2 \mid \Box_i \phi \mid C\phi$$

where $p \in P$ and $i \in A$. As usual the formula $\Box_i \phi$ represents the situation in which the agent i knows the fact represented by the formula ϕ, while $C\phi$ means that ϕ is common

[1] Any real world example can be coded with a finite set of atoms.

knowledge in the group A[2]. The other propositional connectives can be defined in the usual way. The modal connectives $\Diamond_i$, E and B are defined as:

$$\begin{array}{lll} \Diamond_i\phi & \text{means} & \neg\Box_i\neg\phi \\ E\phi & \text{means} & \bigwedge_{i\in A}\Box_i\phi \\ B\phi & \text{means} & \neg C\neg\phi \end{array}$$

The formula $\Diamond_i\phi$ means "it is consistent with agent i's knowledge that ϕ", $E\phi$ means that everyone knows ϕ. The operator B is the dual of C. Although not particularly useful intuitively, we will need it for technical reasons.

It is worth noting that an announcement of ϕ results in common knowledge of ϕ among the hearers, because as well as hearing ϕ they also see that the others have heard it too (we assume throughout that all the agents are perceptive, intelligent, truthful). If one agent secretly informs all the others of ϕ, the result will be that everyone knows ϕ, but ϕ will not be common knowledge.

We will also need the following definitions.

Definition 5.1. *A formula is* universal *if it has only the modalities $C, E, \Box_i$ and no negations outside them. Formally take*

$$\phi \ ::= \ p \mid \neg\phi \mid \phi_1 \wedge \phi_2 \mid \phi_1 \vee \phi_2$$

and define a formula ψ to be universal if it follows the following syntax:

$$\psi \ ::= \ \phi \mid \psi_1 \vee \psi_2 \mid \psi_1 \wedge \psi_2 \mid \Box_i\psi \mid E\psi \mid C\psi.$$

Definition 5.2. *A formula is* safe *if it is universal and no $\Box_i$ and no C appears in the scope of $\vee$. Formally take*

$$\phi \ ::= \ p \mid \neg\phi \mid \phi_1 \wedge \phi_2 \mid \phi_1 \vee \phi_2$$

and define a formula ψ to be safe if it follows the following syntax:

$$\psi \ ::= \ \phi \mid \psi_1 \wedge \psi_2 \mid \Box_i\psi \mid E\psi \mid C\psi.$$

Definition 5.3. *A formula is* disjunction-free *if it is universal and has no $\vee$. Formally take*

$$\phi \ ::= \ p \mid \neg\phi \mid \phi_1 \wedge \phi_2$$

and define a formula ψ to be disjunction-free if it follows the following syntax:

$$\psi \ ::= \ \phi \mid \psi_1 \wedge \psi_2 \mid \Box_i\psi \mid E\psi \mid C\psi.$$

In this chapter we proceed in a slightly different setting. As we know a Kripke model describes a set of possible situations (worlds). In this chapter we sometimes need to be able to identify which is the "actual" world among all the possible ones that the model represents. Formally this is the definition of equivalence model local to this chapter.

Definition 5.4. *Given a set A of agents, an* equivalence Kripke model $M = (W, \sim, \pi, w)$ *is given by:*

1. *A set W, whose elements are called* worlds*;*

[2]Note that, differently from Section 1.3.1.1, we assume $G = A$ here.

2. *An A-indexed family of relations $\sim = \{\sim_i\}_{i \in A}$. For each $1 \leq i \leq n$, $\sim_i$ is an equivalence relation on W ($\sim_i \subseteq W \times W$), called the* accessibility relation*;*

3. *A function $\pi : P \to 2^W$, called the* interpretation*;*

4. *A world $w \in W$, the* actual world.

We will use satisfaction for the basic modal language as defined in Definition 1.2, and satisfaction for the operators of common knowledge and "everybody knows" as in Section 1.3.2.2.

The actual world of the model represents the situation captured by the model and it is where the formulae will be checked for satisfaction. In this chapter we do not use the notion of validity as we are only interested in truth of formulae, especially at the actual world.

Definition 5.5. *Given a model $M = (W, \sim, \pi, w)$ we say that ϕ is true at M, $M \models \phi$, if $M \models_w \phi$.*

Recall Definition 1.15 about points being reachable from each other. The following fact (reported in [FHMV95]) is useful for understanding the technical difference between E and C.

Theorem 5.6.

1. *$M \models_x E^k\phi$ if and only if for all y that are reachable from x in k steps, we have $M \models_y \phi$.*

2. *$M \models_x C\phi$ if and only if for all y that are reachable from x, we have $M \models_y \phi$.*

5.2 Refining Kripke models

Halpern and Vardi propose to refine Kripke models in order to model the evolution of knowledge. They illustrate their method with the muddy children puzzle. This example is a variant of another puzzle presented in [GS58] and is particularly important in the literature. The version we discuss is the one given in [Bar81] and discussed in [FHMV95]; we report it in the following.

5.2.1 The muddy children puzzle

There is a large group of children playing in the garden. A certain number (say k) get mud on their foreheads. Each child can see the mud (if present) on others but not on his own forehead. If $k > 1$ then each child can see another with mud on its forehead, so each one knows that at least one in the group is muddy. The father first announces that at least one of them is muddy (which, if $k > 1$, is something they know already); and then he repeatedly asks them 'Does any of you know whether you have mud on your own forehead?' The first time they all answer 'no'. Indeed, they go on answering 'no' to the first $k - 1$ questions; but at the kth those with muddy foreheads are able to answer 'yes'.

At first sight, it seems rather puzzling that the children are eventually able to answer the father's question positively. The clue to understanding what goes on lies in the notion of common knowledge. Although everyone knows the content of the father's initial announcement, the father's saying it makes it common knowledge among them, so now they all know that everyone else knows it, etc. Consider a few cases of k.

$k = 1$, i.e. just one child has mud. That child is immediately able to answer 'yes', since she has heard the father and doesn't see any other child with mud.

$k = 2$, say a and b have mud. Everyone answers 'no' the first time. Now a thinks: since b answered 'no' the first time, he must see someone with mud. Well, the only person I can see with mud is b, so if b can see someone else it must be me. So a answers 'yes' the second time. b reasons symmetrically about a, and also answers 'yes'.

$k = 3$, say a, b, c. Everyone answers 'no' the first two times. But now a thinks: if it was just b and c with mud, they would have answered 'yes' the second time. So there must be a third person with mud; since I can only see b, c having mud, the third person must be me. So a answers 'yes' the third time. For symmetrical reasons, so do b, c.

And similarly for other cases of k.

To see that it was not common knowledge before the father's announcement that one of the children was muddy, consider again $k = 2$, say a, b. Of course a and b both know someone is muddy (they see each other), but, for example, a doesn't know that b knows that someone is dirty. For all a knows, b might be the only dirty one, and therefore not be able to see a dirty child.

5.2.2 An engineering example

The muddy children puzzle, together with its many variants like the three wise men puzzle, etc. is popular among computer scientists. The reason is that it encodes subtle properties about reasoning, while also being applicable to real life scenarios. We can imagine an example in which an engineering system could benefit from being able to cope with muddy-children-like situations.

Consider a factory in which similar robots collectively manufacture an object while moving in group in a large space. The robots can roughly be thought of being made of two components: the reasoning module and the mechanical actuators, effectively operating on the object. We want to design a fault detection system for the actuators. Given the large area the robots can be in, the installation of cameras to monitor the operational status of the robots' arms is not an option.

Let us suppose that the robots have a visual system directed towards the other robots that can detect faults in their mechanical arms. Note this is quite a reasonable assumption, since it is often problematic to have visual systems that can do self-monitoring as well as monitoring the environment. Suppose now that the factory has a quality control mechanism that can detect if something went wrong during the production of the object and assume this device broadcasts an alarm every time it notices a defect in the production.

This robotic scenario complies with the muddy children example: the children are now robots, the role of the father is taken by the fault detection system. Note that the assumption of communication being common knowledge is not violated because messages are assumed to be broadcasted to all the agents. The task of the robots is then to reason about their status and stop their operation in case they come to know that their mechanical arm is faulty. The evolution of their knowledge proceeds exactly as the case of the muddy children example where we assume the robots to operate synchronously.

Assuming the robots have a reasoning module able to handle the muddy children problem, the group of robots is then effectively able to do collective diagnosis.

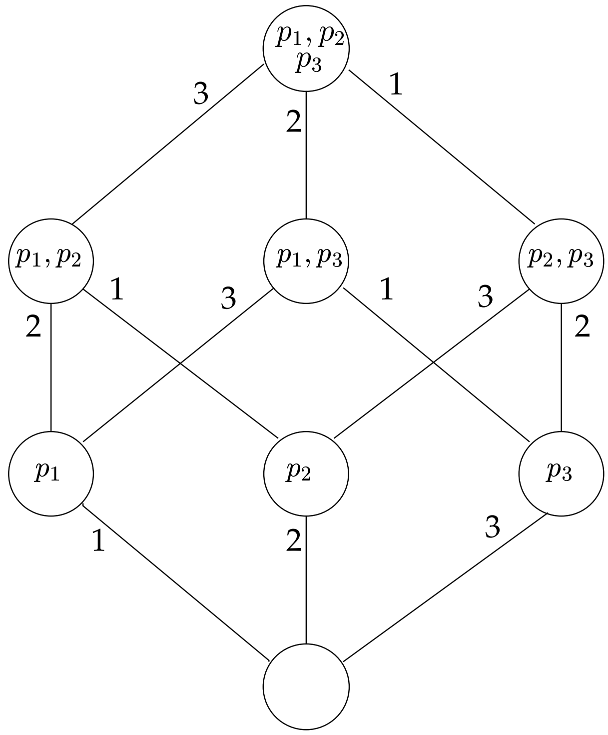

Figure 5.1: M_1: The Kripke model for the muddy children puzzle with $n = 3$.

In the following we refer our discussion to muddy children, but the above scenario can serve equally well.

5.2.3 Halpern and Vardi's formulation

Suppose $A = \{1, \ldots n\}$ and $P = \{p_1, \ldots, p_n\}$; p_i means that the ith child has mud on its forehead. Suppose $n = 3$. The assumption of this puzzle is that each child can see the other children but cannot see itself, so each child knows whether the others have mud or not, but does not know about itself. Under these assumptions, Halpern and Vardi propose the Kripke structure of Figure 5.1 to model the initial situation.

Let w be any world in which there are at least two muddy children (i.e. w is one of the four upper worlds). In w, every child knows that at least one of the children has mud. However, it is not the case that it is common knowledge that each child has mud, since the world at the bottom of the lattice is reachable (cf. Theorem 5.6).

To model the father's announcement, Halpern and Vardi refine the model M_1 in Figure 5.1, arriving at M_2 in Figure 5.2 (these figures also appear in [HV91], [FHMV95]). The refinement process is not precisely defined in [HV91], [FHMV95], though arguments in favour of the transformation from M_1 to M_2 are given.

Suppose now that the father asks the children whether they know whether they are muddy or not, and the children answer simultaneously that that they do not. Halpern and Vardi argue that this renders all models in which there is only one muddy child inaccessible, resulting in M_3 (Figure 5.3).

If there are precisely two children with mud (i.e. the actual world is one of the three in

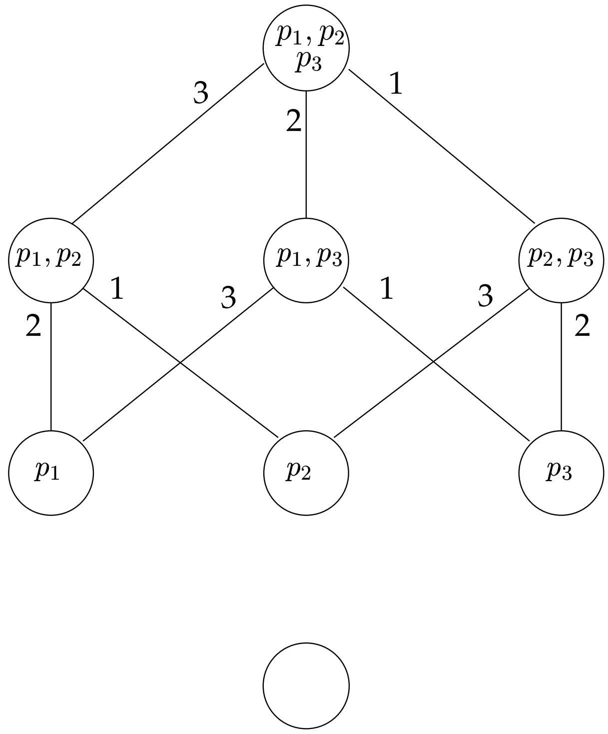

Figure 5.2: M_2: The Kripke structure after the father speaks.

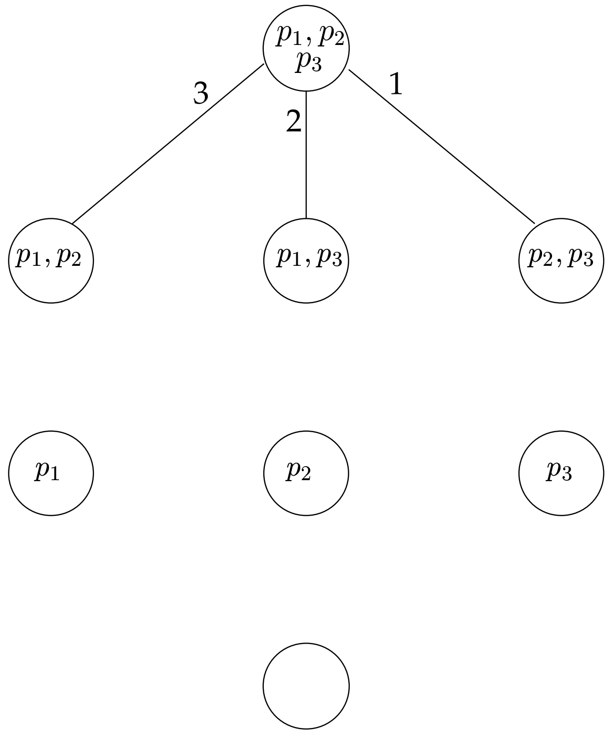

Figure 5.3: M_3: The Kripke structure after the children announce that they don't know whether they are muddy.

the second layer), then each of the muddy children now knows it is muddy. For suppose the actual world is the left one of those three, i.e. w with $\pi(w) = \{p_1, p_2\}$. We easily verify that $M_3 \models_w \Box_1 p_1$ and $M_3 \models_w \Box_2 p_2$.

If all three children are muddy, i.e. the actual world w is the top one, then we are not yet done, for we do not have $M_3 \models_w \Box_i p_i$ for any i. The father again asks each of the children if they know if they are muddy, and the model is refined again according to their answer "no", resulting in M_4 which is M_3 with the last remaining links removed. (M_4 is not illustrated.) We can easily check that $M_4 \models_w \Box_i p_i$ for each i.

In summary, the method proposed by Halpern and Vardi for solving muddy-children-type puzzles is the following. Start with a suitably general model M_1 reflecting the initial set-up of the puzzle. Refine it successively by the announcements made. At the end of the announcements, check formulae against the refined model. In the example above, we refined M_1 first by $\phi_1 = C(p_1 \vee p_2 \vee p_3)$ (the father's announcement), and then twice by

$$\phi_2 = C(\neg\Box_1 p_1 \wedge \neg\Box_1\neg p_1) \wedge C(\neg\Box_2 p_2 \wedge \neg\Box_2\neg p_2) \wedge C(\neg\Box_3 p_3 \wedge \neg\Box_3\neg p_3)$$

which corresponds to each of the three children announcing that they don't know whether they are muddy or not.

Halpern and Vardi do not precisely define what refinement by a formula means. The intuition they give is that refinement removes a minimal set of links of the model, so that the model satisfies the formula at the actual world. Removing links means that epistemic possibilities are removed, that is, knowledge is gained, so this seems intuitively the right thing to do.

5.2.4 Problems with refinement of Kripke models

Let us write $M * \phi$ to denote the result of refining the model M by the formula ϕ. Thus, in the example above, $M_2 = M_1 * \phi_1$, etc.

The muddy-children example discussed above naturally lead us to the question of whether it is possible to make precise the notion of refinement of a Kripke model by a formula, and of what properties this would have. Essentially any refinement procedure will remove the links to the states that are responsible for the non-satisfaction of the formula we are refining with. However, some unexpected problems of any natural procedure operating on Kripke models can be found.

Consider the following examples.

Example 5.2.1. *Let M_5 be the Kripke model illustrated in Figure 5.4, with the left-hand world w the actual world, and consider refining by $\Box_1 p$. However we implement the revision of the model, it would look as though the resulting model should be the model M_6 (see figure). What happens is that agent 1 gains the knowledge of p, and so must eliminate the epistemic possibility of $\neg p$ by removing the link.*

The counterintuitive property of this example is that $M_5 \models_w \Box_3\Diamond_1 p$, while $M_6 \not\models_w \Box_3\Diamond_1 p$. Thus, in M_5, agent 3 knows that p is consistent with 1's knowledge. But after agent 1 learns p for sure in M_6, agent 3 no longer knows this!

Example 5.2.2. *Figure 5.5 shows a model and (the only) two outcomes one could consider for its refinement by $\Box_1\Box_2(p \vee q)$. One must remove either the 1 link or the 2 link in order to prevent the 1–2 path to the world exhibiting $\neg(p \vee q)$. The choice is which link to remove. Both outcomes reveal*

Figure 5.4: M_5 and M_6 (Example 5.2.1).

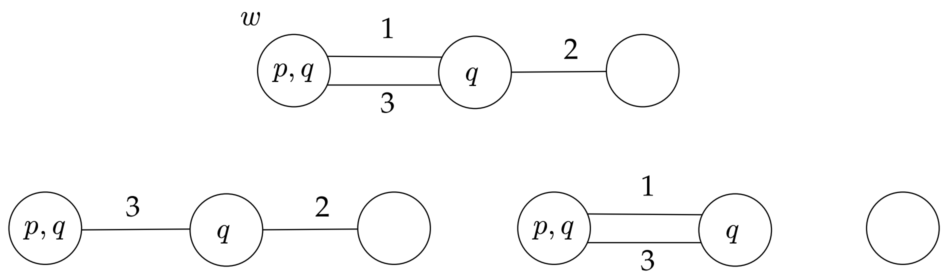

Figure 5.5: Two outcomes for refinement of the top model by $\Box_1\Box_2(p \vee q)$ (Example 5.2.2).

undesirable properties of the refinement operator. In the first case, removing the 1 link adds too much to 1's knowledge (he learns p), while the second case gives us a situation in which a model satisfies $\Box_3\Diamond_2\neg q$ but its refinement by $\Box_1\Box_2(p \vee q)$ does not. It is counterintuitive that 3's knowledge should change in this way when we refine by $\Box_1\Box_2(p \vee q)$.

The second case at least has the desirable property that a minimal change of the knowledge of agents at the actual world w is made, since the set of reachable states from w is maximised (cf. Theorem 5.6).

Example 5.2.3. *Refinement by universal formulae (Definition 5.1) ought to be cumulative, and such formulae ought to commute with each other (i.e. $M * \phi * \psi = M * \psi * \phi$). However, another example shows that this will be hard to achieve. Consider the model M_7 shown at the top of Figure 5.6, and let $\phi = \Box_1 p$ and $\psi = \Box_1\Box_2(p \vee q)$. Whatever way one thinks about defining the refinement operation, the result in the left-hand branch seems clear. Note that $M_7 * \Box_1 p$ already satisfies $\Box_1\Box_2(p \vee q)$ and therefore $M_7 * \Box_1 p * \Box_1\Box_2(p \vee q) = M_7 * \Box_1 p$.*

An argument for the stated result of $M_7 * \Box_1\Box_2(p \vee q)$ was given in Example 5.2.2, and further refining by $\Box_1 p$ leaves little room for maneuver. However, the resulting models differ on whether they satisfy (for example) $\Box_3\Box_2 q$.

Example 5.2.3 shows that even universal formulae, do not enjoy commutativity in any reasonable refinement setting. However, commutativity for universal formulae seems intuitively correct: the order in which ideal agents acquire information should not matter. Non-universal formulae are a different matter, since they can express absence of knowledge, and this will not commute with the acquisition of new knowledge.

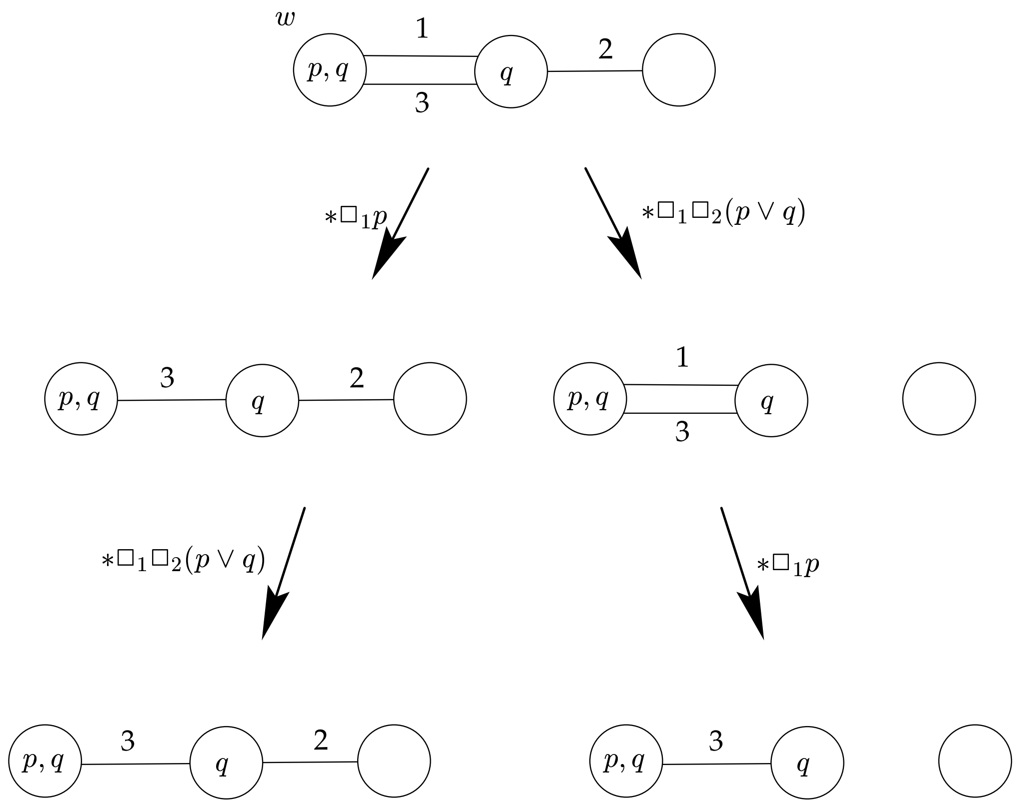

Figure 5.6: Two evolutions of M_7 (Example 5.2.3), showing that $M * \phi * \psi \neq M * \psi * \phi$.

5.3 Refining Kripke trees

Some of the problems exhibited by the three examples at the end of the preceding section seem to be due to the following fact: when we remove a link in a Kripke model in order to block a certain path, we also block other paths that used that link. To overcome this problem, we would like to unravel Kripke models into trees, in which each link participates in just one path. At first sight this looks like it will destroy the finiteness of our models, a feature on which effective refinement operators and model checking operators rely. To retain finiteness, we will need to limit in advance the maximum nesting of boxes that is allowed, and construct a tree to depth greater than this number. Semantic structures similar to Kripke trees have been defined in [HC84]. Our definition differs in detail from the one in [HC84], but it largely agrees with it in spirit.

In this section we define the notion of Kripke tree, show a translation of equivalence Kripke models into Kripke trees and define an algorithm for refining knowledge structures.

5.3.1 Kripke trees: basic definitions

Definition 5.7 (Kripke tree). *A* Kripke tree $T = (V, E, \sigma)$ *is*

- *a set V; elements of V are called* vertices*;*
- *an A-indexed family $E = \{E_i\}_{i \in A}$ of* edges $E_i \subseteq V \times V$*, such that the structure (V, E) forms a tree, that is,*
  - *there is a unique vertex $v_0 \in V$ such that for all $v \in V$ and $i \in A$, $(v, v_0) \notin E_i$. The vertex v_0 is called the* root *of T.*
  - *for every vertex v there is a unique and finite* path *from the root to v, i.e. unique sequences $(v_0, v_1, \ldots, v_k)$ and $(i_1, \ldots, i_k)$ such that $(v_j, v_{j+1}) \in E_{i_{j+1}}$ ($0 \leq j < k$) and $v_k = v$.*
- *a function $\sigma : P \to 2^V$, called* interpretation*.*

We write E^* to mean the transitive closure of the union of relations in E, i.e. $(v, v') \in E^*$ if there is a path from v to v', i.e. sequences $(v_0, v_1, \ldots, v_k)$ and $(i_1, \ldots, i_k)$ such that $(v_j, v_{j+1}) \in E_{i_{j+1}}$, with $0 \leq j < k$, $v_0 = v$ and $v_k = v'$.

We also allow the empty tree $(\emptyset, \emptyset, \emptyset)$ which we write as $\perp$. It has no root.

Definition 5.8 (Generated Kripke tree). *Let $M = (W, \sim, \pi, w_0)$ be an equivalence Kripke model. The Kripke tree $T_M = (V, E, \sigma)$* generated by *M is given as follows:*

- *The set of* vertices *is the set of paths in M:*

$$V = \left\{ (w_0, i_1, w_1, \ldots, w_{k-1}, i_k, w_k) \;\middle|\; (w_j, w_{j+1}) \in \sim_{i_{j+1}}, (0 \leq j < k) \right\}$$

- *E is an A-indexed family of edges. For $s, s' \in V$, there is an i-edge between s, s', written $(s, s') \in E_i$, if s' equals s extended by an i-link, i.e. $s = (w_0, i_1, w_1, \ldots, w_k), s' = (w_0, i_1, \ldots, w_k, i, w)$ for some w.*
- *The* valuation *σ is defined by $\sigma(p) = \{(w_0, i_1, w_1, \ldots, w_k) \mid w_k \in \pi(p)\}$.*

The vertex $w_0 \in V$ is the root of the tree.

When the model M is clear from the context or not relevant we will simply indicate the tree as T.

Generated Kripke trees are irreflexive, intransitive, anti-symmetric, anti-convergent and serial.

If the model M has at least two distinct worlds related by some relation $\sim_i$, then the tree T_M is infinite. For our purposes of model refinement, we usually want to deal with finite trees. The tree T_M^k is the tree T_M with paths truncated at length k. Obviously by truncating the tree we will lose seriality.

Definition 5.9 (Truncated tree of depth k). *Given a tree $T = (V, E, \sigma)$, the truncated tree of depth k is defined as $T^k = (V', E', \sigma')$, where*

- $V' = \{v \in V \mid$ *the distance of v from the root is less or equal than* $k\}$.
- $E' = E|_{V'}$ *is the restriction of E to V',*
- $\sigma' = \sigma|_{V'}$ *is the restriction of σ to V'.*

Infinite and finite trees satisfy modal formulae in the expected way:

Definition 5.10 (Interpretation). *Let ϕ be a formula, and T a tree. The satisfaction of ϕ by T at vertex v, written $T \models_v \phi$, is inductively defined as follows:*

- $T \models_v p$ *if* $v \in \sigma(p)$;
- $T \models_v \neg\phi$ *if not* $T \models_v \phi$;
- $T \models_v \phi \wedge \psi$ *if* $T \models_v \phi$ *and* $T \models_v \psi$;
- $T \models_v \Box_i\phi$ *if for all* $v' \in V$, $(v, v') \in E_i$ *implies* $T \models_{v'} \phi$;
- $T \models_v C\phi$ *if for all* $v' \in V$, $(v, v') \in E^*$ *implies* $T \models_{v'} \phi$.

The tree T satisfies ϕ, written $T \models \phi$, if it satisfies ϕ at its root. The empty tree $\perp$ satisfies no formula.

An infinite tree T_M is semantically equivalent to its generating model M as the following shows:

Lemma 5.11. *Let $M = (W, \sim, \pi, w_0)$ be an equivalence Kripke model and $T_M = (V, E, \sigma)$ its associated Kripke tree. Let $v = (w_0, i_1, w_1, \ldots, w)$ be any vertex ending in w, and ϕ any formula. Then:*

$$M \models_w \phi \quad \textit{if and only if} \quad T_M \models_v \phi.$$

Proof. By induction on the structure of ϕ. The result holds for atoms by construction. For the inductive case observe that there is a one-to-one correspondence between paths in M from w and extensions of the path represented by vertex v. □ □

In particular, Lemma 5.11 applies to valuations at the root of the tree, corresponding to the actual world of the model.

Corollary 5.12. $M \models \phi$ *if and only if* $T_M \models \phi$.

For the case of truncated trees, Lemma 5.11 is not valid. However, we can prove a related result for formulae up to a certain level of modal nesting.

We inductively define the *rank* of a formula as follows:

Definition 5.13 (Rank of a formula). *The rank $rank(\phi)$ of a formula ϕ is defined as follows:*

- $rank(p) = 0$, *where p is a propositional atom.*
- $rank(\neg\phi) = rank(\phi)$.
- $rank(\phi_1 \wedge \phi_2) = max\{rank(\phi_1), rank(\phi_2)\}$.
- $rank(\phi_1 \vee \phi_2) = max\{rank(\phi_1), rank(\phi_2)\}$.
- $rank(\Box_i\phi) = rank(\phi) + 1$.
- $rank(C\phi) = \infty$.

The rank of a formula ϕ intuitively represents the maximum number of nested modalities that occur in ϕ. If an operator C occurs in ϕ we take the value of $rank(\phi)$ to be infinite. The rank of a formula reflects the maximal length of any path that needs to be explored to evaluate ϕ on an infinite tree. In other words, to evaluate a formula ϕ of rank k at w_0 we need not examine worlds whose distance from w_0 is greater than k.

Lemma 5.14. *If $rank(\phi) \leq k$, $M \models \phi$ if and only if $T_M^k \models \phi$.*

Proof. By corollary 5.12, $M \models \phi$ if and only if $T_M \models \phi$, but, by induction, the evaluation of a formula of $rank(\phi) \leq k$ does not involve the evaluation of nodes of depth greater than k. So $T_M \models \phi$ if and only if $T_M^k \models \phi$, which gives the result. $\Box$

In the following we shift our attention from an equivalence Kripke model to its truncated generated tree. Truncated generated trees satisfy $S5_n$-axioms provided that the rank of the formulae is sufficiently small compared to the size on the tree. The following clarifies under which circumstances $S5_n$-axioms are satisfied at the root of the tree and that $S5_n$-inference rules are sound.

Lemma 5.15. *Let M be an equivalence model and T_M^k its generated tree truncated at k.*

1. $T_M^k \models \phi$, *where ϕ is a tautology, and $rank(\phi) \leq k$.*
2. $T_M^k \models \Box_i(\phi \Rightarrow \psi) \Rightarrow \Box_i\phi \Rightarrow \Box_i\psi$, *where* $\max\{rank(\phi), rank(\psi)\} \leq k - 1$.
3. $T_M^k \models \Box_i\phi \Rightarrow \phi$, *where $rank(\phi) \leq k - 1$.*
4. $T_M^k \models \Box_i\phi \Rightarrow \Box_i\Box_i\phi$, *where $rank(\phi) \leq k - 2$.*
5. $T_M^k \models \Diamond_i\phi \Rightarrow \Box_i\Diamond_i\phi$, *where $rank(\phi) \leq k - 2$.*
6. *If for every vertex $v \in V$ of T_M^k, $T_M^k \models_v \phi$, then for every $v \in V$, $T_M^k \models_v \Box_i\phi$, for any $i \in A$.*
7. *If for every vertex $v \in V$ of T_M^k, $T_M^k \models_v \phi$, and $T_M^k \models_v \phi \Rightarrow \psi$ then $T_M^k \models_v \psi$.*

Proof. We prove item number 4; the others can be done similarly. Suppose $T_M^k \not\models \Box_i\phi \Rightarrow \Box_i\Box_i\phi$, where $rank(\phi) \leq k - 2$. Since T_M^k is generated by M, and $rank(\Box_i\phi \Rightarrow \Box_i\Box_i\phi) \leq k$, then by Lemma 5.14 we have $M \not\models \Box_i\phi \Rightarrow \Box_i\Box_i\phi$. But by hypothesis M is an equivalence model. This is absurd. $\Box$

Before we proceed further, we introduce a few basic definitions and operations on subtrees.

Definition 5.16 (Rooted-subtrees). *Let $T' = (V', E', \sigma')$, $T = (V, E, \sigma)$ be trees with roots v_0', v_0. The tree T' is a* rooted subtree *of T, written $T' \leq T$, if $v_0 \in V', V' \subseteq V$, $E|_{V'} = E'$, and $\sigma|_{V'} = \sigma'$.*

Definition 5.17 (Intersection of trees). *Let $T' = (V', E', \sigma')$ and $T = (V, E, \sigma)$ be trees such that $\sigma|_{V' \cap V} = \sigma'|_{V' \cap V}$. The intersection of T and T' is $T \sqcap T' = (V' \cap V, E' \cap E, \sigma'|_{V' \cap V})$.*

It is easy to see that definition 5.17 (when applicable) defines a tree.

Definition 5.18 (Restriction of trees). *Let $T = (V, E, \sigma)$ be a tree with root v, and V' a subset of V. The restriction of T to V', written $T|_{V'}$, is the largest rooted subtree of T generated by v whose vertices are in V'. If the root of T is not in V', then $T|_{V'} = \bot$.*

5.3.2 Kripke trees: refinement

In Section 5.2.4, we discussed the difficulties that arise when using equivalence Kripke models as knowledge structures for refinement. Example 5.2.3 showed that any straightforward procedure to refine an equivalence Kripke model will be non-commutative even for universal formulae, i.e. there will be universal α, β, such that $M * \alpha * \beta \not\equiv M * \beta * \alpha$.

Commutativity for universal formulae can be achieved by shifting to Kripke trees. Before we can show this, we must define refinement on Kripke trees.

The typical working scenario in which we operate is the same one as that advocated by [HV91], except that we refine T_M^k instead of M. It can be described as follows: we are given an initial configuration of a MAS, and a set of formulae $\{\phi_1, \ldots, \phi_m\}$ that represent the update of the scenario. The question is whether the updated configuration will validate a set of formulae $\{\psi_1, \ldots, \psi_l\}$. We assume every ψ to have finite rank, i.e. we cannot check a formula containing the operator of common knowledge. There is no restriction on the formulas ϕs.

Our method operates as follows:

1. Start from the most general equivalence Kripke model M that represents the MAS.

2. Generate the infinite tree T_M, as given in Definition 5.8.

3. Generate from T_M^k, the truncated tree of depth k, for some sufficiently large k.

4. Sequentially refine T_M^k with $\{\phi_1, \ldots, \phi_m\}$.

5. Check whether the resulting tree structure satisfies $\{\psi_1, \ldots, \psi_l\}$.

The method described above needs some further explanation. First, what is the most general Kripke model representing a MAS configuration? How are we to build it? Our answer is the same as that given by Halpern and Vardi. Assume the set of atoms P is finite, as we set it to be in Section 5.1.1. We take the model whose universe W is equal to 2^P with an interpretation that covers all the possible assignments to the atoms. We take the relations $\sim_i, i \in A$ to be the universal relations on $W \times W$, and w_0 to be the actual world of the given MAS.

In general we will require that M is more specific than the most general model, e.g. some agent will have certain knowledge about the world. We can add all the formulae that need be satisfied to the set of updates $\{\phi_1, \ldots, \phi_m\}$. For example in the muddy children example we can start from the model with universal relations and add

$$\bigwedge_{\substack{i,j \in \{1,2,3\} \\ i \neq j}} C(p_i \Rightarrow K_j p_i)$$

to the set of updates.

We have already explained how to execute steps 1, 2, 3, and 5. We now present a notion of refinement to execute step 4.

Definition 5.19 (Refinement of Kripke tree structures). *Given a truncated Kripke tree $T = (V, E, \sigma)$, a point $v \in V$, and a formula ϕ, the result $T' = (T, v) * \phi$ of refining T by ϕ at v is procedurally defined as follows. We assume that the negation symbols in ϕ apply only to atomic propositions (to achieve this, negations may be pushed inwards using de Morgan laws and dualities $\Box/\Diamond$ and C/B).*

- *If $T = \bot$, then $T' = \bot$.*
- *If $T \models_v \phi$, then $T' = T$.*
- *Otherwise the result is defined inductively on ϕ:*
  - $\phi = p$. $T' = \bot$.
  - $\phi = \neg p$. $T' = \bot$.
  - $\phi = \psi \wedge \chi$. $T' = ((T, v) * \psi) \sqcap ((T, v) * \chi)$.
  - $\phi = \psi \vee \chi$. *If $(T, v) * \psi \leq (T, v) * \chi$ then $T' = (T, v) * \chi$, and if $(T, v) * \chi \leq (T, v) * \psi$ then $T' = (T, v) * \psi$. Otherwise T' is non-deterministically given as $(T, v) * \psi$ or $(T, v) * \chi$.*
  - $\phi = \Box_i \psi$. *T' is given by computing as follows:*

    $T' := T$;
    for each v' such that $(v, v') \in E_i$ do
        if $(T', v') * \psi = \bot$, then
            $T' := T'|_{V - \{v'\}}$
        else
            $T' := (T', v') * \psi$

  - $\phi = \Diamond_i \psi$. *Let X be the set $X = \{(T, v') * \psi \mid (v, v') \in E_i\}$.*

    If $X = \emptyset$, then $T' = \bot$,
    else T' is non-deterministically chosen to be a $\leq$-maximal element of X.

  - $\phi = C\psi$. *T' is given by computing as follows:*

$T' := T$;
for each v' *such that* $(v, v') \in E^*$ *do*
  if $(T', v') * \psi = \bot$, *then*
    $T' := T'|_{V-\{v'\}}$
  else
    $T' := (T', v') * \psi$

– $\phi = B\psi$. *Let* X *be the set* $X = \{(T, v') * \psi \mid (v, v') \in E^*\}$.

If $X = \emptyset$, *then* $T' = \bot$,
else T' *is non-deterministically chosen to be a* $\leq$*-maximal element of* X.

$T * \phi$ *means* $(T, v) * \phi$, *where* v *is the root of* T.

Lemma 5.20. *Given a tree* T, *a formula* α *and a point* v, $(T, v) * \alpha$ *is a tree.*

Proof. It follows from the fact that if T is a tree then $T|_{V'}$ is also a tree. □

The intuition behind $(T, v) * \phi$ is that it is obtained by removing as small a set of links from T as possible, in order to satisfy ϕ. Note that, due to the clauses for the connectives $\vee, \Diamond_i, B$, we have that the tree $(T, v) * \phi$ is not uniquely defined. However, we will see that running the procedure on the muddy children example does not introduce non-determinism.

5.4 The muddy children puzzle using Kripke trees

In Section 5.2.1, we described the muddy children puzzle and we reported the formalisation that was given in [FHMV95], [HV91]. The aim of the present section is to solve an instance of it (where the actual situation is coded by the tuple p_1, p_2, p_3 that we equivalently write as $(1, 1, 1)$; all the children are muddy) by using Kripke trees and the methods we introduced in Section 3.

We start with the most general model to represent the puzzle: this is the model M_1 of Figure 5.1[3]. Given M_1, we generate the infinite tree T_{M_1} for M_1 and then the truncation T_1 of T_{M_1}. In this example we truncate at level, say, ten. The starting tree and the three successive refinements are in Figure 5.7, and 5.8 (shown to three levels). Let $\phi_1 = C(p_1 \vee p_2 \vee p_3)$ (this is the father's announcement), and $\phi_2 = C(\neg\Box_1 p_1 \wedge \neg\Box_1\neg p_1) \wedge C(\neg\Box_2 p_2 \wedge \neg\Box_2\neg p_2) \wedge C(\neg\Box_3 p_3 \wedge \neg\Box_3\neg p_3)$ (the children's simultaneous reply that they don't know whether or not they are muddy). We now sequentially update T_1 by ϕ_1 and then by ϕ_2 three times. Note that since all children are muddy, they will have to speak three times before everyone knows he is muddy.

Consider the algorithm of Definition 5.19 and T_1. Following the algorithm, the refined tree $T_1 * \phi_1 = T_2$ in Figure 5.7 is T_1 in which the links to states where no children are muddy

[3]According the the notion of most general model as described in Section 5.3.2 the model M should actually be $M = (2^{\{p_1,p_2,p_3\}}, U, \pi, w)$, where U is a family of universal relations on $W \times W$, and $\pi(w) = \{p_1, p_2, p_3\}$. The model M_1 we analyse is the result of the update of M by

$$C(p_i \Rightarrow K_j p_i) : i \neq j; i, j \in \{1, 2, 3\},$$

where the formula above represents the fact that children can see each other. For brevity (as in [FHMV95], [HV91]) we start our analysis from M_1; i.e. rather than building the tree for M and update it first by $C(p_i \Rightarrow K_j p_i)$, we directly build the tree for M_1. The reader can check that this leads to the same result.

have been removed. The tree $T_3 = T_2 * \phi_2$ (shown in Figure 5.8) is then constructed by isolating worlds that do not see two worlds for every relation. In fact, only in this case one of the formulae $\Diamond_i p_i \wedge \Diamond_i \neg p_i$ can fail on a point of T_2. We can now obtain T_4 similarly.

Having made all the refinements, we can now check whether or not the muddy children know that they are muddy. This involves checking

$$T_4 \models \bigwedge_{i=1}^{3} (p_i \Rightarrow K_i p_i),$$

which is indeed the case.

Analogously we can prove that the procedure given in Section 5.3 produces solutions for the other cases of the muddy children.

Note that had we decided to consider the Kripke tree truncated at $n \geq 4$, the formula $\bigwedge_{i=1}^{3}(p_i \Rightarrow K_i p_i)$ would still be satisfied at the root after three refinements.

Let us now consider the example presented in Section 5.2.2. By following the above described procedure with the assumption of synchronicity, the k faulty robots will announce their fault and disconnect from the system after k rounds, allowing the system to start normal production again and substitute the faulty units.

5.5 Properties of refinement on Kripke trees

In the rest of this chapter we analyse some more properties of the refinement procedure that we defined in Definition 5.19.

The first remark that we should make is that refining a scenario by some agent's knowledge cannot affect other agents' knowledge, as was the case in Example 5.2.1 for Kripke models. This is because by unravelling a Kripke model we produce a tree whose leaves are in a bijection with paths of the original model. We formalise this as follows:

Theorem 5.21. *Let T be a tree, and ϕ, ψ two formulae, we have the following:*

$$\textit{If } T \models \Box_i \phi \textit{ then } T * \Box_j \psi \models \Box_i \phi, \textit{ with } i \neq j.$$

Proof. Nodes of a Kripke tree are in a bijection with paths of the generating model. Therefore by removing some j-links we cannot affect the interpretation of any modality whose index is not j. The only problematic case would arise if $i = j$ and $T * \Box_j \psi = \bot$, but this is excluded by hypothesis. $\Box$

Although the theorem above refers to infinite trees, an analogue version can be proved for truncated trees.

The second point worth stressing is that Kripke trees solve the problem of Example 5.2.3, i.e. we can prove commutativity although the result is limited to safe formulae (Definition 5.2). We need a few results before achieving this.

Lemma 5.22. *Let $T_1 = (V_1, E_1, \sigma_1), T_2 = (V_2, E_2, \sigma_2)$ be trees. The following hold.*

1. $(T, v) * \phi \leq T$.
2. *If α is disjunction-free, then $T_1 \leq T_2$ implies $(T_1, v) * \alpha \leq (T_2, v) * \alpha$, where $v \in V_1 \cap V_2$.*

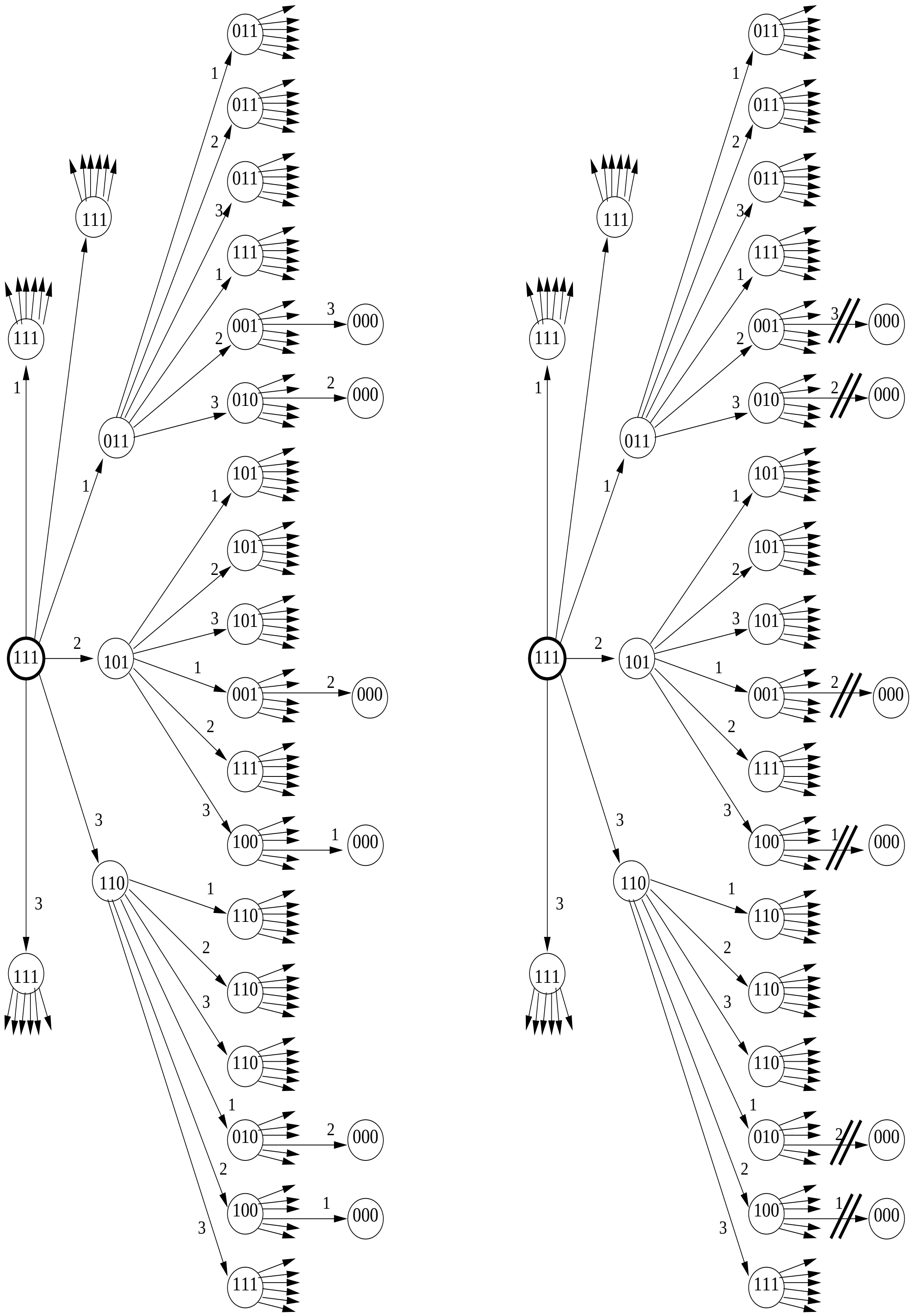

Figure 5.7: T_1, T_2: The Kripke trees before and after the father speaks.

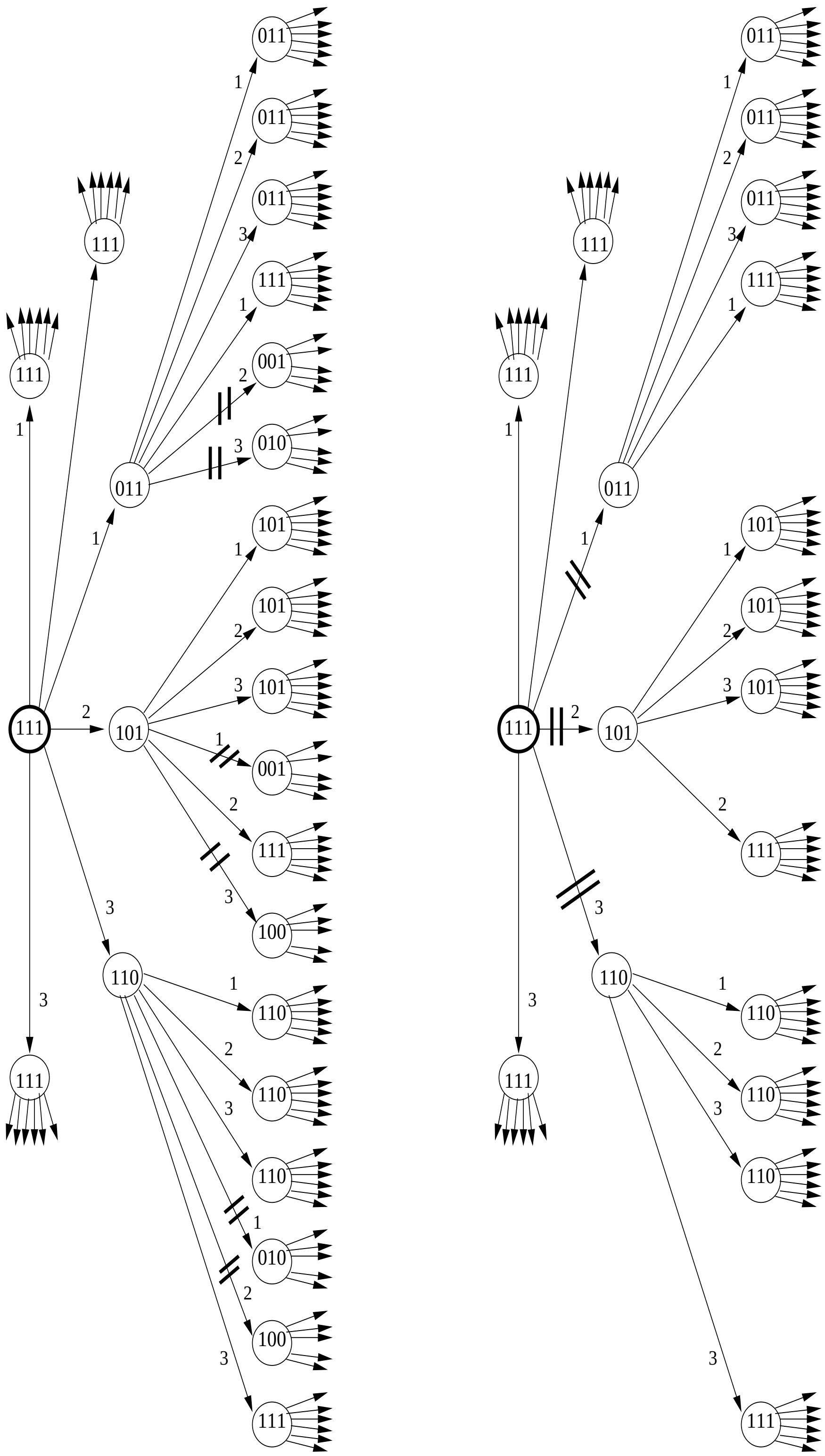

Figure 5.8: T_3, T_4: The Kripke trees after the children speak the first and second time.

3. *If α is universal then $T \models \alpha$, $T' \neq \bot$, $T' \leq T$ imply $T' \models \alpha$.*

Proof. 1. The procedure for obtaining $(T, v) * \phi$ only removes links or produces the empty tree. Therefore we have the result.

2. We perform structural induction on α. Let $T_1' = (T_1, v) * \alpha$ and $T_2' = (T_2, v) * \alpha$. Suppose α is of the form:

   - $\alpha = p$. If $v \in \sigma(p)$ then $T_1' = T_1$, $T_2' = T_2$; else $T_1' = T_2' = \bot$.
   - $\alpha = \neg p$. If $v \notin \sigma(p)$ then $T_1' = T_2' = \bot$; else $T_1' = T_1$, $T_2' = T_2$.
   - $\beta \wedge \gamma$.
   $$\begin{array}{rcll} (T_1, v) * \alpha & = & (T_1, v) * \beta \sqcap (T_1, v) * \gamma & \\ & \leq & (T_2, v) * \beta \sqcap (T_2, v) * \gamma & \text{Induction hypothesis} \\ & = & (T_2, v) * \alpha & \end{array}$$
   - $\alpha = \Box_i \beta$. Set $T_1' = T_1$ and $T_2' = T_2$ and we execute the loops of Definition 5.19 ($\Box_i$-case) synchronously. We will show that $T_1' \leq T_2'$ is an invariant of the execution. Suppose $(v, v') \in E_{2i}$.
     - If $(v, v') \in E_{1i}$, then consider the following cases:
       * $(T_1', v') * \beta = \bot$ and $(T_2', v') * \beta = \bot$.
         $T_1' := T_1'|_{V-\{v'\}}$ and $T_2' := T_2'|_{V-\{v'\}}$, so $T_1' \leq T_2'$ is not violated.
       * $(T_1', v') * \beta = \bot$ and $(T_2', v') * \beta \neq \bot$.
         $T_1' := T_1'|_{V-\{v'\}}$ and $T_2' := (T_2', v') * \beta$; so $T_1' \leq T_2'$.
       * $(T_1', v') * \beta \neq \bot$ and $(T_2', v') * \beta = \bot$.
         Contradicts hypothesis that $T_1' \leq T_2'$.
       * $(T_1', v') * \beta \neq \bot$ and $(T_2', v') * \beta \neq \bot$.
         $T_1' := (T_1', v') * \beta$, $T_2' := (T_2', v') * \beta$, and $T_1' \leq T_2'$ by induction hypothesis.
     - If $(v, v') \notin E_{1i}$ then T_1' is unchanged by the body of the loop, while T_2' becomes one of $T_2' := T_2'|_{V-\{v'\}}$ and $(T_2', v') * \beta$. In either case, we are removing links in T_2 which are not present in T_1, so $T_1' \leq T_2'$ is preserved.
   - $\alpha = E\beta$. It follows by induction hypothesis by noting that $E\beta = \wedge_{i=1}^{n} K_i \beta$.
   - $\alpha = C\beta$. Similar to $\Box_i \beta$, but with proofs related to E^*.

3. It follows from structural induction on α.

□

Theorem 5.23 (Success). *If α is universal, $(T, v) * \alpha = \bot$ or $(T, v) * \alpha \models_v \alpha$.*

Proof. Induction on α. The cases $\alpha = p, \neg p, \psi \vee \chi, \Box_i \psi, E\phi, C\psi$ are straightforward; we prove the case $\alpha = \psi \wedge \chi$.

$(T, v) * (\psi \wedge \chi) = (T, v) * \psi \sqcap (T, v) * \chi$. But by induction hypothesis we have that $(T, v) * \psi \models \psi$ and that $(T, v) * \chi \models \chi$. Since $(T, v) * \psi \leq (T, v) * \psi \sqcap (T, v) * \chi$ and $(T, v) * \chi \leq (T, v) * \psi \sqcap (T, v) * \chi$, by part 3 of Lemma 5.22, we have that $(T, v) * \psi \sqcap (T, v) * \chi \models \psi$ and that $(T, v) * \psi \sqcap (T, v) * \chi \models \chi$. So we have that $(T, v) * \psi \sqcap (T, v) * \chi \models \phi \wedge \chi$. □

Lemma 5.24. *If α is safe, then the outcome of $(T, v) * \alpha$ is deterministically defined.*

Proof. Suppose ϕ contains no $\Box_i, C$ operators. Then it is an easy induction to see that $(T, v) * \phi$ is either T or $\perp$. Now consider $(T, v) * (\phi \vee \psi)$, where ϕ, ψ are $\Box_i, C$-free. We see that either $(T, v) * \phi \leq (T, v) * \psi$ or $(T, v) * \psi \leq (T, v) * \phi$, so the result is again T or $\perp$. The cases $\Box_i \phi, C\phi$ do not introduce non-determinism. ☐

We show that, for universal formulae, the change made by a refinement is the minimal one possible in order to satisfy the formula:

Theorem 5.25. *If α is safe, then the tree $(T, v) * \alpha$ is $\leq$-maximum in $\{T' \leq T \mid T' \models_v \alpha$ or $T' = \perp\}$.*

Proof. Let $T' = (T, v) * \alpha$. By part 1 of Lemma 5.22 and Theorem 5.23, we know T' is in the set. To prove that it is maximum, take any T'' in the set; we will show $T'' \leq T'$. If $T'' = \perp$ the result is immediate; otherwise, we have $T'' \models_v \alpha$ and $T'' \leq T$. Since $T'' \leq T$, we get $(T'', v) * \alpha \leq (T, v) * \alpha$ by part 2 of Lemma 5.22. But $(T'', v) * \alpha = T''$ (since it is already $T'' \models_v \alpha$) and $(T, v) * \alpha = T'$; so $T'' \leq T'$. ☐

Theorem 5.26. *If α, β are safe, then the tree $(T, v) * \alpha * \beta$ is maximum in $\{T' \leq T \mid T' \models_v \alpha \wedge \beta$ or $T' = \perp\}$.*

Proof. Let $T' = (T, v) * \alpha * \beta$. By parts 1 and 3 of Lemma 5.22 and Theorem 5.23, we know T' is in the set. The argument that it is maximum is similar to the proof of Theorem 5.25. Take any T'' in the set; we will show $T'' \leq T'$. If $T'' = \perp$ the result is immediate; otherwise, we have $T'' \models_v \alpha \wedge \beta$ and $T'' \leq T$. Since $T'' \leq T$, we get $(T'', v) * \alpha * \beta \leq (T, v) * \alpha * \beta$ by part 2 of Lemma 5.22. But since $T'' \models_v \alpha$ we have $(T'', v) * \alpha = T''$ and since $T'' \models_v \beta$ we have $(T'', v) * \beta = T''$. So $(T'', v) * \alpha * \beta = T''$ and $(T, v) * \alpha * \beta = T'$. Therefore we have $T'' \leq T'$. ☐

Theorem 5.27 (Commutativity). *If α, β are safe, then $T * \alpha * \beta = T * \beta * \alpha$.*

Proof. By Theorem 5.26, $T * \alpha * \beta$ and $T * \beta * \alpha$ are maxima in the same set. Therefore they are equal. ☐

It is worth mentioning an example of which non-universal formulae can make commutativity to fail, independently of non-determinism.

Example 5.5.1. *Commutativity can fail for arbitrary formulae. The problem is that if the formulae are non-universal, the order of updating can play a role in the outcome of the update and we might have that one of the two cases fail. The example we report here is the tree T_5, illustrated in Figure 5.9, where the root is the top vertex. Consider now $T_6 = T_5 * \Diamond_1 \neg p * \Box_1(p \vee \neg q)$, illustrated, and $T_7 = T_5 * \Box_1(p \vee \neg q) * \Diamond_1 \neg p = \perp$.*

5.6 Conclusions

This chapter is related to Section 3.8. While that section was focused on the study of hypercubes, there we saw a low-level formal technique for reasoning explicitly about the evolution of the local states of the agents. Here we isolated the case of internal actions consisting in the update of the agents' knowledge and we showed that Kripke semantics can still be a useful modelling tool.

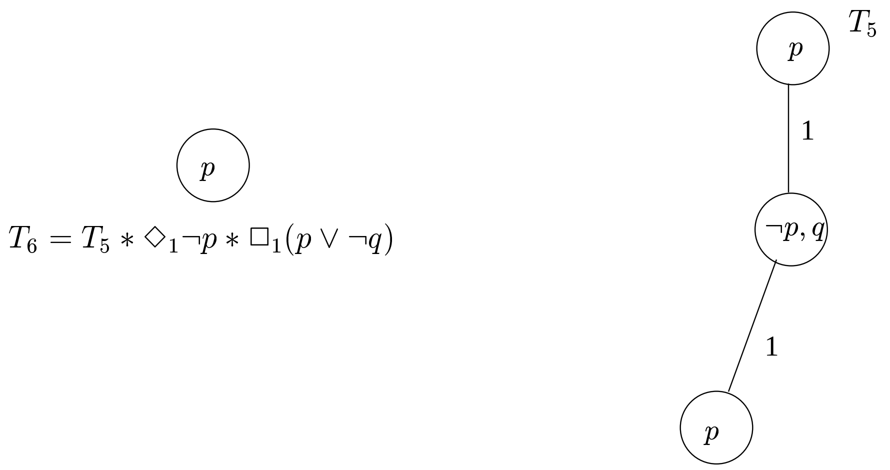

Figure 5.9: T_5 and T_6 discussed in Example 5.5.1. While $T_6 = T_5 * \Diamond_1 \neg p * \Box_1 (p \vee \neg q)$ is defined and shown above, $T_7 = T_5 * \Box_1 (p \vee \neg q) * \Diamond_1 \neg p$ is undefined.

The work of this chapter builds upon the influential [HV91] in which model refinement and model checking were first proposed in the case of modelling knowledge for MAS. With respect to that paper, this chapter should be seen as an attempt to clarify issues such as which are the appropriate semantic structures to use, which algorithms for refinement seem to be promising and what properties this refinement should have. We argued that model refinement could not be defined satisfactorily on Kripke models, and proposed a definition on Kripke trees obtained from Kripke models instead. The shift from Kripke models to Kripke trees let us achieve two technical results. First, we showed that it is possible to refine trees by a formula expressing knowledge of an agent without affecting the knowledge of the other agents (Theorem 5.27). Note that this was not apparently possible on standard Kripke models (see Example 5.2.1). Secondly, while it seems impossible to obtain commutativity even for safe formulae on Kripke models, we showed this is possible for Kripke trees.

More generally, the chapter (by following the approach suggested in [HV91]) tries to answer one of the critics that are often addressed to using Kripke semantics for representing knowledge in MAS scenarios. These critics involve essentially the fact that Kripke semantics is "ungrounded" (see for example [Woo97]). By "ungrounded" it is meant that there is no clear correspondence between possible worlds semantics and the physical state (in this case the knowledge of the agents) the agents are in. By providing a notion of refinement on Kripke trees and study how these evolve it is hoped that a more direct correspondence between physical processes and Kripke semantics can be drawn.

It should be added that many of the issues discussed in this chapter seem worth investigating further. For example, although it is reasonable to assume that the logic of Kripke trees (as they were defined in Definition 5.7) is simply K_n because of the too weak characteristics being imposed on the frames, it remains to be seen whether there is actually a logic for generated Kripke trees (Definition 5.8). What about the Kripke frames that result from the process of updating? We know that generated Kripke frames satisfy axioms of $S5_n$ (Lemma 5.15) up to a certain depth, but it is hard to think of a suitable restriction that will work for arbitrary truncated trees that result from an update.

Another interesting point would be to investigate whether there exist circumstances in which commutativity fails for universal formulae in which no non-deterministic choices have to be performed. In order to investigate further properties of the algorithm given in Definition 5.19 it would seem appropriate to build an implementation of it.

In this chapter, the algorithm of Definition 5.19 was tested only on the muddy children problem. In [Dav99] this algorithm has been implemented in JAVA and results for the muddy children problem have been verified. [Dav99] confirmed that the algorithm produces the expected result also for other scenarios such as the three wise men puzzle.

It would also seem promising to compare the approach presented in this chapter to a recent work [Ger99], where similar themes are investigated by using non-well-founded set theory. A solution to the problem of refinement on Kripke models by the use of trees is also reported there, although the technical machinery is different.

Also related to this chapter is some recent work by Baltag and colleagues ([BMS98]). [BMS98] develops a logic of epistemic actions. The machinery for updating Kripke models developed there involves making copies of worlds when needed. The work is more concerned with the expressivity of a formal language that can express announcements and it has a somewhat different focus from this chapter. It would seem promising to extend the multi-modal syntax used in this chapter to include the operators discussed in [BMS98].

The work presented in this chapter was developed independently from [Ger99] and [BMS98] and the authors were not aware of each other's ongoing research.

A final note concerns Section 5.2.2 in which we discussed an example in which the ideas of this chapter can be applied. This consisted in a collective diagnosis problem among a group of homogeneous robots working at a factory. It should be clear that the scenarios commonly analysed in collective diagnosis research (see for example [FNS98], [BD93], [JW92], [Sch98]) are somehow different from our example. Our example is much closer to ideal scenarios coming directly from robotics.

Chapter 6

Discussion: combining logics for MAS theories

6.1 Introduction

In Chapter 1 we have argued that completeness and decidability are very important features of a MAS theory. Indeed these two properties played an important role throughout this thesis. Most of Chapter 3 was focused on proving completeness for the logics that capture epistemic properties of hypercubes. Similarly, Chapter 4 investigated whether a relatively simple class of extensions of the epistemic system $S5_2$ enjoy completeness. Although we did not ask the question of decidability in Chapter 4, this was the topic of some of the discussion we carried out in Chapter 3.

It would be fair to say that proving these two properties for the logics we have examined in this thesis has not always been straightforward. In particular in Chapter 3 we have seen that the canonical model for the logic $S5WD_n$ is based on a DI frame and we had to rely on some more subtle arguments to prove its completeness. In Chapter 4 we had to leave open the completeness problem for two logics.

These were *not* unfortunate occurrences. Many of the logics for MAS proposed in the literature have complex completeness and decidability proofs. Indeed this intrinsic difficulty has seriously slowed down research in this area that can nowadays hardly keep the pace with new developments in the applications. Given this thematic problem in developing MAS theories, one could ask the question of whether or not it is possible to develop some more abstract (and hopefully powerful) tools to prove properties about these. After all, recall from Section 1.3.2 that when designing MAS theories we are very often dealing with quite similar extensions of very basic normal modal logics, so how hard can it be to prove some general results about them? These pages will discuss this question.

In Section 1.3.2.3 and Section 1.4.1 we argued that any fully-fledged MAS theory has to deal with the multiplicity of the agents and of their mental states and with the interactions between these. As an example, in Section 1.3.2.3.1 we noted that the multi-modal system presented in [KL88] for modelling knowledge and belief can be seen as composed by two different logical layers (specifically $S5_n$ for knowledge and $KD45_n$ for belief) linked together by specific interaction axioms and enriched by operators for group properties. Many other MAS logical theories can be seen in this way. They are basically well understood basic modal logics "combined together" to model different facets of the agents.

The observation above should remind us of "combining logics". Combining logics is an emerging area in logic that deals exactly with the problem of combining logic systems from an abstract point of view. Some techniques have been developed in recent years for this goal and the field is undergoing rapid development. These techniques aim to define a syntax, a semantics and a proof theory of a combination of logics. Most crucially the area investigates the problem of the transfer of properties such as completeness and decidability from the basic components into the combination.

It should be evident how relevant this area can potentially be for MAS theories. If tools for combining logics suitable to the problems of MAS theories could be identified, this would even let us re-define MAS theories altogether. MAS theorists could build formal models of agency simply by considering basic well-understood logics and study the possible interplays of the different components when combined together. If operating under the guidelines of a well-developed theory of logic combination, properties of the logics such as completeness and decidability would be inherited from the basic components. This would permit us to develop not just a general theory for MAS (that would possibly turn out to be so general that it would not model any realistic intelligent agent), but to define the criteria in order to define *appropriate* specifications of the system the AI-user has in mind.

It is fair to say that what we described above would be quite an extraordinary development; this chapter aims at answering the question of its feasibility, not just in the context of knowledge theories but more broadly for any mental aspects of MAS. It is organised as follows. In Section 6.2.2 some of the best known techniques for combining logics are presented followed by some negative results in Section 6.2.3. In Section 6.3 we will discuss whether or not some existing MAS theories can be seen as results of the application of the combining logics tools of Section 6.2.2. Section 6.4 concludes with a discussion.

Differently from the material presented so far, this chapter does not present original research by the author. Instead it is a review, carried out from the point of view of the agent theorist, of material published in the field of combining logics. Still I believe it is relevant to the rest of this thesis and it actually constitutes its natural end as it aims to provide a glimpse onto the possible future of theories for MAS.

6.2 Combining logics

6.2.1 Introduction

Combining logics is an emerging (see for example: [dRB96, BdR97a, BdR97b, Gab99]) area in logic. The discipline addresses the problem of defining techniques that permit to integrate two or more logic systems into a more expressive one.

The idea of combining logics is not new. As we have noticed before even the systems we discussed in Section 1.3.2 have in-built the idea of a combination between more logics and many other "combined" systems have been proposed in the past. The novelty of combining logics is the aim to develop *general techniques* that allow to produce combinations of *existing* and well understood logics.

Why should we interested in exploring such general techniques? For many reasons, but most importantly because it is the natural answer to the need of formalising complex systems in a *systematic* way. Combining logics recognises that the big problems of the formalisation of complex logics such those required in many areas (Artificial Intelligence but also Linguistics and Computer Science in general) require a *divide and conquer* strategy.

Combining logics suggests identifying the basic components of a system, define the corresponding formal tools and combine them to produce a *scalable* formalisation of the whole system. This methodology would allow the user to re-use previously defined and understood components (logics) and would guide him on how to combine these in a proficient way.

Technically, the discipline addresses the following problem. Given two logics L_1 and L_2 define a logic $L_1 \otimes L_2$, which is more expressive than L_1 and L_2. Note the generality of the setting. For example, the logics may be radically different: L_1 may be a system of modal logic, and L_2 a first-order fuzzy logic. Even if the logics are somehow similar, they may be presented in very different ways; for example L_1 may be described by an axiomatisation, while we may have only the semantics of L_2. In order to be able to reason about combinations, we should then identify a suitable working definition of logic system that allows to define combinations of them.

The problem is indeed extremely complex to be solved in this general setting and we cannot expect that a complete solution will be available in the near future; in the following we will discuss more modest contributions that have been proposed recently.

Let us focus our attention to well-studied logics L_1, L_2 for which we have a syntax, a semantics and a proof theory. A general technique for combining logics should then address the following points:

1. Define the syntax of $L_1 \otimes L_2$ from the syntax of L_1 and L_2.
2. Define the semantics of $L_1 \otimes L_2$ from the semantics of L_1 and L_2.
3. Define the proof theory of $L_1 \otimes L_2$ from the proof theory of L_1 and L_2.
4. Prove the transfer of important properties (soundness, completeness, decidability, fmp, etc.) of the logics L_1 and L_2 into $L_1 \otimes L_2$.

A few different techniques are discussed in the literature ([FS77, Fit69, FS96, KW91, GP92, Pfa87]) and some papers (for example [BdR97a]) relate them in order to make the overall picture of the discipline. Here we will discuss only the ones that seem to be the most promising with respect to combinations of multi-agent theories. After this we will present some negative results.

6.2.2 Positive results

6.2.2.1 Temporalising a logic

The first of the techniques we discuss is the one known as "embedding a logic into another", "adding a temporal dimension to a logic" or simply "temporalising a logic" as it is described in [FG92].

Finger and Gabbay propose a case of limited combination between a logic L and a propositional temporal logic T. In [FG92] a logic is supposed to comprehend a syntax, a semantics and a proof theory. L can be any classical logic system, while T is a classical propositional temporal logic with the binary modal operators since ($\mathcal{S}$) and until ($\mathcal{U}$). The semantics of T is given in terms of possible worlds with linear flows of time, i.e. the accessibility relation of the Kripke frames is irreflexive, transitive and total (see Section 1.2.2).

Given the logic L, the temporalisation process defines a 'temporalised' logic T(L) . Syntax, semantics and proof theory of T(L) are defined as follows.

6.2.2.1.1 Syntax. The language of the temporalised logic is structurally limited.

Definition 6.1 (Syntax of the temporalised logic). *Let $\mathcal{L}^*_{\mathrm{L}}$ be the set of monolithic formulae of L, i.e. the set of formulae of the language of L , $\mathcal{L}_{\mathrm{L}}$, that do not contain boolean connectives.*
The language $\mathcal{L}_{\mathrm{T(L)}}$ of T(L) is defined as the smallest set such that:

- *if $\phi \in \mathcal{L}^*_{\mathrm{L}}$ then $\phi \in \mathcal{L}_{\mathrm{T(L)}}$,*
- *if $\phi, \psi \in \mathcal{L}_{\mathrm{L}}$ then $\neg\phi, (\phi \wedge \psi) \in \mathcal{L}_{\mathrm{T(L)}}$,*
- *if $\phi, \psi \in \mathcal{L}_{\mathrm{L}}$ then $\phi \mathcal{U} \psi, \phi \mathcal{S} \psi \in \mathcal{L}_{\mathrm{T(L)}}$.*

Note that although the language $\mathcal{L}_{\mathrm{T(L)}}$ of the combined logic T(L) includes the two languages $\mathcal{L}_{\mathrm{L}}$ and $\mathcal{L}_{\mathrm{T}}$, not all the formulae that we could build by freely mixing the two languages are in $\mathcal{L}_{\mathrm{T(L)}}$. For example, if $\mathcal{L}_L$ is built from a set of operators containing a unary $\Box$, and ϕ and ψ are two monolithic formulae of L then the formula $\phi \mathcal{U} (\Box\psi)$ is not in $\mathcal{L}_{\mathrm{T(L)}}$.

The language $\mathcal{L}_{\mathrm{T(L)}}$ can only express temporal properties of a system specified with L; as a consequence of this, in a formula of $\mathcal{L}_{\mathrm{T(L)}}$ all the temporal operators must precede any operator of L.

6.2.2.1.2 Semantics. Consider the logic T to have a semantics defined on the class of temporal models $\mathcal{M}_{\mathrm{T}}$, where each temporal model in the class is a triple $M_{\mathrm{T}} = (W, \prec, g)$, where W is a set whose elements represent time points; the relation $\prec \subseteq W \times W$ is a linear order (i.e. an irreflexive, total and transitive relation) on such points; the function $g : P \to 2^W$ is an interpretation function for the atoms in P. Pairs $(W, \prec)$ are called temporal frames. Let $\mathcal{M}_{\mathrm{L}}$ be the class of models M_{L} for L .

Models of the temporalised logic are then defined as follows.

Definition 6.2 (Models of the temporalised logic). *A model $M_{\mathrm{T(L)}}$ for T(L) is defined as a triple $M_{\mathrm{T(L)}} = (W, \prec, h)$, where the frame $(W, \prec)$ is a temporal frame and $h : W \to \mathcal{M}_L$ is a function mapping a time point to a model of L .*

The models of the combination T(L) are combined models of L and T and can be seen as composed by two layers: an upper flow of time, and a set of models of L . The two levels are connected via the function h which associates a model of L to every point of time. The upper level is used to interpret the temporal operators, while the monolithic formulae of L are interpreted on the underlying layer of models.

More precisely, the satisfaction relation $\models$ between models of T(L), points of such models, and formulae is inductively defined as follows:

Definition 6.3 (Interpretation of the temporalised logic).

$M_{\mathrm{T(L)}} \models_t \phi$, *where* $\phi \in \mathcal{L}^*_{\mathrm{L}}$	*if*	$h(t) = M_L$ *and* $M_L \models_t \phi$.
$M_{\mathrm{T(L)}} \models_t \neg\phi$	*if*	*it is not the case that* $M_{\mathrm{T(L)}} \models_t \phi$.
$M_{\mathrm{T(L)}} \models_t (\phi \wedge \psi)$	*if*	$M_{\mathrm{T(L)}} \models_t \phi$ *and* $M_{\mathrm{T(L)}} \models_t \psi$.
$M_{\mathrm{T(L)}} \models_t \phi \mathcal{S} \psi$	*if*	*there exists a point* s, $s \prec t$ *and* $M_{\mathrm{T(L)}} \models_s \psi$ *and for all points* $u \in T$, *if* $s \prec u \prec t$ *then* $M_{\mathrm{T(L)}} \models_u \phi$.

$M_{\mathrm{T(L)}} \models_t \phi \mathcal{U} \psi$ *if there exists a point* s, $t \prec s$ *and* $M_{\mathrm{T(L)}} \models_s \psi$ *and for all points* $u \in T$, *if* $t \prec u \prec s$ *then* $M_{\mathrm{T(L)}} \models_u \phi$.

6.2.2.1.3 Proof theory. Finger and Gabbay also give an axiomatisation for T(L) .

Definition 6.4 (Axiomatisation of T(L)). *If* $(A_{\mathrm{T}}, I_{\mathrm{T}})$[1] *is an axiomatisation for T, then* $(A_T, I_T \cup$ **Preserve**) *is an axiomatisation for T(L) , where* **Preserve** *is an inference rule defined as:*

$$\text{If } \phi \text{ is in } \mathcal{L}_{\mathrm{L}} \text{ and } \vdash_{\mathrm{L}} \phi, \text{ then } \vdash_{\mathrm{T(L)}} \phi. \qquad \textbf{Preserve}$$

For the embedding technique Finger and Gabbay prove the following important result:

Theorem 6.5 (Properties of the temporalisation).

1. *If L is sound and T is sound with respect to* $\mathcal{M}_T$, *then T(L) is sound with respect to* $\mathcal{M}_{\mathrm{T(L)}}$.
2. *If L is complete with respect to* $\mathcal{M}_{\mathrm{L}}$ *and T is complete with respect to* $\mathcal{M}_{\mathrm{T}}$, *then T(L) is complete with respect to* $\mathcal{M}_{\mathrm{T(L)}}$.
3. *If L is decidable and complete with respect to* $\mathcal{M}_{\mathrm{L}}$ *and T is decidable and complete with respect to* $\mathcal{M}_{\mathrm{T}}$, *then T(L) is decidable with respect to* $\mathcal{M}_{\mathrm{T(L)}}$.
4. *If T is complete with respect to* $\mathcal{M}_T$, *then T(L) is a conservative extension of L and of T .*

The work does not analyse the transfer of other properties such as fmp or compactness. Notwithstanding this, Theorem 6.5 is a very strong result: every classical logic can be temporalised in a standard way by means of this technique.

For example, we can temporalise the predicate calculus obtaining a limited first-order temporal logic, we can temporalise a modal logic obtaining a limited multi-modal logic, etc. Finger and Gabbay are particularly interested in applying the method to a temporal logic itself ([FG92, FG96a]), that is to temporalise a temporal logic. The idea is to define a framework that allows one to reason not only about the temporal evolution of a system, but also about the temporal change of the temporal description. This can be useful in a number of cases; for example to model two computer systems with different clocks that observe each other or to represent the historiography of a system.

Another interesting result is that the embedding process can be iterated. For example, if we temporalise by using T_1a linear temporal logic T_2, which includes the since ($\mathcal{S}$) and until ($\mathcal{U}$) operators we obtain a logic $\mathrm{T}_1(\mathrm{T}_2)$, whose language $\mathcal{L}_{\mathrm{T}_1(\mathrm{T}_2)}$ comprehends two families of temporal operators ($\mathcal{S}_1, \mathcal{U}_1$ and $\mathcal{S}_2, \mathcal{U}_2$), each of them to be evaluated on the corresponding layer. According to Theorem 6.5, the logic $\mathrm{T}_1(\mathrm{T}_2)$ can be described axiomatically by the axiomatisation of T_1 and the inference rule **Preserve**. Therefore, it is now possible to temporalise the logic $\mathrm{T}_1(\mathrm{T}_2)$ using an element of $\mathcal{M}_{\mathrm{L}_2}$ as upper layer, obtaining the logic $\mathrm{T}_2(\mathrm{T}_1(\mathrm{T}_2))$. The language $\mathcal{L}_{\mathrm{T}_2(\mathrm{T}_1(\mathrm{T}_2))}$ will be composed by sentences which have some operators of T_2 followed by some operators of $\mathcal{L}_{\mathrm{T}_1}$ and then other operators of $\mathcal{L}_{\mathrm{T}_2}$. Although this is still a subset of the union of the two languages, we have that the language $\mathcal{L}_{\mathrm{T}_2(\mathrm{T}_1(\mathrm{T}_2))}$ includes $\mathcal{L}_{\mathrm{T}_1(\mathrm{T}_2)}$.

[1]We assume A_{T} to be the set of axioms and I_{T} to be the set of inference rules.

It is not really clear what is the purpose of adding more and more temporal dimension to a system can be; notwithstanding this an iteration of the embedding process does enrich the expressivity of the language and it may be very meaningful when the logics being combined are not homogeneous. This requires to drop the assumption on the external logic that would need not be temporal; the authors claim this is possible and that the role of the 'upper-level' logic can be taken by any classical modal logic, provided it is sound and complete. Indeed a general technique allowing 'to modalise' a logic would be interesting on its own.

The powerful results of Theorem 6.5 have been achieved by imposing very strong limitation on the language $\mathcal{L}_{\mathrm{T(L)}}$ of the combined logic T(L) . Unfortunately, this leads to serious problems for the expressivity of the logic that hinder many applications of the technique. For example, as discussed in Chapter 3.6 one of the key interactions that one might want to consider by combining two logics is the commutativity of two operators. If $\Box_1$ and $\Box_2$ are one of the operators that $\mathcal{L}_{\mathrm{L}_1}$ and $\mathcal{L}_{\mathrm{L}_2}$ respectively are built from, then formulae like $\Box_1\Box_2 p \Rightarrow \Box_2\Box_1 p$[2] are not in the language $\mathcal{L}_{\mathrm{L}_1(\mathrm{L}_2)}$, nor in $\mathcal{L}_{\mathrm{L}_2(\mathrm{L}_1)}$. In order to allow for this sort of formulae to be in the language, the two logics cannot play different roles and a more "neutral" approach is required.

6.2.2.2 Fusion of modal logics

[FG92] is very liberal about the requirements on the logics to be combined, but for doing so it must pay a heavy price in terms of expressiveness of the language of the composition. [KW91] takes a radically different perspective: the logics are restricted, but it aims to achieve the richest expressivity. Specifically, Kracht and Wolter, extending some results already obtained by Thomason in [Tho80b] and Fine and Schurz in [FS96][3], investigate the transfer properties for a combination between two mono-modal normal logics into a particular normal bi-modal logic.

Consider a language $\mathcal{L}$ built from a set of propositional variables P and the connectives $\neg, \wedge, \Box_1, \Box_2$. Recall from Definition 1.8 that a normal bi-modal logic L is a set of formulae closed under the axiom K and the inference rules of necessitation (for both connectives) and uniform substitution. Consider now two mono-modal logics L_1 and L_2 defined from the languages $\mathcal{L}_{\mathrm{L}_1}$ and $\mathcal{L}_{\mathrm{L}_2}$ respectively and assume them to be built from the set of connectives $\neg, \wedge$ and the mono-modal operators $\Box_1$ and $\Box_2$ respectively. The projections of a normal bimodal logic L onto its two constituent languages are normal mono-modal logics; that is $\mathrm{L} \cap \mathcal{L}_{\mathrm{L}_1} = \mathrm{L}_1$ and $\mathrm{L} \cap \mathcal{L}_{\mathrm{L}_2} = \mathrm{L}_2$. In the following we use $\mathrm{L}_{\mathcal{L}_{\mathrm{L}_1}}$ as an abbreviation for $\mathrm{L} \cap \mathcal{L}_{\mathrm{L}_1}$.

What is particularly interesting, though, is the converse of this construction, that is to start with two mono-modal normal logics L_1 and L_2 and investigate the properties of the bi-modal logics containing the two. There are many bi-modal logics that we can consider, Kracht and Wolter investigate the transfer properties in respect to a special one that they call *fusion*.

Definition 6.6. *If* $\mathrm{L}_1, \mathrm{L}_2$ *are normal modal logics, the* fusion $\mathrm{L}_1 \oplus \mathrm{L}_2$ *is the least normal bimodal logic containing the two.*

If a logic L can be expressed as a fusion then L is also called *independently axiomatisable* in consequence of a result that we report shortly. The fusion operator is commutative up to isomorphisms; that is $(\mathrm{L}_1 \oplus \mathrm{L}_2) \cong (\mathrm{L}_2 \oplus \mathrm{L}_1)$.

[2] And indeed many of the axiom schemas analysed in Chapter 4.

[3] Although published later, this paper was actually written before [KW91].

Kracht and Wolter do not use the standard possible worlds semantics to define the interpretation of the logics. Instead, they use Boolean algebras [BS84] as in the tradition of mathematical logic. For the purposes of this review it is not important to discuss these here.

This framework allows the authors to import a theorem which was already proved in [FS96] and [Tho80b].

Theorem 6.7. *If* $\mathrm{L}_1, \mathrm{L}_2$ *are mono-modal logics,* $(\mathrm{L}_1 \oplus \mathrm{L}_2)_{\mathcal{L}_{\mathrm{L}_1}} = \mathrm{L}_1$ *except precisely when the logic* L_2 *is inconsistent and the logic* L_1 *is not.*

This gives the flavour of the fusion: the restriction to its own language of the fusion of a mono-modal logic with another is the logic itself. In other words, the fusion does not add any new theorem on the corresponding dimensions. Formally we say that a fusion $\mathrm{L}_1 \oplus \mathrm{L}_2$ of L_1 with L_2 is a *conservative extension* of both. Still, it should be noted that the fusion of two logics $\mathrm{L}_1, \mathrm{L}_2$ is not equal to the union of L_1 with L_2. This is because $\mathrm{L} = (\mathrm{L}_1 \oplus \mathrm{L}_2)$ is closed under substitution. For example: if $\mathrm{L}_1, \mathrm{L}_2$ are normal mono-modal logics with operators $\Box_1, \Box_2$, then $\Box_2(\Box_1 p \Rightarrow q) \Rightarrow (\Box_2\Box_1 p) \Rightarrow (\Box_2 q)$ belongs to $\mathrm{L}_1 \oplus \mathrm{L}_2$.

Kracht and Wolter prove a number of properties for the fusion. The most important are the following.

Theorem 6.8 (Properties of fusion). *Let* L_1 L_2 *be two consistent normal modal logics. Then the following holds.*

1. *The logic* $\mathrm{L}_1 \oplus \mathrm{L}_2$ *is finitely axiomatisable if and only if the logics* L_1 *and* L_2 *are.*
2. *The logic* $\mathrm{L}_1 \oplus \mathrm{L}_2$ *is complete if and only if the logics* L_1 *and* L_2 *are.*
3. *The logic* $\mathrm{L}_1 \oplus \mathrm{L}_2$ *is compact if and only if the logics* L_1 *and* L_2 *are.*
4. *If the logics* $\mathrm{L}_1, \mathrm{L}_2$ *are complete, then the logic* $\mathrm{L}_1 \oplus \mathrm{L}_2$ *is decidable if and only if the logics* L_1 *and* L_2 *are.*
5. *If the logics* $\mathrm{L}_1, \mathrm{L}_2$ *are complete, then the logic* $\mathrm{L}_1 \oplus \mathrm{L}_2$ *is Halldén-complete*[4] *if and only if the logics* L_1 *and* L_2 *are.*

The results in [KW91] are very strong and the process of fusion can be extended to the fusion of normal n-ary modal logics.

Notwithstanding this, the fusion technique works only for independently axiomatisable modal logics and it does not address more general cases. As an example of what is not covered, Kracht and Wolter discuss the minimal tense extension of a logic. Given a logic L, whose modal language is built from the modality $\Box_1$, the minimal tense extension of L is defined as: $\mathrm{T_L} = \mathrm{L} \oplus \mathrm{K}(p \Rightarrow \Box_1\Diamond_2 p, p \Rightarrow \Box_2\Diamond_1 p)$. In other words $\mathrm{T_L}$ is the fusion of a normal logic $\mathrm{L}_{\Box_1}$ with the system $\mathrm{K}_{\Box_2}$ *plus* the two axioms $\{p \Rightarrow \Box_1\Diamond_2 p\}$ and $\{p \Rightarrow \Box_2\Diamond_1 p\}$. Kracht and Wolter report that their technique did not seem to help in order to prove the completeness of $\mathrm{T_L}$, given the completeness of L, although this is a generally considered an expected result.

Apparently, the fusion technique provides a more expressive language than the embedding, but it does not give any result if we want to express interactions between the two components.

[4] A logic L is said to be Halldén-complete if $\phi \vee \psi \in L$ and $var(\phi) \cap var(\psi) = \emptyset$ implies $\phi \in L$ or $\psi \in L$, where $var(\phi)$ and $var(\psi)$ are the sets of propositional variables that appear in ϕ and ψ respectively. The notion of Halldén-completeness is related to the concept of "relevance".

6.2.2.3 Independent combination of linear temporal logics

In this section we discuss a technique called "independent combination", which has been developed on the purpose of the combination of linear temporal logics.

The independent combination was developed by Finger and Gabbay [FG96a] to overcome the expressivity limitations of the embedding (Section 6.2.2.1) and it is similar in definition and results to the fusion of Kracht and Wolter.

Although it is designed especially for temporal combinations, it is worth discussing because it shows an interesting relation between embedding and fusion. Given the similarities that will arise between fusion and independent combination we will denote by $\mathrm{T}_1 \oplus \mathrm{T}_2$ the independent combination of T_1 and T_2.

In the following the logics $\mathrm{T}_1, \mathrm{T}_2$ are modal temporal logics for the operators of since and until defined on their respective classes of linear frames (see Section 6.2.2.1 and Section 1.3.1.4).

6.2.2.3.1 Syntax. The fully combined language $\mathcal{L}_{\mathrm{T}_1 \oplus \mathrm{T}_2}$ of two logics T_1 and T_2 is obtained by making the union of the formation rules of the two languages $\mathcal{L}_{\mathrm{T}_1}, \mathcal{L}_{\mathrm{T}_2}$ over a single set of propositional variables P.

Definition 6.9 (Syntax of fully combined temporal language).

- *if p is a propositional variable, then p is in* $\mathcal{L}_{\mathrm{T}_1 \oplus \mathrm{T}_2}$*,*
- *if* ϕ, ψ *are in* $\mathcal{L}_{\mathrm{T}_1 \oplus \mathrm{T}_2}$*, then* $\neg\phi, (\phi \wedge \psi)$ *are in* $\mathcal{L}_{\mathrm{T}_1 \oplus \mathrm{T}_2}$*,*
- *if* ϕ, ψ *are in* $\mathcal{L}_{\mathrm{T}_1 \oplus \mathrm{T}_2}$*, then* $(\phi\, \mathcal{S}_1 \psi), (\phi \mathcal{U}_1 \psi)$ *are in* $\mathcal{L}_{\mathrm{T}_1 \oplus \mathrm{T}_2}$*,*
- *if* ϕ, ψ *are in* $\mathcal{L}_{\mathrm{T}_1 \oplus \mathrm{T}_2}$*, then* $(\phi\, \mathcal{S}_2 \psi), (\phi \mathcal{U}_2 \psi)$ *are in* $\mathcal{L}_{\mathrm{T}_1 \oplus \mathrm{T}_2}$*.*

This language is indeed very similar to the one of fusion, presented in [KW91]. The only difference is that [FG96a] discusses the case of *binary* modal operators, while [KW91] is restricted to more traditional unary operators.

6.2.2.3.2 Semantics. Assume $\mathrm{T}_1, \mathrm{T}_2$ have associated two classes of temporal models $\mathcal{M}_{\mathrm{T}_1}$ and $\mathcal{M}_{\mathrm{T}_2}$. The class $\mathcal{M}_{\mathrm{T}_1}$ is made by models $M(T, \prec_1, \pi_1)$, where T is a set of time points, the relation $\prec_1$ is a linear order on T^2 and π_1 is an interpretation for the atoms P. The class $\mathcal{M}_{\mathrm{T}_2}$ is built similarly.

A model for $\mathrm{T}_1 \oplus \mathrm{T}_2$ is a tuple $M_{\mathrm{T}_1 \oplus \mathrm{T}_2} = (T, \prec_1, \prec_2, \pi)$ where $(T, \prec_1)$ is a linear frame for T_1, $(T, \prec_2)$ is a linear frame for T_2, and $\pi : T \to 2^P$ is an interpretation function for the atoms. In the following we use the symbol M to refer to models of the fusion $M_{\mathrm{T}_1 \oplus \mathrm{T}_2}$. By $\mathcal{M}_{\mathrm{T}_1 \oplus \mathrm{T}_2}$ we will denote the class of models of the independently combined logics T_1 and T_2.

The definition of satisfiability for formulae of $\mathcal{L}_{\mathrm{T}_1 \oplus \mathrm{T}_2}$ can then be defined by merging the definitions for the two temporal logics:

Definition 6.10 (Satisfiability for $\mathrm{T}_1 \oplus \mathrm{T}_2$).

$M \models_t p$*, where* $p \in P$	*if*	$p \in P$.
$M \models_t \neg\phi$	*if*	*it is not the case that* $M \models_t \phi$.
$M \models_t (\phi \wedge \psi)$	*if*	$M \models_t \phi$ *and* $M \models_t \psi$.

$M \models_t (\phi\,\mathcal{S}_1\psi)$	*if*	*there exists a point* s*,* $s \prec_1 t$ *and* $M \models_s \psi$ *and* *for all points* $u \in T$*, if* $s \prec_1 u \prec_1 t$ *then* $M \models_u \phi$*.*
$M \models_t (\phi\,\mathcal{U}_1\psi)$	*if*	*there exists a point* s*,* $t \prec_1 s$ *and* $M \models_s \psi$ *and* *for all points* $u \in T$*, if* $\mathrm{T}_1 \prec_1 u \prec_1 s$ *then* $M \models_u \phi$*.*
$M \models_t (\phi\,\mathcal{S}_2\psi)$	*if*	*there exists a point* s*,* $s \prec_2 t$ *and* $M \models_s \psi$ *and* *for all points* $u \in T$*, if* $s \prec_2 u \prec_2 t$ *then* $M \models_u \phi$*.*
$M \models_t (\phi\,\mathcal{U}_2\psi)$	*if*	*there exists a point* s*,* $t \prec_2 s$ *and* $M \models_s \psi$ *and* *for all points* $u \in T$*, if* $t \prec_2 u \prec_2 s$ *then* $M \models_u \phi$*.*

6.2.2.3.3 Proof Theory. The proof theory of the independent combination is defined in terms of the union of the proof theories of the components. Assume once again that Hilbert-style proof theories (Section 1.2.3) for $\mathrm{T}_1, \mathrm{T}_2$ are defined. These are pairs (A, I) of axiom schemas and inference rules.

Definition 6.11. *Let* $(A_{\mathrm{T}_1}, I_{\mathrm{T}_1})$ *and* $(A_{\mathrm{T}_2}, I_{\mathrm{T}_2})$ *be the axiomatisation of* T_1 *and* T_2 *respectively and let* A_{T_1} *and* A_{T_2} *be disjoints. The axiomatisation of* $\mathrm{T}_1 \oplus \mathrm{T}_2$ *is then defined as* $(A_{\mathrm{T}_1} \cup A_{\mathrm{T}_2}, I_{\mathrm{T}_1} \cup I_{\mathrm{T}_2})$*.*

Finger and Gabbay claim that Thomason's theorem[5] that we presented for fusion in Theorem 6.7 applies for independent combination as well. The difference is that [Tho80b] and [KW91] prove the theorem for normal unary modal operators, while Finger and Gabbay are working with the binary operators of since and until.

Similarly to Kracht and Wolter, Finger and Gabbay prove a number of transfer properties from the mono-temporal logics into the combination. What is interesting to note is that the proofs of the transfer properties use the embedding technique. In fact, consider the following.

Given a formula ϕ in $\mathcal{L}_{\mathrm{T}_1 \oplus \mathrm{T}_2}$ it is possible to define a *degree of alteration*, *dg*(ϕ), which is the minimum number of alternate temporalisation between T_1 and T_2 which is necessary to make in order to see ϕ as part of an iteratively temporalised language. The degree of alteration can be defined inductively on the structure of the formula; for the purpose of this review we do not give the details.

Lemma 6.12. *Let* ϕ *be in* $\mathcal{L}_{\mathrm{T}_1 \oplus \mathrm{T}_2}$ *with* $dg(\phi) = n$*. The formula* ϕ *is a theorem of* $\mathrm{T}_1 \oplus \mathrm{T}_2$ *if and only if* ϕ *is a theorem of the logic obtained by* n *alternate temporalisation of* T_1 *and* T_2*.*

The lemma is indeed revealing. It is not only the case that by embedding alternately a logic inside another a finite amount of times it is possible to express any formula of the language of the independent combination (or fusion), but any theorem of the independent combination is a theorem of an iterated temporalisation and vice versa. This means that the independent combination of two logics can be considered the infinite union of the alternate temporalisations of the two logics. By using this result, Finger and Gabbay can prove the transfer of soundness, completeness and decidability more simply than Kracht and Wolter.

Theorem 6.13 (Properties of the independent combination).

1. *If* T_1 *and* T_2 *are sound and complete logics with respect to* $\mathcal{M}_{\mathrm{T}_1}$ *and* $\mathcal{M}_{\mathrm{T}_2}$ *respectively, then* $\mathrm{T}_1 \oplus \mathrm{T}_2$ *is sound and complete with respect to* $\mathcal{M}_{\mathrm{T}_1 \oplus \mathrm{T}_2}$*.*

[5] If a formula is valid in T_1 then it is valid in every independent combination of T_1.

2. *If* T_1 *and* T_2 *are decidable and complete with respect to* $\mathcal{M}_{\mathrm{T}_1}$ *and* $\mathcal{M}_{\mathrm{T}_2}$ *respectively, then* $\mathrm{T}_1 \oplus \mathrm{T}_2$ *is decidable.*

Finger and Gabbay do not discuss the transfer of other properties. Notwithstanding this, their analysis covers the basic properties and builds interestingly upon their previously defined methodology.

Of course, the drawbacks of the fusion technique are all present in the independent combination. In particular, the lack of interplay between the different dimensions of time is reflected in the absence of interaction theorems in the logic. Finger and Gabbay also suggest that a broadly two dimensional logic, should have a fully two dimensional definition of satisfiability. This means that formulae should be evaluated at pairs of time points instead of single points. While this is theoretically very interesting, we cannot see any possible use of this in agents theory and therefore we do not report here this aspect of their work here (see [FG96a] for details).

6.2.2.4 Fibring and dovetailing

The most general technique for combining logics is potentially the one conceived by Dov Gabbay in the 1990s and generally referred to with the term "fibring". In fact, [Gab96b], [Gab96a] and the forthcoming [Gab99] present a general methodology for combining logics irrespective of their presentation and type.

The methodology allows the user to combine (henceforth *fibre*) the semantics of the two systems and *weave* the two proof-theories into a combined logic that preserves the basic properties of the components. In order to do so, the two logics should be fully presentable, through their syntax, semantics and proof-theory. If that is not the case, the methodology is still usable. The strategy is then to extract the consequence relations from the two logics, however they are presented, give a basic point relational semantics to the two logics (through a method conceived by Gabbay) and finally fibre the relational semantics into the combined logic.

The fibring methodology is sufficiently general to be applied to any combination, but this generality renders any attempt of applying it not a straightforward one. For the aims of this section we will not discuss the general methodology, but we will focus on how to fibre two modal logics.

Given two modal logics, L_1 and L_2 and their corresponding languages $\mathcal{L}_{\mathrm{L}_1}, \mathcal{L}_{\mathrm{L}_2}$, classes of models $\mathcal{M}_{\mathrm{L}_1}, \mathcal{M}_{\mathrm{L}_2}$ and satisfaction relations $\models_1, \models_2$, we would like to define a logic that has the expressivity of both. As we have seen above the problem is not to define the syntax of of the combination: for example we can consider the formation rules of both languages $\mathcal{L}_{\mathrm{L}_1}, \mathcal{L}_{\mathrm{L}_2}$. The real problem is to define the models and the satisfaction relation. The basic idea of fibring is to perform a model construction while calculating the interpretation a formula. So, consider a formula as being composed by a preamble operator (inherited from one of the two logics, say $\Box_1$ in this case) and the rest. We can inductively interpret $\Box_1$ in one of the models for L_1 and then interpret the rest. If the rest has $\Box_1$ as preamble (or there are no more operators at that level of parsing), we can interpret it on that model, otherwise in order to interpret the other modal operator $\Box_2$, we have to use a model for L_2. In order to do so, we are to link (*to fibre*), via a *fibring* function, the model for L_1 with a model for L_2 and in the process of doing so, we build a *fibred* model of the combination. In one line, the interpretation of a formula ϕ of the combined language in the fibred model at a state w can

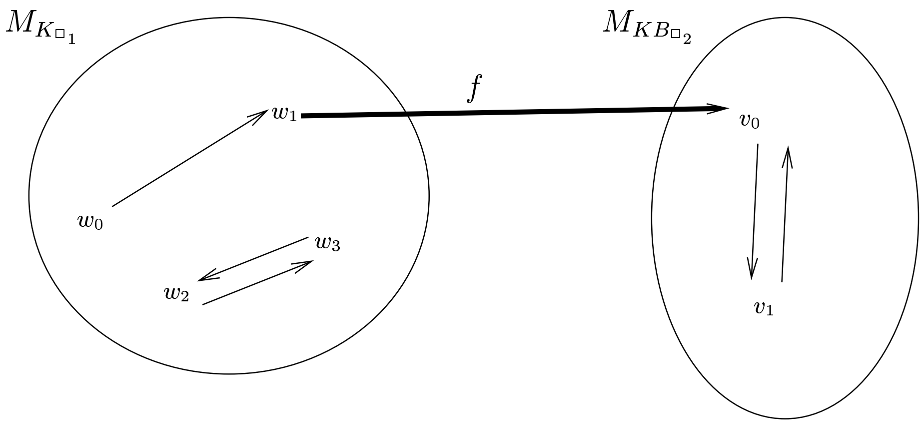

Figure 6.1: An example of fibring.

be summarised as:

$$\models_w \phi \text{ if and only if } \models^*_{f(w)} \phi,$$

where f is a fibring function that maps a world to a model *suitable for interpreting* ϕ and $\models_*$ is the corresponding satisfaction relation (either $\models_1$ for L_1 or $\models_2$ for L_2). We present an example to convey the main thrust of the idea.

Let us consider two modal logics, $K_{\Box_1}$ and $KB_{\Box_2}$. Suppose we are to interpret the formula $\Box_1\Diamond_2 p$ on a world of the fibred semantics, say w_0. The situation is represented in Figure 6.1. We start by evaluating the preamble $\Box_1$ of $\Box_1\Diamond_2 p$, $\Box_1$, at w_0. According to the standard definitions, we have to check whether $\Diamond_2 p$ is true at every w_1 accessible from w_0. Intuitively, at w_1 we cannot interpret the operator $\Diamond_2$, because we are in a model of $K_{\Box_1}$, not of $KB_{\Box_2}$. This is when the fibring function f comes to help. The function f at w_1 points to v_0, a world in a model suitable to interpret formulae with $\Box_2$ as preamble. Now, all we have to check is whether $\Diamond_2 p$, is true at v_0 in this last model, and this can be done in the usual way.

In order to present the general case of the fibring of n modal logics, let $L_i, i \in I$ mono-modal logics. L_i need not be normal. In the following we denote by $\Box_i$ the modal operator of the logic L_i.

The syntax of the fibred logic L_i^F is defined as follows[6].

Definition 6.14 (Syntax of the fibred modal logic). *Given a family* $L_i, i \in I$ *of modal logics defined from the same set P of propositional atoms, the language* $\mathcal{L}_{L_i^F}$ *of the logic* L_i^F *is defined by taking the union of the formation rules.*

Since we will not have the chance to use the semantics of fibring I do not feel the need of presenting here with its complex details. I therefore refer the interested reader to [Gab96b, Gab99] for the semantic definitions that define the class $\mathcal{M}_{L_i^F}$ of fibred models.

Axiomatically the fibred logic L_I^F is defined as follows.

Definition 6.15. *Let* $L_i, i \in I$ *be modal logics; let* L_I^F *be defined as follows:*

[6]This is actually a simplification on the definition given in [Gab99, GG98] where different sets of atoms and connectives are assumed for the logics. Since this is particularly compelling for MAS theories it makes sense to simplify the definition.

1. $\mathrm{L}_i \subseteq \mathrm{L}_I^F$ *for any* $i \in I$.
2. *(Modal fibring rule):*
   If $i \neq j$, *and* $\chi = (\bigwedge_{k=1}^{n} \Box_i \phi_k \Rightarrow \bigvee_{k=1}^{m} \Box_i \psi_k) \in \mathrm{L}_I^F$, *then for all* d, $\Box_j^d \chi \in \mathrm{L}_I^F$;
3. L_I^F *is the smallest set closed under item 1 of this list, Modus Ponens, uniform substitution, and modal fibring rule.*

Gabbay proves the following:

Theorem 6.16 (Properties of the fibred modal logic).

1. L_I^F *is the set of valid formulae of every model in* $\mathcal{M}_{\mathrm{L}_i^F}$.
2. *If all the* $\mathrm{L}_i, i \in I$, *satisfy finite model property, then so does* L_i^F.
3. *If all the* L_i *are finitely axiomatisable, so is* L_i^F.

A special case of fibring, called dovetailing, arises when the fibring function is set to map the actual world of the target model. Dovetailing has the same syntax, a much easier semantics (that for the same reason as before we do not report here) and it generates a different logic L_I^D, defined on class of dovetailed model $\mathcal{M}_{\mathrm{L}_I^D}$.

Definition 6.17. *Let* $\mathrm{L}_i, i \in I$ *be modal logics; let* L_I^D *be defined as follows:*

1. $\mathrm{L}_i \subseteq \mathrm{L}_I^D$ *for any* i *in* I.
2. *(Modal dovetailing rule):*
   If $i \neq j$, *and* $\chi = (\bigwedge_{k=1}^{n} \Box_i \phi_k \wedge \bigwedge_{k=1}^{m} \Diamond_i \neg \psi_k \Rightarrow \bigvee_{k=1}^{m} q_k) \in \mathrm{L}_I^D$, *then for all* d, $\Box_j^d \chi \in \mathrm{L}_I^D$, *where* q_k *are atoms (or their negations) and* $q_1, \ldots, q_r$ *list all the atoms (or their negations) in any* ϕ_k, ψ_k, $k = 1, \ldots$
3. L_I^D *is the smallest set closed under 1, Modus Ponens, Substitution, and the Modal Dovetailing Rule.*

We have the following properties.

Theorem 6.18 (Properties of the dovetailed modal logic).

1. L_I^D *is the set of valid formulae of every model in* $\mathcal{M}_{\mathrm{L}_I^D}$.
2. *If each of the* $\mathrm{L}_i, i \in I$ *include* K *and can be formulated by an Hilbert style system with necessitation rule, then* L_I^D *can be axiomatised by taking the union of the axiomatisations.*
3. *If all the* $\mathrm{L}_i, i \in I$ *satisfy finite model property, then so does* L_I^D.

So, from a proof-theoretical point of view, dovetailing is the special case of fibring that generates simply the fusion of the logics. In some cases fibring and dovetailing produce the same result. The following addresses one of such cases:

Theorem 6.19. *If each of the* L_i, i *in* I *admits necessitation and satisfies the disjunction property*[7], *then* $\mathrm{L}_I^F = \mathrm{L}_I^D$.

[7] If $\vdash \Box\phi \Rightarrow (\Box\psi \vee \Box\chi)$, then $\vdash \Box\phi \Rightarrow \Box\psi$ or $\vdash \Box\phi \Rightarrow \Box\chi$.

Gabbay shows the extent to which some nonstandard logic systems can be analysed either as fibrings or dovetailings of two or more logics. For some systems ([Wij90], [Fit48], [Ono77], [Ewa86]) this analysis is successful, for some others (for example [FS77]) it is not.

Adaptations of the fibring technique can apparently be applied to many other cases we did not discuss here. In particular, it is possible to fibre a logic with a fuzzy logic, fibre a first-order logic with itself or with a non-standard logic, combine a logic with its meta-level (for example by bringing inside the logic its own consequence relation), and many more. Self fibring has also been used very recently ([GN97]) to model context.

But the big question which still remains without an answer is to what extent fibring, when applied to modal logics, extends the results we presented for fusion and independent combination. In other words, is there a way to specify interaction between the logics in the fibred logic, or is it just another way to define fusion? How do interactions between the logics translate into restrictions of the fibring function? To my knowledge these hard questions have not been answered yet. What we can notice is that it is sometimes possible to recognise some existing combined systems as fibrings or dovetailings but difficulties arise when the combination is not a simple fusion, but an interaction between the components is present.

Perhaps the fibring technique is still too young to give all the answers we need and by investigating some class of fibring functions it will be possible to give a basic class of interactions and prove properties about these. This is very much part of the ongoing research carried out by Gabbay and colleagues.

6.2.3 Negative results

Apart from the technical difficulties of the previous section, the story we have presented so far of combining logics is one of positive results. After all, completeness and decidability do transfer for all the cases we have examined in Section 6.2.2. Still, the feeling that a casual listener would get from overhearing a conversation on the subject would possibly be not quite the same. What follow are some of the reasons for that.

Since the discovery of incomplete logics, advanced research on modal logic has also focused on negative results. This has proven to be as useful as research on positive results and it has helped understanding notions such as completeness, fmp, decidability, etc. in more detail. It was therefore natural that when the problem of identifying classes of combinations for which transfer of properties occurs came about, many logicians tried to find examples for which this does *not* happen.

In [dRB96] at least two strong negative results are present and others have been published even more recently. It would not be appropriate to discuss here all this material and I only report three representative results.

6.2.3.1 Incomplete minimal tense extensions of K4

A normal bimodal logic with the operators $\Box^+, \Box^-$ is a tense logic if it contains the axioms

$$\{\phi \Rightarrow \Box^+\Diamond^-\phi\}, \{\phi \Rightarrow \Box^-\Diamond^+\phi\}.$$

Consider a modal logic Γ containing K4 and its minimal tense extension Γ^+, which is the smallest tense logic containing Γ. [Wol96] shows that completeness and the fmp do not transfer in general from Γ to Γ^+.

Since minimal tense extensions can be intuitively seen as one of the less demanding ways of temporalising a logic and since K4 is contained in most of the logics for representing mental states, the paper shows that even some natural combinations can suffer from undecidability and incompleteness.

6.2.3.2 Undecidability of modal logics with reflexive transitive operators

We now report on an important negative results published in [Hem96].

Consider a complete and decidable modal logic Γ with a modal operator $\Box$, whose interpretation is defined on a relation R in the usual Kripke semantics. Consider now Γ enriched by the reflexive transitive modality $\boxast$, which is defined in terms of the reflexive transitive closure of the relation associated with $\Box$ on the Kripke models.

From examples in the literature ([FL79, Pra79, EH85, BAHP82]) one could think that decidability of a logic Γ should not be lost by adding $\boxast$ (in fact this has also been conjectured in [GP92]). Hemaspaandra shows that this is not the case by proving that there exists a unimodal decidable, canonical logic for which adding the reflexive transitive modality $\boxast$ causes undecidability.

$\boxast$–modalities are used in MAS theories to express, for example, common knowledge and common belief (Section 1.3.2). We know from Section 1.3.2.2 that epistemic logics like $S5_n$ and doxastic logics like $KD45_n$ are not affected by this problem. Nonetheless this could indeed fail in any combination of knowledge with other mental states.

6.2.3.3 Failure of interpolation in combined modal logics

This result is by Marx and Areces and it appears in [MA99]. Recall that a logic L has interpolation if whenever $\phi_0 \wedge \phi_1 \models_{\mathrm{L}} \psi$, then there exists a formula θ such that $\phi_0 \models_{\mathrm{L}} \theta$ and $\phi_1 \wedge \theta \models_{\mathrm{L}} \psi$ and $P(\theta) \subseteq P(\phi_0) \cap (P(\psi) \cup P(\phi_1))$, where $P(\phi)$ is the set of atomic symbols occurring in ϕ, and $\models_{\mathrm{L}}$ is a global (on models) consequence relation for the class of models for L[8].

As mentioned in [MA99] the importance of interpolation in normal modal logics is connected to the fact that it is equivalent to the property that whenever two theories T_1, T_2 both have a model and one does not contradict the other on the shared language, then the union of the two has also a model. This property has been used in specification of software (see [MS85] cited in [MA99]) and it is therefore one of the properties that we would like to transfer into the combinations. Note that interpolation holds for all the logics of MAS theories discussed in Section 1.3.2.

[MA99] shows that interpolation does not in general transfer neither into union (Section 6.2.2.2) of modal logics nor in the product (see [GS98]). It is been discussed ([MS85]) that when a logic enjoys the interpolation property the efficiency of an automatic theorem proving technique can be dramatically increased. Given the importance of automatic theorem proving to the verification of agent specifications, this could be a problem.

[8]There are actually a few variants of the definition of interpolation; for this section the one presented above is adequate. See the cited paper for more details.

6.3 MAS theories as combined logics

In Section 6.2.2 we have examined some of the most successful techniques to combine modal logics. In this section we investigate the extent to which the MAS theories we discussed in Section 1.3.2 can be seen as particular examples of logic combination.

6.3.1 Knowledge and other single mental state theories

Let us consider once again the logic $S5_n$ that we have been using for most of this thesis as a base for an epistemic theory[9]. Recall from Section 1.3.2 that $S5_n$ was motivated as the indexed extension of the logic S5 which was proposed long ago to model knowledge.

Alternately, let us consider n-copies of the the mono-modal logic S5, one for every agent of the group: $S5_{K_1}, \ldots, S5_{K_n}$. Having done so, consider now the fusion (as defined in Section 6.2.2.2) of these. We obtain the logic $S5_{K_1} \otimes \cdots \otimes S5_{K_n}$. Note that the fusion is well-defined because all the logics involved are normal logics. We can then apply the results of Theorem 6.2.2.2 and Theorem 6.2.2.3. We have that the fusion can be axiomatised by taking the union of the axiomatisations of the n copies of S5, one for every modal box. We now notice that the union of the n axiomatisations gives exactly the logic $S5_n$. In other words $S5_n$ can effectively be seen as the fusion (precisely as defined in Section 6.2.2.2) of n S5-components, one for every agent of the model:

$$S5_n \equiv S5_{K_1} \otimes \cdots \otimes S5_{K_n}.$$

Indeed, by Theorem 6.13, it follows that the logic $S5_n$ is complete and decidable, which is something that had already been proved by more standard means.

Consider now the operators of distributed knowledge and common knowledge of Section 1.3.2.2. The semantics of these is defined by using the accessibility relations of the private knowledge operators and so they cannot be seen as independent as in the above case of $S5_n$. Moreover, interaction axioms between common knowledge and private knowledge are also present and so we are unable to recognise the system $S5_n^{E,C}$ as either a fusion or as an independent combination of logics. That is certainly not the result of an embedding because of the syntactic limitations of this technique. It might be the result of a fibring or dovetailing but that is not clear either.

So, we conclude that the logic $S5_n$ is a fusion or independent combination of n copies of S5, but it is far from being obvious whether any of its extensions devised to deal with group properties can be thought of as the result of a logic combination.

6.3.2 BDI logics

In Section 1.3.2.3.3 we briefly discussed a family of logics commonly known as BDI logics. There we introduced the very basic BDI logic. This was a multi-modal logic for belief, desire and intention skimmed of the temporal dimension (usually CTL*).

In Section 1.3.2.3.3 we discussed that the basic BDI logic is defined from three logics: $KD45_n$ for belief, and KD_n for desires and intentions. In top of this, the two interaction axioms

$$\{I_i\phi \Rightarrow D_i\phi\}, \{D_i\phi \Rightarrow B_i\phi\}$$

[9]It should be noted that what follows actually applies equally well to any logic modelling a single mental state of a group of agents, e.g. all the logics discussed in Section 1.3.

that regulate the relation between desires, intentions and beliefs of every agent are imposed.

Following the observation regarding $S5_n$ of the previous section we can now see the three different component logics as being fusions of n copies for each logic. The basic BDI logic L can then be seen as the fusion of these plus the two interaction axioms. In symbols:

$$\mathrm{L} \equiv (\oplus_{i=1}^{n}\mathrm{KD45}_{B_i}) \oplus (\oplus_{i=1}^{n}\mathrm{KD}_{D_i}) \oplus (\oplus_{i=1}^{n}\mathrm{KD}_{I_i}) + \{I_i\phi \Rightarrow D_i\phi\} + \{D_i\phi \Rightarrow B_i\phi\}.$$

We can now apply the results of Section 6.2.2.2 or Section 6.2.2.3 to prove that the logic

$$\mathrm{L}' \equiv (\oplus_{i=1}^{n}\mathrm{KD45}_{B_i}) \oplus (\oplus_{i=1}^{n}\mathrm{KD}_{D_i}) \oplus (\oplus_{i=1}^{n}\mathrm{KD}_{I_i})$$

is sound, complete and decidable, but note that these results do not transfer to the logic L when we we add the required interaction axioms.

So, combining logics for this case is of limited use. We have the transfer of properties for the three components fused together but we do not have a result that we can apply in the case of interaction axioms being present in the combination.

6.3.3 Knowledge and belief

In Section 1.3.2.3.1 we discussed the system introduced by Kraus and Lehmann to model knowledge and belief in a community of ideal agents. As noted there this system includes operators for common knowledge and common belief. By the same considerations of Section 6.3.2, we are unable to identify this system as the result of a logic combination.

In this case, differently from the BDI case, it is not even possible to recognise in a straightforward way even *a fragment* of the system, private belief and knowledge for example, as a fusion plus interaction axioms. To see this, note that in the axiomatisation of [KL88] the inference rule of necessitation for belief is not present. Still, the rule is actually sound as we can apply necessitation for common knowledge that via Modus Ponens implies necessitation for common belief, from which we can derive "everybody believes" which in turn implies the belief of any agent. Therefore by restricting the axiomatisation to the fragment defined only on the bi-modal family of operators B_i and K_i, we would not be getting all the theorems that we have in the combined system, restricted to the language of B_i and K_i.

Our conclusion must then be that this system, *as it was presented* by Kraus and Lehmann, is so interconnected that it cannot even be easily recognised as a fusion of mono-modal logics plus extra interaction axioms. It may be that the logic L defined on the language of private knowledge and belief as

$$\mathrm{L} \equiv \mathrm{KD45}_{B_1} \oplus \cdots \oplus \mathrm{KD45}_{B_n} \oplus \mathrm{S5}_{K_1} \oplus \cdots \oplus \mathrm{S5}_{K_n} + \{K_i\phi \Rightarrow B_i\phi\} + \{B_i\phi \Rightarrow K_iB_i\phi\}$$

is complete with respect to the class of models considered in [KL88] but in order to prove that, as things stand, it looks easier to refer to traditional methods.

6.3.4 Knowledge and time

So far we have had limited success in applying the techniques of Section 6.2.2 to our MAS theories. As a last example, we analyse the system for knowledge and time the we explored in Section 1.3.2.3.2.

Although also in this case the logic is clearly a combination between the epistemic $S5_n$ and a basic temporal logic, this time it is even less obvious than before whether the system can actually be seen as an application of combining logics. Indeed, apart from the above mentioned lack of formal tools to handle interaction axioms (required in this case to model perfect recall, no learning, etc.) we have two extra technical problems here.

Firstly, the semantics in use in Section 1.3.2.3.2 (based on interpreted systems) is different from the one used in the analysis of [FG96a] and [KW91] (based on Kripke models). In Chapter 2 we studied a technique for mapping one into another, but that was limited to the case of hypercubes. So it is not clear how to proceed in this case.

Secondly, the logic $S5_n^U$ cannot straightforwardly be seen as a normal modal logic, simply because the temporal part with its binary operators of since and until does not conform to the usual definition of normality.

To try to overcome these difficulties two ways may be promising.

The first is to develop a technique (perhaps similar to the one presented in Section 2.4) for relating any class of interpreted systems to Kripke models. This is necessary in order to translate the classes of agents defined with interpreted systems into the semantics used by fusion or independent combination. Once this tool is available and normality can be defined for binary operators in a way such that the results of Section 6.2.2.2 hold we still face the problem of the interaction between the temporal and epistemic dimension. This interaction would this time be represented by semantic conditions between the temporal and epistemic relations on the Kripke models that result from the interpreted systems (see page 34).

Alternately we can think of using the technique of fibring. This may be more appropriate in this case because it does not require the logics to be normal. The procedure would be the same as before: translate the semantics of interpreted systems into Kripke models and apply the fibring technique. Still, as we suggested in Section 6.2.2.4, fibring does not seem to be easily applicable when we need to impose interaction axioms. In principle it should even be possible to fibre directly the components on a semantics based on interpreted system, although it should be said that adapting fibring for a new semantics is itself a very challenging task.

So, quite predictably the existing combining techniques cannot be applied straightforwardly in this case either.

6.4 Conclusion

In this chapter we reviewed some promising tools for combining logics with respect to their applicability to agent theories. We argued that, in principle, they are very relevant for the needs of MAS theorists as they aim at tackling the same technical problems but in a more general context.

The embedding technique gives a way of temporalising any logic for agents. This technique has been used in the literature to model temporal evolution of epistemic states in a MAS (see [Pre96] for example). The authors claim that embedding is equally applicable in other cases as well, like an external epistemic logic referring to a doxastic setting; I am not aware of any work that follows this line. With respect to agent theories, the main drawback of embedding is clearly that the language of the combination is strictly limited and therefore not suitable for expressing temporal-epistemic constraints of a MAS, such as perfect recall. At present is hard to see whether the result relating independent combination (or fusion) to multi-embedded logics (Lemma 6.12) could help overcome this strong limitation.

Both fusion and independent combination clarify the issue of completeness and decidability for the extension of normal modal logic to the multi-modal setting. These results do not really add anything new in the context of agent theories, as it has been known for some time that all the single agent theories presented in Section 1.3 extend to the MAS case. It must be said though that these techniques offer a clear explanation of the reasons behind this and draw a line on where these extensions might fail, i.e. in presence of interaction axioms.

As for the technique of fibring, the literature shows that most uses of it actually boil down to fusion, which, as noted above, has the problem of not being able to express interaction axioms, much needed for MAS theories. Still, at least in principle, fibring is more powerful than fusion because of the possibility of adding conditions on the fibring function. These conditions could encode interactions between the two classes of models that are being combined and therefore could represent interaction axioms between the two logics. At the time of writing, this is still an area of research with little or no results and it is therefore hard to say whether it will be successful. If so, I believe MAS theories with interactions axioms will be one of the ideal test-beds for fibring and, as a result, MAS theories could much benefit from this.

In view of all these considerations, our conclusion has unfortunately to be that the automatic transfer of meta-properties of the logics into non-trivial combinations of the type studied in this thesis is not possible with the current techniques. The only results that we have been able to achieve with combining logic methods refer to the MAS theories with no interaction axioms and these had already been proven by other means. The techniques presented in Section 6.2.2 were not applicable neither to any of the complex systems defined in Section 1.3.2 nor to the logic $S5WD_n$ examined in Chapter 3.

Clearly, the main difficulty with trying to apply combining logics to MAS theories is that while the latter heavily depends on interaction axioms between the various logics to be combined, the former hardly addresses this case at all. The reasons for this are not in a supposed indifference of the theorists about more applied problems, but in the *intrinsic difficulties* of the subject. As seen in Section 6.2.3, negative results of all sorts are just around the corner as soon as we start dealing with interaction axioms. For example, the transfer of one of the key properties for a MAS theory, decidability, is known to be particularly hard, as the following words by of one of the world-experts on this subject, Marcus Kracht, demonstrate:

> Polymodal logics are not just alleys, they are highways into undecidability. ([Kra95], page 93)

Notwithstanding this gloomy picture we have discussed, I still believe that the construction of cultural bridges between the two areas is a useful exercise. On the one hand, this could encourage MAS theorists to recognise the need of shifting the attention from the single case analysis to the general problem of the combination of mental states. On the other hand, the problem of interaction axioms (expressed in whatever format) between logic fragments in a combined logic might become more central in combining logics. Results of properties transfer, even if limited to very small classes of interaction axioms, would be of extreme importance to the whole area of MAS theories.

Appendix A

Some proofs for Chapter 4

The proofs that do not appear in Chapter 4 are reported here.

A.1 Interaction axioms of the form $\boxdot p \Rightarrow \boxdot \boxdot p$

A.1.1 Interaction axioms of the form $\Diamond_1 p \Rightarrow \boxdot \boxdot p$

A.1.1.0.1 $\Diamond_1 p \Rightarrow \Diamond_1 \Box_2 p$

Lemma A.1. *$F \models \Diamond_1 p \Rightarrow \Diamond_1 \Box_2 p$ if and only if $F = (W, \sim_1, \sim_2)$ is such that $\sim_2 = id_W$.*

Proof. From left to right. Consider a frame $F = (W, \sim_1, \sim_2)$ and two points $w, w' \in W$ such that $w \sim_2 w'$. Consider a valuation π such that $\pi(p) = \{w'\}$. We have $(F, \pi) \models_{w'} \Diamond_1 p$. Then $(F, \pi) \models_{w'} \Diamond_1 \Box_2 p$, and since p is true only at w' and the relation $\sim_2$ is reflexive this implies that $w = w'$.

From right to left. Consider any equivalence model $M = (W, \sim_1, \sim_2, \pi)$ such that $M \models_w \Diamond_1 p$. Then there exists a point $w' \in W$ such that $w \sim_1 w'$ and $M \models_{w'} p$. But since $\sim_2 = id_W$, we also have $M \models_{w'} \Box_2 p$ and so $M \models_w \Diamond_1 \Box_2 p$. □

Lemma A.2. $\vdash_{S5_2 + \{\Diamond_1 p \Rightarrow \Diamond_1 \Box_2 p\}} \Diamond_2 p \Rightarrow \Box_2 p$ *and* $\vdash_{S5_2 + \{\Diamond_2 p \Rightarrow \Box_2 p\}} \Diamond_1 p \Rightarrow \Diamond_1 \Box_2 p$.

Proof. First part. Suppose $\Diamond_1 p \Rightarrow \Diamond_1 \Box_2 p$; so $\Box_1 \Diamond_2 p \Rightarrow \Box_1 p$. Substitute the term $(p \Rightarrow \Box_2 p)$ for p uniformly in the axiom above; we obtain $\Box_1 \Diamond_2 (p \Rightarrow \Box_2 p) \Rightarrow \Box_1 (p \Rightarrow \Box_2 p)$. We prove that the antecedent of this formula is a theorem of $S5_2$. In fact we have $\neg \Box_2 p \vee \Box_2 p$; so from the axiom above and by Lemma 1.30 we have $\Diamond_2 \neg p \vee \Diamond_2 \Box_2 p$. Now since, as it can easily be verified, diamond distributes over logical or, we have $\Diamond_2 (\neg p \vee \Box_2 p)$, which by necessitating by $\Box_1$ leads to $\Box_1 \Diamond_2 (\neg p \vee \Box_2 p)$. So, it follows that $\Box_1 (p \Rightarrow \Box_2 p)$, which gives $p \Rightarrow \Box_2 p$. Now, from this formula it follows from Lemma 1.29 that $\Diamond_2 p \Rightarrow \Diamond_2 \Box_2 p$, which is equivalent to $\Diamond_2 p \Rightarrow \Box_2 p$.

Second part. Suppose $\Diamond_2 p \Rightarrow \Box_2 p$. We have $\Diamond_1 p \Rightarrow \Diamond_1 \Diamond_2 p$. So by using the assumption and Lemma 1.29 we obtain $\Diamond_1 p \Rightarrow \Diamond_1 \Box_2 p$.

□

A.1.1.0.2 $\Diamond_1 p \Rightarrow \Box_1 \Box_2 p$

Lemma A.3. *$F \models \Diamond_1 p \Rightarrow \Box_1 \Box_2 p$ if and only if $F = (W, \sim_1, \sim_2)$ is such that $\sim_2 = \sim_1 = id_W$.*

Proof. From left to right. Consider a frame $F = (W, \sim_1, \sim_2)$ and two points $w, w' \in W$ such that $w \sim_i w'$ for any $i \in \{1, 2\}$. Consider a valuation π such that $\pi(p) = \{w\}$. By reflexivity, we have $(F, \pi) \models_w \Diamond_1 p$. Then $(F, \pi) \models_w \Box_1\Box_2 p$, and so $(F, \pi) \models_w \Box_1 p \wedge \Box_2 p$. But since p is false at w', this implies that $w = w'$.

From right to left. Consider any equivalence model $M = (W, \sim_1, \sim_2, \pi)$ such that $M \models_w \Diamond_1 p$. Then there exists a point $w' \in W$ such that $w \sim_1 w'$ and $M \models_{w'} p$. But since $\sim_2 = \sim_1 = id_W$, we have that $w = w'$ and $M \models_w \Box_1\Box_2 p$. □

Lemma A.4. *$\vdash_{S5_2+\{\Diamond_1 p \Rightarrow \Box_1\Box_2 p\}} \Diamond_1 p \Rightarrow \Box_2 p$ and $\vdash_{S5_2+\{\Diamond_1 p \Rightarrow \Box_2 p\}} \Diamond_1 p \Rightarrow \Box_1\Box_2 p$.*

Proof. First part. Suppose $\Diamond_1 p \Rightarrow \Box_1\Box_2 p$. But $\Box_1\Box_2 p \Rightarrow \Box_2 p$ and so the result follows.

Second part. Suppose $\Diamond_1 p \Rightarrow \Box_2 p$ and necessitate by $\Box_1$. We have $\Box_1\Diamond_1 p \Rightarrow \Box_1\Box_2 p$, so we obtain $\Diamond_1 p \Rightarrow \Box_1\Box_2 p$.

□

A.1.1.0.3 $\Diamond_1 p \Rightarrow \Box_1\Diamond_2 p$

Lemma A.5. *$F \models \Diamond_1 p \Rightarrow \Box_1\Diamond_2 p$ if and only if $F = (W, \sim_1, \sim_2)$ is such that $\sim_1 \subseteq \sim_2$.*

Proof. From left to right. Consider a frame $F = (W, \sim_1, \sim_2)$ and two points $w, w' \in W$ such that $w \sim_1 w'$. Consider a valuation π such that $\pi(p) = \{w'\}$. We have $(F, \pi) \models_w \Diamond_1 p$. Then $(F, \pi) \models_w \Box_1\Diamond_2 p$, and so in particular $(F, \pi) \models_w \Diamond_2 p$. But p is true only at w' this implies that $w \sim_2 w'$.

From right to left. Consider any equivalence model $M = (W, \sim_1, \sim_2, \pi)$ such that $M \models_w \Diamond_1 p$. Then there exists a point $w' \in W$ such that $w \sim_1 w'$ and $M \models_{w'} p$. It remains to show that $M \models_w \Box_1\Diamond_2 p$ for which it is enough to show that for any point $w'' \in W$ such that $w \sim_1 w''$ we have $w'' \sim_2 w'$. But since $\sim_1 \subseteq \sim_2$ from $w \sim_1 w'$ we have $w \sim_2 w'$ and similarly for $w \sim_1 w''$ and $w \sim_2 w''$. Therefore by transitivity we have $w'' \sim_2 w'$. □

Lemma A.6. *$\vdash_{S5_2+\{\Diamond_1 p \Rightarrow \Box_1\Diamond_2 p\}} \Diamond_1 p \Rightarrow \Diamond_2 p$ and $\vdash_{S5_2+\{\Diamond_1 p \Rightarrow \Diamond_2 p\}} \Diamond_1 p \Rightarrow \Box_1\Diamond_2 p$.*

Proof. First part. Suppose $\Diamond_1 p \Rightarrow \Box_1\Diamond_2 p$. But $\Box_1\Diamond_2 p \Rightarrow \Diamond_2 p$ and so we have $\Diamond_1 p \Rightarrow \Diamond_2 p$.

Second part. Suppose $\Diamond_1 p \Rightarrow \Diamond_2 p$ and necessitate by $\Box_1$. We have $\Box_1\Diamond_1 p \Rightarrow \Box_1\Diamond_2 p$, so we obtain $\Diamond_1 p \Rightarrow \Box_1\Diamond_2 p$.

□

A.1.1.0.4 $\Diamond_1 p \Rightarrow \Box_2\Box_1 p$

Lemma A.7. *$F \models \Diamond_1 p \Rightarrow \Box_2\Box_1 p$ if and only if $F = (W, \sim_1, \sim_2)$ is such that $\sim_2 = \sim_1 = id_W$.*

Proof. From left to right. Consider a frame $F = (W, \sim_1, \sim_2)$ and two points $w, w' \in W$ such that $w \sim_i w'$ for any $i \in \{1, 2\}$. Consider a valuation π such that $\pi(p) = \{w\}$. By reflexivity, we have $(F, \pi) \models_w \Diamond_1 p$. Then $(F, \pi) \models_w \Box_2\Box_1 p$, and so $(F, \pi) \models_w \Box_1 p \wedge \Box_2 p$. But since p is false at w', this implies that $w = w'$.

From right to left. Consider any equivalence model $M = (W, \sim_1, \sim_2, \pi)$ such that $M \models_w \Diamond_1 p$. Then there exists a point $w' \in W$ such that $w \sim_1 w'$ and $M \models_{w'} p$. But since $\sim_2 = \sim_1 = id_W$, we have that $w = w'$ and $M \models_w \Box_2\Box_1 p$. □

Lemma A.8. *$\vdash_{S5_2+\{\Diamond_1 p \Rightarrow \Box_2\Box_1 p\}} \Diamond_1 p \Rightarrow \Box_2 p$ and $\vdash_{S5_2+\{\Diamond_1 p \Rightarrow \Box_2 p\}} \Diamond_1 p \Rightarrow \Box_2\Box_1 p$.*

Proof. First part. Suppose $\Diamond_1 p \Rightarrow \Box_2\Box_1 p$. But $\Box_2\Box_1 p \Rightarrow \Box_2 p$ and so the result follows.

Second part. Suppose $\Diamond_1 p \Rightarrow \Box_2 p$; but $\Box_2 p \Leftrightarrow \Box_2\Diamond_2 p$ and $\Diamond_2 p \Rightarrow \Box_1 p$. So we have $\Box_2\Diamond_2 p \Rightarrow \Box_2\Box_1 p$. So $\Diamond_1 p \Rightarrow \Box_2\Box_1 p$.

□

A.1.1.0.5 $\Diamond_1 p \Rightarrow \Box_2\Diamond_1 p$

Lemma A.9. *$F \models \Diamond_1 p \Rightarrow \Box_2\Diamond_1 p$ if and only if $F = (W, \sim_1, \sim_2)$ is such that $\sim_2 \subseteq \sim_1$.*

Proof. From left to right. Consider a frame $F = (W, \sim_1, \sim_2)$ and two points $w, w' \in W$ such that $w \sim_2 w'$. Consider a valuation π such that $\pi(p) = \{w\}$. By reflexivity, we have $(F, \pi) \models_w \Diamond_1 p$. Then $(F, \pi) \models_w \Box_2\Diamond_1 p$, and so in particular $(F, \pi) \models_{w'} \Diamond_1 p$. But p is true only at w, which implies that $w \sim_1 w'$.

From right to left. Consider any equivalence model $M = (W, \sim_1, \sim_2, \pi)$ such that $M \models_w \Diamond_1 p$. Then there exists a point $w' \in W$ such that $w \sim_1 w'$ and $M \models_{w'} p$. It remains to show that $M \models_w \Box_2\Diamond_1 p$ for which it is enough to show that for any point $w'' \in W$ such that $w \sim_2 w''$ we have $w'' \sim_1 w'$. But since $\sim_2 \subseteq \sim_1$, we have that $w \sim_1 w''$ and therefore we have $w'' \sim_1 w'$. □

Lemma A.10. *$\vdash_{S5_2+\{\Diamond_1 p \Rightarrow \Box_2\Diamond_1 p\}} \Diamond_2 p \Rightarrow \Diamond_1 p$ and $\vdash_{S5_2+\{\Diamond_2 p \Rightarrow \Diamond_1 p\}} \Diamond_1 p \Rightarrow \Box_2\Diamond_1 p$.*

Proof. First part. Suppose $\Diamond_1 p \Rightarrow \Box_2\Diamond_1 p$. We have $\Diamond_2 p \Rightarrow \Diamond_2\Diamond_1 p$, but $\Diamond_1 p \Leftrightarrow \Box_1\Diamond_1 p$; so we have $\Diamond_2 p \Rightarrow \Diamond_2\Box_1\Diamond_1 p$. But by using the contrapositive of our assumption we deduce $\Diamond_2 p \Rightarrow \Box_1\Diamond_1 p$, which gives $\Diamond_2 p \Rightarrow \Diamond_1 p$.

Second part. Suppose $\Diamond_2 p \Rightarrow \Diamond_1 p$ and substitute in there uniformly $\Box_1 p$ for p. We obtain $\Diamond_2\Box_1 p \Rightarrow \Diamond_1\Box_1 p$; but then we have $\Diamond_2\Box_1 p \Rightarrow \Box_1 p$, which is the contrapositive of what we needed to prove.

□

A.1.2 Interaction axioms of the form $\Box_1 p \Rightarrow \boxdot\boxdot p$

A.1.2.0.6 $\Box_1 p \Rightarrow \Box_1\Box_2 p$

Lemma A.11. *$F \models \Box_1 p \Rightarrow \Box_1\Box_2 p$ if and only if $F = (W, \sim_1, \sim_2)$ is such that $\sim_2 \subseteq \sim_1$.*

Proof. From left to right. Consider a frame $F = (W, \sim_1, \sim_2)$ and two points $w, w' \in W$ such that $w \sim_2 w'$. Consider a valuation π such that $\pi(p) = \{w'\}$. By reflexivity, we have $(F, \pi) \models_w \Diamond_1\Diamond_2 p$. Then $(F, \pi) \models_w \Diamond_1 p$. But p is true only at w', which implies that $w \sim_1 w'$.

From right to left. Consider any equivalence model $M = (W, \sim_1, \sim_2, \pi)$ such that $M \models_w \Box_1 p$. So for every w' such that $w \sim_1 w'$ we have that $M \models_{w'} p$; indeed since the model is an equivalence model we have $M \models_{w'} \Box_1 p$, for any $w' \in [w]_{\sim_1}$. Consider now $[w']_{\sim_2}$; by hypothesis we have $[w']_{\sim_2} \subseteq [w']_{\sim_1}$. So $[w']_{\sim_2} \subseteq [w]_{\sim_1}$ and so $M \models_w \Box_1\Box_2 p$. □

Lemma A.12. *$\vdash_{S5_2+\{\Box_1 p \Rightarrow \Box_1\Box_2 p\}} \Box_1 p \Rightarrow \Box_2 p$ and $\vdash_{S5_2+\{\Box_1 p \Rightarrow \Box_2 p\}} \Box_1 p \Rightarrow \Box_1\Box_2 p$.*

Proof. First part. Suppose $\Box_1 p \Rightarrow \Box_1\Box_2 p$. But $\Box_1\Box_2 p \Rightarrow \Box_2 p$, so we have $\Box_1 p \Rightarrow \Box_2 p$
Second part. Suppose $\Box_1 p \Rightarrow \Box_2 p$ and necessitate by $\Box_1$. We have $\Box_1\Box_1 p \Rightarrow \Box_1\Box_2 p$, so we obtain $\Box_1 p \Rightarrow \Box_1\Box_2 p$.

□

A.1.2.0.7 $\Box_1 p \Rightarrow \Box_2\Box_1 p$

Lemma A.13. *$F \models \Box_1 p \Rightarrow \Box_2\Box_1 p$ if and only if $F = (W, \sim_1, \sim_2)$ is such that $\sim_2 \subseteq \sim_1$.*

Proof. From left to right. Consider a frame $F = (W, \sim_1, \sim_2)$ and two points $w, w' \in W$ such that $w \sim_2 w'$. Consider a valuation π such that $\pi(p) = \{w'\}$. By reflexivity, we have $(F, \pi) \models_w \Diamond_2\Diamond_1 p$. Then $(F, \pi) \models_w \Diamond_1 p$. But p is true only at w', which implies that $w \sim_1 w'$.

From right to left. Consider any equivalence model $M = (W, \sim_1, \sim_2, \pi)$ such that $M \models_w \Box_1 p$. So for every w' such that $w \sim_1 w'$ we have that $M \models_{w'} p$; indeed since the model is an equivalence model we have $M \models_{w'} \Box_1 p$, for any $w' \in [w]_{\sim_1}$. Consider now $[w]_{\sim_2}$; by hypothesis we have $[w]_{\sim_2} \subseteq [w]_{\sim_1}$. So $M \models_w \Box_2\Box_1 p$. $\Box$

Lemma A.14. *$\vdash_{S5_2+\{\Box_1 p \Rightarrow \Box_2\Box_1 p\}} \Box_1 p \Rightarrow \Box_2 p$ and $\vdash_{S5_2+\{\Box_1 p \Rightarrow \Box_2 p\}} \Box_1 p \Rightarrow \Box_2\Box_1 p$.*

Proof. First part. Suppose $\Box_1 p \Rightarrow \Box_2\Box_1 p$. But $\Box_1 p \Rightarrow p$, and so $\Box_2\Box_1 p \Rightarrow \Box_2 p$ and so by transitivity we have the result.

Second part. Suppose $\Box_1 p \Rightarrow \Box_2 p$ and substitute uniformly in this formula $\Box_1 p$ for p. We have $\Box_1\Box_1 p \Rightarrow \Box_2\Box_1 p$, so we obtain the desired result $\Box_1 p \Rightarrow \Box_2\Box_1 p$.

$\Box$

A.1.2.0.8 $\Box_1 p \Rightarrow \Box_2\Diamond_1 p$

Lemma A.15. *$F \models \Box_1 p \Rightarrow \Box_2\Diamond_1 p$ if and only if $F = (W, \sim_1, \sim_2)$ is such that $\sim_2 \subseteq \sim_1$.*

Proof. From left to right. Consider a frame $F = (W, \sim_1, \sim_2)$ and two points $w, w' \in W$ such that $w \sim_2 w'$. Consider a valuation π such that $\pi(p) = \{[w']_{\sim_1}\}$. We have $(F, \pi) \models_w \Diamond_2\Box_1 p$. Then $(F, \pi) \models_w \Diamond_1 p$. So there exists a point $w'' \in [w']_{\sim_1}$ such that $w \sim_1 w''$; but then $w \sim_1 w'$.

From right to left. Consider any equivalence model $M = (W, \sim_1, \sim_2, \pi)$ such that $M \models_w \Box_1 p$. Since $\sim_2 \subseteq \sim_1$ we have also $M \models_w \Box_2 p$. But then $M \models_w \Box_2\Diamond_1 p$ by reflexivity. $\Box$

Lemma A.16. *$\vdash_{S5_2+\{\Box_1 p \Rightarrow \Box_2\Diamond_1 p\}} \Box_1 p \Rightarrow \Box_2 p$ and $\vdash_{S5_2+\{\Box_1 p \Rightarrow \Box_2 p\}} \Box_1 p \Rightarrow \Box_2\Diamond_1 p$.*

Proof. First part. Suppose $\Box_1 p \Rightarrow \Box_2\Diamond_1 p$. We have $\Diamond_2 p \Rightarrow \Diamond_2\Diamond_1 p$ and so, by Lemma 1.30, we have $\Diamond_2 p \Rightarrow \Diamond_2\Box_1\Diamond_1 p$. But then by using our assumption we can infer $\Diamond_2\Box_1\Diamond_1 p \Rightarrow \Diamond_1\Diamond_1 p$. So, by transitivity and $\vdash \Diamond_1 p \Leftrightarrow \Diamond_1\Diamond_1 p$ we have $\Diamond_2 p \Rightarrow \Diamond_1 p$, which is the result we aimed for.

Second part. Suppose $\Box_1 p \Rightarrow \Box_2 p$. But $\Box_2 p \Rightarrow \Box_2\Diamond_1 p$. So we have $\Box_1 p \Rightarrow \Box_2\Diamond_1 p$.

$\Box$

A.2 Interaction axioms of the form $\boxdot \boxdot p \Rightarrow \boxdot \boxdot p$

A.2.1 Interaction axioms of the form $\Diamond_1\Box_2 p \Rightarrow \boxdot \boxdot p$

A.2.1.0.9 $\Diamond_1\Box_2 p \Rightarrow \Box_1\Box_2 p$

Lemma A.17. *$\vdash_{S5_2+\{\Diamond_1\Box_2 p \Rightarrow \Box_1\Box_2 p\}} \Box_2 p \Rightarrow \Box_1 p$ and $\vdash_{S5_2+\{\Box_2 p \Rightarrow \Box_1 p\}} \Diamond_1\Box_2 p \Rightarrow \Box_1\Box_2 p$.*

Proof. First part. Suppose $\Diamond_1\Box_2 p \Rightarrow \Box_1\Box_2 p$. We have $\Diamond_1 p \Rightarrow \Diamond_1\Diamond_2 p$. But since we have $\Diamond_1\Diamond_2 p \Rightarrow \Box_1\Diamond_2 p$, we obtain $\Diamond_1 p \Rightarrow \Box_1\Diamond_2 p$, which in turn implies $\Diamond_1 p \Rightarrow \Diamond_2 p$.

Second part. Suppose $\Diamond_1 p \Rightarrow \Diamond_2 p$ and substitute $\Box_2 p$ for p in it. We obtain $\Diamond_1\Box_2 p \Rightarrow \Diamond_2\Box_2 p$, which is equivalent to $\Diamond_1\Box_2 p \Rightarrow \Box_2 p$. Now, by Lemma 1.29 we obtain $\Box_1\Diamond_1\Box_2 p \Rightarrow \Box_1\Box_2 p$, which, given Lemma 1.30 gives us to the result $\Diamond_1\Box_2 p \Rightarrow \Box_1\Box_2 p$.

$\Box$

A.2.1.0.10 $\Diamond_1 \Box_2 p \Rightarrow \Box_1 \Diamond_2 p$

Lemma A.18. $\vdash_{S5_2 + \{\Diamond_1\Box_2 p \Rightarrow \Box_1\Diamond_2 p\}} \Box_2 p \Rightarrow \Box_1 p$ *and* $\vdash_{S5_2+\{\Box_2 p \Rightarrow \Box_1 p\}} \Diamond_1\Box_2 p \Rightarrow \Box_1\Diamond_2 p$.

Proof. First part. Suppose $\Diamond_1\Box_2 p \Rightarrow \Box_1\Diamond_2 p$. We have $\Box_2 p \Rightarrow \Diamond_1\Box_2 p$. So, by Lemma 1.30 we have $\Box_2 p \Rightarrow \Diamond_1\Box_2\Box_2 p$. From this, by applying our assumption, we obtain $\Box_2 p \Rightarrow \Box_1\Diamond_2\Box_2 p$, which is equivalent to $\Box_2 p \Rightarrow \Box_1\Box_2 p$. By an instance of axiom T we then obtain $\Box_2 p \Rightarrow \Box_1 p$.

Second part. Suppose $\Box_2 p \Rightarrow \Box_1 p$. By Lemma 1.29 we have $\Diamond_1\Box_2 p \Rightarrow \Diamond_1\Box_1 p$, which is equivalent to $\Diamond_1\Box_2 p \Rightarrow \Box_1 p$. But then by an instance of axiom T we obtain $\Diamond_1\Box_2 p \Rightarrow \Box_1\Diamond_2 p$. $\Box$

A.2.1.0.11 $\Diamond_1\Box_2 p \Rightarrow \Box_2\Box_1 p$

Lemma A.19. $\vdash_{S5_2+\{\Diamond_1\Box_2 p \Rightarrow \Box_2\Box_1 p\}} \Box_2 p \Rightarrow \Box_1 p$ *and* $\vdash_{S5_2+\{\Box_2 p \Rightarrow \Box_1 p\}} \Diamond_1\Box_2 p \Rightarrow \Box_2\Box_1 p$.

Proof. First part. Suppose $\Diamond_1\Box_2 p \Rightarrow \Box_2\Box_1 p$. By axiom T we have $\Diamond_1 p \Rightarrow \Diamond_2\Diamond_1 p$. By the contrapositive of our assumption we have the implication $\Diamond_2\Diamond_1 p \Rightarrow \Box_1\Diamond_2 p$. So we obtain $\Diamond_1 p \Rightarrow \Diamond_2 p$.

Second part. Suppose $\Diamond_1 p \Rightarrow \Diamond_2 p$ and substitute $\Diamond_2 p$ for p in it. We have $\Diamond_1\Diamond_2 p \Rightarrow \Diamond_2\Diamond_2 p$, which by Lemma 1.30 is equivalent to $\Diamond_1\Diamond_2 p \Rightarrow \Diamond_2 p$. Now by Lemma 1.29 we have $\Box_1\Diamond_1\Diamond_2 p \Rightarrow \Box_1\Diamond_2 p$, which is equivalent to $\Diamond_1\Diamond_2 p \Rightarrow \Box_1\Diamond_2 p$. By taking the contrapositive of this formula we have the formula $\Diamond_1\Box_2 p \Rightarrow \Box_1\Box_2 p$. All it remains to prove is that under the assumption we also have $\Box_1\Box_2 p \Rightarrow \Box_2\Box_1 p$, i.e. $\Diamond_2\Diamond_1 p \Rightarrow \Diamond_1\Diamond_2 p$. To see this apply Lemma 1.29 to the formula $\Diamond_1 p \Rightarrow \Diamond_2 p$ to get $\Diamond_2\Diamond_1 p \Rightarrow \Diamond_2\Diamond_2 p$. By Lemma 1.30 we get $\Diamond_2\Diamond_1 p \Rightarrow \Diamond_2 p$. By applying axiom T and transitivity to this formula we can get $\Diamond_2\Diamond_1 p \Rightarrow \Diamond_1\Diamond_2 p$, which by the observation above is the result we need.

$\Box$

A.2.1.0.12 $\Diamond_1\Box_2 p \Rightarrow \Box_2\Diamond_1 p$

Lemma A.20. *The logic $S5_2 + \{\Diamond_1\Box_2 p \Rightarrow \Box_2\Diamond_1 p\}$ is sound and complete with respect to equivalence 2WD frames, i.e. frames $F = (W, \sim_1, \sim_2)$ satisfying the property for all $w, w', w'' \in W$ such that $w \sim_1 w', w \sim_2 w''$ there exists a point $\overline{w}$ such that $w' \sim_2 \overline{w}, w'' \sim_1 \overline{w}$.*

Proof. The lemma follows from Theorem 3.12 for the case $n = 2$. $\Box$

A.2.2 Interaction axioms of the form $\Diamond_1\Diamond_2 p \Rightarrow \boxdot \boxdot p$

A.2.2.0.13 $\Diamond_1\Diamond_2 p \Rightarrow \Diamond_1\Box_2 p$

Lemma A.21. $\vdash_{S5_2+\{\Diamond_1\Diamond_2 p \Rightarrow \Diamond_1\Box_2 p\}} \Diamond_2 p \Rightarrow \Diamond_1\Box_2 p$ *and* $\vdash_{S5_2+\{\Diamond_2 p \Rightarrow \Diamond_1\Box_2 p\}} \Diamond_1\Diamond_2 p \Rightarrow \Diamond_1\Box_2 p$.

Proof. First part. Suppose $\Diamond_1\Diamond_2 p \Rightarrow \Diamond_1\Box_2 p$. We have $\Diamond_2 p \Rightarrow \Diamond_1\Diamond_2 p$, so by applying our assumption we obtain $\Diamond_2 p \Rightarrow \Diamond_1\Box_2 p$.

Second part. Suppose $\Diamond_2 p \Rightarrow \Diamond_1\Box_2 p$. From Lemma 1.29 we obtain $\Diamond_1\Diamond_2 p \Rightarrow \Diamond_1\Diamond_1\Box_2 p$; so by Lemma 1.30 we have $\Diamond_1\Diamond_2 p \Rightarrow \Diamond_1\Box_2 p$.

$\Box$

A.2.2.0.14 $\Diamond_1\Diamond_2 p \Rightarrow \Box_1\Box_2 p$

Lemma A.22. $\vdash_{S5_2+\{\Diamond_1\Diamond_2 p\Rightarrow\Box_1\Box_2 p\}} \Diamond_1 p \Rightarrow \Box_2 p$ *and* $\vdash_{S5_2+\{\Diamond_1 p\Rightarrow\Box_2 p\}} \Diamond_1\Diamond_2 p \Rightarrow \Box_1\Box_2 p$.

Proof. First part. Suppose $\Diamond_1\Diamond_2 p \Rightarrow \Box_1\Box_2 p$. We have $\Diamond_1 p \Rightarrow \Diamond_1\Diamond_2 p$, so by applying the contrapositive of our assumption and transitivity we obtain $\Diamond_1 p \Rightarrow \Box_1\Box_2 p$, from which we obtain $\Diamond_1 p \Rightarrow \Box_2 p$.

Second part. Suppose $\Diamond_1 p \Rightarrow \Box_2 p$. So, by axiom T we obtain $\Diamond_1 p \Leftrightarrow \Box_2 p \Leftrightarrow \Diamond_2 p \Leftrightarrow \Box_1 p$. So we have $\Diamond_1\Diamond_2 p \Rightarrow \Box_1\Box_2 p$.

□

A.2.2.0.15 $\Diamond_1\Diamond_2 p \Rightarrow \Diamond_2\Box_1 p$

Lemma A.23. $\vdash_{S5_2+\{\Diamond_1\Diamond_2 p\Rightarrow\Diamond_2\Box_1 p\}} \Diamond_2 p \Rightarrow \Diamond_2\Box_1 p$ *and* $\vdash_{S5_2+\{\Diamond_2 p\Rightarrow\Diamond_2\Box_1 p\}} \Diamond_1\Diamond_2 p \Rightarrow \Diamond_2\Box_1 p$.

Proof. First part. Suppose $\Diamond_1\Diamond_2 p \Rightarrow \Diamond_2\Box_1 p$. We have $\Diamond_2 p \Rightarrow \Diamond_1\Diamond_2 p$, so by applying our assumption we obtain $\Diamond_2 p \Rightarrow \Diamond_2\Box_1 p$.

Second part. Suppose $\Diamond_2 p \Rightarrow \Diamond_2\Box_1 p$. So by Lemma 1.29 we have $\Diamond_1\Diamond_2 p \Rightarrow \Diamond_1\Diamond_2\Box_1 p$. But by a proof equivalent to the one presented in Lemma A.2 it can be shown that when $\Diamond_2 p \Rightarrow \Diamond_2\Box_1 p$ holds also $\Diamond_1 p \Rightarrow \Box_1 p$ holds. So by using this observation we obtain the formula $\Diamond_1\Diamond_2\Box_1 p \Rightarrow \Box_1\Diamond_2\Box_1 p$; so by axiom T we have $\Diamond_1\Diamond_2 p \Rightarrow \Diamond_2\Box_1 p$.

□

A.2.2.0.16 $\Diamond_1\Diamond_2 p \Rightarrow \Diamond_2\Diamond_1 p$

Lemma A.24. $\vdash_{S5_2+\{\Diamond_1\Diamond_2 p\Rightarrow\Diamond_2\Diamond_1 p\}}\Diamond_1\Box_2 p \Rightarrow \Box_2\Diamond_1 p$ *and* $\vdash_{S5_2+\{\Diamond_1\Box_2 p\Rightarrow\Box_2\Diamond_1 p\}}\Diamond_1\Diamond_2 p \Rightarrow \Diamond_2\Diamond_1 p$.

Proof. It follows from Lemma 4.24 by considering the contrapositives of the two formulae.

□

A.2.2.0.17 $\Diamond_1\Diamond_2 p \Rightarrow \Box_2\Box_1 p$

Lemma A.25. $\vdash_{S5_2+\{\Diamond_1\Diamond_2 p\Rightarrow\Box_2\Box_1 p\}} \Diamond_1 p \Rightarrow \Box_2 p$ *and* $\vdash_{S5_2+\{\Diamond_1 p\Rightarrow\Box_2 p\}} \Diamond_1\Diamond_2 p \Rightarrow \Box_2\Box_1 p$.

Proof. First part. Assume $\Diamond_1\Diamond_2 p \Rightarrow \Box_2\Box_1 p$. So, we have $\Diamond_1 p \Rightarrow \Diamond_1\Diamond_2 p$. But by the assumption we infer $\Diamond_1 p \Rightarrow \Box_2\Box_1 p$. So by axiom T we have $\Diamond_1 p \Rightarrow \Box_2 p$.

Second part. Assume $\Diamond_1 p \Rightarrow \Box_2 p$ and substitute $\Diamond_2 p$ for p. We have $\Diamond_1\Diamond_2 p \Rightarrow \Box_2\Diamond_2 p$. But under the hypothesis $\Diamond_2 p \Rightarrow \Box_1 p$, so by Lemma 1.29 we obtain $\Diamond_1\Diamond_2 p \Rightarrow \Box_2\Box_1 p$.

□

Bibliography

[AGB] Why are agents important?
http://www.agentbuilder.com/Documentation/whyAgents.html.

[Aus62] J. L. Austin. *How To Do Things with Words*. Harvard University Press, Cambridge, Massachusetts, 1962.

[BAHP82] M. Ben-Ari, J. Y. Halpern, and A. Pnueli. Deterministic propositional dynamic logic: Finite models, complexity, and completeness. *Journal of Computer and System Sciences*, 25:402–417, 1982.

[Bal98] M. Baldoni. *Normal Multimodal Logics: Automatic Deduction and Logic Programming Extension*. PhD thesis, Dipartimento di Informatica, Universitá degli Studi di Torino, Italy, 1998.

[Bar81] J. Barwise. Scense and other situations. *The Journal of Philosophy*, 78(7):369–397, July 1981.

[Bay58] A. Bayart. La correction de la logique modale du premier et second ordre S5. *Logique et Analyse*, 1:28–44, 1958.

[BD93] C. Böttcher and O. Dressler. Diagnosis process dynamics: Holding the diagnostic trackhound in leash. In R. Bajcsy, editor, *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI'93)*, pages 1460–1471. Morgan Kauffmann, 1993.

[BdR97a] P. Blackburn and M. de Rijke. Why combine logics? *Studia Logica*, 59:5–27, 1997. Edited by D. Gabbay and F. Pirri.

[BdR97b] P. Blackburn and M. de Rijke. Zooming in, zooming out. *Journal of Logic, Language and Information*, 6:5–31, 1997.

[BdRV99] P. Blackburn, Maarten de Rijke, and Y. Venema. Modal logic. Draft, 1999.

[Bei84] B. Beizer. *Software System Testing and Quality Assurance*. Van Nostrand Reinhold, New York, 1984.

[Bei90] B. Beizer. *Software Testing Techniques*. Van Nostrand Reinhold, New York, 2nd edition, 1990.

[Ben75] J. F. K. A. van Benthem. A note on modal formulae and relational properties. *The Journal of Symbolic Logic*, 40:55–58, 1975.

[Ben83] J. van Benthem. *The Logic of Time*. D. Reidel, Dordrecht, The Netherlands, 1983.

[Ben84] J. van Benthem. Correspondence theory. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, Volume II: Extensions of Classical Logic*, volume 165 of *Synthese Library*, chapter II.4, pages 167–247. D. Reidel Publ. Co., Dordrecht, 1984.

[Ber49] G. Bergmann. The finite representations of S5. *Methodos*, 1:217–219, 1949.

[BMS98] A. Baltag, L. S. Moss, and S. Solecki. The logic of public announcement, common knowledge, and private suspicions. In I. Gilboa, editor, *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK-98)*, pages 125–132, San Francisco, July 22–24 1998. Morgan Kaufmann.

[Boo86] G. Booch. Object-Oriented Development. *IEEE Transactions on Software Engineering*, SE-12(2):211–221, February 1986.

[Boo93] G. Boolos. *The Logic of Provability*. Cambridge University Press, 1993.

[BP84] J. Barwise and J. Perry. *Situations and attitudes*. MIT Press/Barford Boks, Cambridge MA, 1984.

[Bra83] M. E. Bratman. Casteñada's theory of thought and action. In J. Toberlin, editor, *Agent, Language, and the structure of the world: Essays presented to Hector-Neri Casteñada with his replies*, pages 149–169. Hackett, Indianapolis, 1983.

[Bra87] M. E. Bratman. *Intentions, Plans, and Practical Reason*. Harvard University Press: Cambridge, MA, 1987.

[Bra90] M. E. Bratman. What is intention? In P. R. Cohen, J. L. Morgan, and M. E. Pollack, editors, *Intentions in Communication*, pages 15–32. The MIT Press: Cambridge, MA, 1990.

[Bra97] J. M. Bradshaw. *Software Agents*. AAAI Press, Mento Prak, USA, 1997.

[BS84] R. A. Bull and K. Segerberg. Basic modal logic. In D. Gabbay and F. Guenther, editors, *Handbook of Philosophical Logic Volume II — Extensions of Classical Logic*, pages 1–88. D. Reidel Publishing Company: Dordrecht, The Netherlands, 1984. (Synthese library Volume 164).

[Car46] R. Carnap. Modalities and quantification. *Journal of Symbolic Logic*, 11:33–64, 1946.

[Cas75] N. H. Casteñada. *Thinking and Doing*. Reiter, Dodrecht, Netherlands, 1975.

[Cas98] C. Castelfranchi. Modelling social action for AI agents. *Artificial Intelligence*, 103(1-2):95–116, August 1998.

[Cat88] L. Catach. Normal multimodal logics. In T. M. Smith and G. R. Mitchell, editor, *Proceedings of the 7th National Conference on Artificial Intelligence (AAAI'88)*, pages 491–495, St. Paul, MN, August 1988. Morgan Kaufmann.

[CC95] R. Conte and C. Castelfranchi. *Cognitive and Social Action*. UCL Press, London, 1995.

[CE81] E. M. Clarke and E. A. Emerson. Design and synthesis of synchronization skeletons using branching time temporal logic. In *Proceedings of the Workshop on Logics of Programs*, volume 131 of *Lecture Notes in Computer Science*, pages 52–71. Springer, 1981.

[CES86] E. M. Clarke, E. A. Emerson, and A. P. Sistla. Automatic verification of finite-state concurrent systems using temporal logic specifications. *ACM Trans. Programming Languages and Systems*, 8(2):244–263, 1986.

[Che80] B. Chellas. *Modal Logic*. Cambridge University Press, Cambridge, 1980.

[CL90] P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42(2-3):213–261, March 1990.

[CM96] A. Chavez and P. Maes. Kasbah: An agent marketplace for buying and selling goods. In *First International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology*, pages 75–90, 1996.

[Col93] M. Colombetti. Formal semantics for mutual belief. *Artificial Intelligence*, 62:341–353, 1993.

[Col96] M. Colombetti. A formal model of intentional communication. In *Preprints of the 2nd Conference on Logic and the Foundations of the Theory of Games and Decisions (LOFT 2)*, Turin, December 1996.

[CP79] P. R. Cohen and C. R. Perrault. Elements of a plan-based theory of speech acts. *Cognitive Science*, 3(3):177–212, 1979.

[CZ97] A. Chagorov and M Zakharyaschev. *Modal Logic*, volume 35 of *Oxford Logic Guides*. Clarendon Press, Oxford, 1997.

[Dav93] N. J. Davies. *Truth, Modality, and Action*. PhD thesis, Department of Computer Science, University of Essex, Colchester, UK, March 1993.

[Dav99] H. David. The application of Kripke trees to model refinement and model checking for $S5_n$. Final year project. School of Computer Science. University of Birmingham, July 1999.

[Den87] D. Dennet. *The Intensional Stance*. MIT Press, 1987.

[dFL$^+$97] M. d'Inverno, M. Fisher, A. Lomuscio, M. Luck, M. de Rijke, M. Ryan, and M. Wooldridge. Formalisms for multi-agent systems. *Knowledge Engineering Review*, 12(3):315–321, 1997.

[dMAE$^+$97] R. López de Mantaras, J. Amat, F. Esteva, M. López, and C. Sierra. Generation of unknown environment maps by cooperative low-cost robots. In W. L. Johnson and B. Hayes-Roth, editors, *Proceedings of the 1st International Conference on Autonomous Agents*, pages 164–169, New York, February5–8 1997. ACM Press.

[Dow97] M. Dowson. The ARIANE 5 sofware failure. *ACM SIGSOFT Software Engineering Notes*, 22(2):84, March 1997.

[dRB96] M. de Rijke and P. Blackburn, editors. *Notre Dame Journal of Formal Logic*, volume 37(2), Special Issue on Combining Logics. University of Notre Dame, Indiana, Spring 1996.

[Dun86] J. M. Dunn. Relevant logic and entailment. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume III, pages 117–224. Reidel, Dordrecht, 1986.

[EH85] E. A. Emerson and J. Y. Halpern. Decision procedures and expressiveness in the temporal logic of branching time. *Journal of Computer and System Sciences*, 30:1–24, 1985.

[Eme90] E. A. Emerson. Temporal and modal logic. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, chapter 16, pages 996 – 1071. Elsevier Science Publishers, 1990.

[End72] H. B. Enderton. *A Mathematical Introduction to Logic*. Academic Press, Inc., Orlando, Florida, 1972.

[EV98] J. Engelfriet and Y. Venema. A modal logic of information change. In I. Gilboa, editor, *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK-98)*, pages 125–132, San Francisco, July 22–24 1998. Morgan Kaufmann.

[EW94] O. Etzioni and D. Weld. A softbot-based interface to the internet. *Communications of the ACM*, 37(7):72–76, July 1994.

[Ewa86] W. B. Ewald. Intuitionistic tense and modal logic. *Journal of Symbolic Logic*, 51:166–179, 1986.

[FG92] M. Finger and D. M. Gabbay. Adding a temporal dimension to a logic. *Journal of Logic, Language and Information*, 1:203–233, 1992.

[FG96a] M. Finger and D. M. Gabbay. Combining temporal logic systems. *Notre Dame Journal of Formal Logic*, 37(2), Special Issue on Combining Logics, Spring 1996.

[FG96b] S. Franklin and A. Graesser. Is it an agent, or just a program? In J. P. Müller, M. J. Wooldridge, and N. R. Jennings, editors, *Intelligent Agents III — Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages (ATAL-96)*, volume 1193 of *Lecture Notes in Artificial Intelligence*, pages 193–206. Springer-Verlag, Heidelberg, 1996.

[FH88] R. Fagin and J. Y. Halpern. Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34(1):39–76, January 1988.

[FH94] N. Friedman and J. Y. Halpern. A knowledge-based framework for belief change, part I: Foundations. In R. Fagin, editor, *Proceedings of the 5th Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 44–64, San Francisco, CA, USA, March 1994. Morgan Kaufmann Publishers Inc.

[FHMV95] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about Knowledge*. MIT Press, Cambridge, 1995.

[FHMV97] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. Knowledge-based programs. *Distributed Computing*, 10(4):199–225, 1997.

[FHV92] R. Fagin, J. Y. Halpern, and M. Y. Vardi. What can machines know? On the properties of knowledge in distributed systems. *Journal of the ACM*, 39(2):328–376, April 1992.

[FHV95] R. Fagin, J. Y. Halpern, and M. Vardi. A nonstandard approach to the logical omniscience problem. *Artificial Intelligence*, 79, 1995.

[FIP] FIPA: Foundation for intelligent physical agents.
http://www.fipa.org.

[Fit48] F. B. Fitch. Intuitionistic modal logic with quantifiers. *Portugalia Mathematica*, 7:113–118, 1948.

[Fit69] M. Fitting. Logics with several modal operators. *Theoria*, 35:259–266, 1969.

[Fit83] M. Fitting. *Proof Methods for Modal and Intuitionistic Logics*. Reidel, Dordrecht, 1983.

[FL79] M. J. Fischer and R. E. Ladner. Propositional dynamic logic of regular programs. *Journal of Computer and System Sciences*, 18(2):194–211, April 1979.

[FNS98] P. Fröhlich, W. Nejdl, and M. Schroeder. Strategies in model-based diagnosis. *Journal of Automated Reasoning*, 20(1,2):81–105, 1998.

[FS77] G. Fischer-Servi. On modal logic with an intuitionistic modal calculi. *Studia Logica*, 36, 1977.

[FS96] K. Fine and G. Schurz. Transfer theorems for stratified multimodal logics. In J. Copeland, editor, *Logic and Reality, Essays in Pure and Applied Logic. In memory of Arthur Prior*, pages 169–213. Oxford University Press, 1996.

[Gab96a] D. Gabbay. An overview of fibred semantics and the combination of logics. In F. Baader and K. U. Schulz, editors, *Frontiers of Combining Systems: Proceedings of the 1st International Workshop, Munich (Germany)*, Applied Logic, pages 1–56. Kluwer, March 1996.

[Gab96b] D. M. Gabbay. Fibred semantics and the weaving of logics. Part 1. Modal and intuitionistic logics. *Journal of Symbolic Logic*, 61(4):1057–1120, December 1996.

[Gab99] D. Gabbay. *Fibring Logics*. Cambridge University Press, 1999. Forthcoming.

[Gam91] L. T. F. Gamut. *Logic, Language and Meaning*, volume 2. The University of Chicago Press, 1991.

[Gar84] J. Garson. Quantification in modal logic. In *Handbook of Philosophical Logic*, volume 2, pages 249–307. D. Gabbay and F. Guenthner Eds., 1984.

[Ger99] J. Gerbrandy. *Bisimulations on planet Kripke*. PhD thesis, Institute for Logic, Language and Computation, University of Amsterdam, 1999.

[GG98] D. M. Gabbay and G. Governatori. Dealing with label dependent deontic modalities. In P. McNamara and H. Prakken, editors, *Norms, Logics and Information Systems*, pages 311–330. IOS Press, Amsterdam, 1998.

[GHR93] D. M. Gabbay, I. M. Hodkinson, and M. A. Reynolds. *Temporal Logic: Mathematical Foundations and Computational Aspects, Volume 1:Mathematical Foundations*. Oxford University Press, 1993.

[GL87] M. P. Georgeff and A. L. Lansky. Reactive reasoning and planning. In *The Proceedings of AAAI-87*, pages 677–682, Seattle, 1987.

[GN97] D. Gabbay and R. Nossum. Structured context with fibred semantics. Unpublished, 1997.

[Gol75] R. Goldblatt. First-order definability in modal logic. *The Journal of Symbolic Logic*, 40:35–40, 1975.

[Gol91] R. Goldblatt. The McKinsey axiom is not canonical. *The Journal of Symbolic Logic*, 56(2):554–562, June 1991.

[Gol92] R. Goldblatt. *Logics of Time and Computation, Second Edition, Revised and Expanded*, volume 7 of *CSLI Lecture Notes*. CSLI, Stanford, 1992. Distributed by University of Chicago Press.

[GP92] V. Goranko and S. Passy. Using the universal modality: Gains and questions. *Journal of Logic and Computation*, 2(1):5–30, March 1992.

[GPSS80] D. M. Gabbay, A. Pnueli, S. Shelah, and J. Stavi. On the temporal basis of fairness. In *Conference Record of the Seventh Annual ACM Symposium on Principles of Programming Languages*, pages 163–173, Las Vegas, Nevada, January 1980.

[GS58] G. Gamow and M. Stern. *Puzzle Math*. Viking Press, New York, 1958.

[GS98] D. Gabbay and V. Shehtman. Products of modal logics, part 1. *Logic Journal of the IGPL*, 6(1):73–146, 1998.

[Ham78] A. G. Hamilton. *Logic for Mathematicians*. Cambridge University Press, 1978.

[HB99a] A. L. G. Hayzelden and J. Bigham, editors. *Software Agents for Future Communications Systems*. Springer Verlag, 1999.

[HB99b] A. L. G. Hayzelden and J. Bigham. Software agents in communications network management: An overview. *Knowledge Engineering Review Journal*, 1999. To appear.

[HC84] G. E. Hughes and M. J. Cresswell. *A Companion to Modal Logic*. Methuen, London, 1984.

[HC96] G. E. Hughes and M. J. Cresswell. *A New Introduction to Modal Logic*. Routledge, New York, 1996.

[Hem96] E. Hemaspaandra. The price of universality. *Notre Dame Journal of Formal Logic*, 37(2):174–203, Spring 1996.

[HF89] J. Y. Halpern and R. Fagin. Modelling knowledge and action in distributed systems. *Distributed Computing*, 3(4):159–179, 1989.

[Hin57] J. Hintikka. Quantifiers in deontic logic. *Societas Scientiarumi Fennica, Commentationes Humanarum Literarum*, 23(4):1–23, 1957.

[Hin62] J. Hintikka. *Knowledge and Belief, an introduction to the logic of the two notions*. Cornell University Press, Ithaca (NY) and London, 1962.

[HJF95] J. Huang, N. R. Jennings, and J. Fox. An agent-based approach to health care management. *Applied Artificial Intelligence: An International Journal*, 9(4):401–420, 1995.

[HLM96] W. van der Hoek, B. Van Linder, and J.-J. Meyer. Group knowledge isn't always distributed (neither is always implicit). In *Preprints of the 2nd Conference on Logic and the Foundations of the Theory of Games and Decisions (LOFT 2)*, Turin, December 1996.

[HM89] W. van der Hoek and J. J. Ch. Meyer. Possible logics for belief. *Logique et Analyse*, 1989.

[HM90] J. Halpern and Y. Moses. Knowledge and common knowledge in a distributed environment. *Journal of the ACM*, 37(3):549–587, 1990. A preliminary version appeared in *Proc. 3rd ACM Symposium on Principles of Distributed Computing*, 1984.

[HM92a] J. Halpern and Y. Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54:319–379, 1992.

[HM92b] W. van der Hoek and J.-J. Ch. Meyer. Making some issues of implicit knowledge explicit. *International Journal of Foundations of Computer Science*, 3(2):193–223, 1992.

[HM97] W. van der Hoek and J.-J. Ch. Meyer. A complete epistemic logic for multiple agents: Combining distributed and common knowledge. In M.O.L. Bacharach, L.-A. Girard-Varet, P. Mongin, and H.S. Shin, editors, *Epistemic Logic and the Theory of Games and Decisions*, pages 35–68. Kluwer, 1997.

[HMT94] W. van der Hoek, J.-J. Ch. Meyer, and J. Treur. Formal semantics of temporal epistemic reflection. In *Logic Program Synthesis and Transformation — Meta-Programming in Logic, 4th International Workshops (LOPSTR'94 and META'94)*, volume 883 of *Lecture Notes in Computer Science*, pages 332–352, Pisa, 1994.

[HMV97] J. Halpern, R. van der Meyden, and M. Y. Vardi. Complete axiomatisations for reasoning about knowledge and time. Submitted, 1997.

[Hoe93] W. van der Hoek. Sytems for knowledge and belief. *Journal of Logic and Computation*, 3(2):173–195, April 1993.

[HR95] B. Hayes-Roth. An architecture for adaptive intelligent systems. *Artificial Intelligence*, 72(1-2):329–365, January 1995.

[HV86] J. Y. Halpern and M. Y. Vardi. The complexity of reasoning about knowledge and time. In *ACM Symposium on Theory of Computing (STOC '86)*, pages 304–315, Baltimore, USA, May 1986. ACM Press.

[HV89] J. Y. Halpern and M. Y. Vardi. The complexity of reasoning about knowledge and time 1: lower bounds. *Journal of Computer and System Sciences*, 38(1):195–237, 1989.

[HV91] J. Halpern and M. Vardi. *Model checking vs. theorem proving: a manifesto*, pages 151–176. Artificial Intelligence and Mathematical Theory of Computation. Academic Press, Inc, 1991.

[Jas91] J. Jaspars. Fused modal logic and inconsistent belief. In M. DeGlas and D. Gabbay, editors, *Proceedings of the 1st World Conference on the Fundamentals of Artificial Intelligence*, pages 267–276, Paris, France, July 1991. Angkor.

[Jas93] J. O. M. Jaspars. Logical omniscience and inconsistent beliefs. In M. de Rijke, editor, *Diamonds and Defaults*, pages 129–146. Kluwer Academic Publishers, 1993.

[JCL95] N. R. Jennings, J. M. Corera, and I. Laresgotti. Developing industrial multi–agent systems. In V. Lesser, editor, *Proceedings of the First International Conference on Multi–Agent Systems*, pages 423–431, San Francisco, CA, 1995. MIT Press. (invited speaker talk).

[Je98] N. R. Jennings and M. Wooldridge (eds). *Agent Technology: Foundations, Applications, and Markets*. Springer-Verlag, March 1998.

[JW92] N. R. Jennings and T. Wittig. ARCHON: Theory and practice. In N. M. Avouris and L. Gasser, editors, *Distributed Artificial Intelligence: Theory and Praxis*, pages 179–195. Kluwer Academic Press, 1992.

[JW95] N. R. Jennings and M. Wooldridge. Applying agent technology. *International Journal of Applied Artificial Intelligence*, 9(4):351–361, 1995.

[JW98] N. R. Jennings and M. Wooldridge. Application of intelligent agents. In N. R. Jennings and M. Wooldridge, editors, *Agent Technology: Foundations, Applications, and Markets*. Springer-Verlag, March 1998.

[Kan57] S. Kanger. *Provability in logic*. Number 1 in Stockholm Studies in Philosophy. Almqvist & Wiksell, Stockholm, 1957.

[Kap66] D. Kaplan. Review of "A semantical analysis of modal logic I: normal modal propositional calculi". *Journal of Symbolic Logic*, 31:120–122, 1966.

[KG97] D. Kinny and M. Georgeff. Modelling and design of multi-agent systems. In J. P. Müller, M. J. Wooldridge, and N. R. Jennings, editors, *Proceedings of the ECAI'96 Workshop on Agent Theories, Architectures, and Languages: Intelligent Agents III*, volume 1193 of *LNAI*, pages 1–20, Berlin, August12–13 1997. Springer.

[KL88] S. Kraus and D. J. Lehmann. Knowledge, belief, and time. *Theoretical Computer Science*, 58:155–174, 1988.

[Kon86] K. Konolige. *A Deduction Model of Belief*. Brown University Press, 1986.

[KP81] D. Kozen and R. Parikh. An elementary proof of completness for PDL. *Theoretical Computer Science*, 14:113–118, 1981.

[Kra91] M. Kracht. *Internal Definability and Completeness in Modal Logic*. PhD thesis, FU Berlin, 1991.

[Kra93] M. Kracht. How completeness and correspondence theory got married. In M. de Rijke, editor, *Diamonds and Defaults*, pages 175–214. Kluwer Academic Publishers, Dordrecht, 1993.

[Kra95] M. Kracht. Highway to the danger zone. *Journal of Logic and Computation*, 5(1):93–109, February 1995.

[Kri59] S. A. Kripke. Semantic analysis of modal logic (abstract). *Journal of Symbolic Logic*, 24:323–324, 1959.

[Kri63] S. Kripke. A semantical analysis of modal logic I: normal modal propositional calculi. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 9:67–96, 1963. Announced in *Journal of Symbolic Logic*, **24**, 1959, p. 323.

[KW91] M. Kracht and F. Wolter. Properties of independently axiomatizable bimodal logics. *The Journal of Symbolic Logic*, 56(4), 1991.

[Lan84] S. Lang. *Algebra*. Addison-Wesley, Reading, 2 edition, 1984.

[LC96] A. Lomuscio and M. Colombetti. QLB: a quantified logic for belief. In J. P. Müller, M. J. Wooldridge, and N. R. Jennings, editors, *Intelligent Agents III — Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages (ATAL-96)*, volume 1193 of *Lecture Notes in Artificial Intelligence*. Springer-Verlag, Heidelberg, 1996.

[Ld95] M. Luck and M. d'Inverno. A formal framework for agency and autonomy. In Victor Lesser, editor, *Proceedings of the First International Conference on Multi–Agent Systems*, pages 254–268, San Francisco, CA, 1995. MIT Press.

[Leh84] D. J. Lehmann. Knowledge, common knowledge and related puzzles. In *Proceedings of the Third Annual ACM Symposium on Principles of Distributed Computing*, pages 62–67, Vancouver, B.C., Canada, 27–29 August 1984.

[Lem77] E. J. Lemmon. *The "Lemmon Notes": An Introduction to Modal Logic*. Basil Blackwell, Oxford, 1977. Written in collaboration with D. Scott; edited by K. Segerberg. American Philosophical Quarterly Monograph Series. Monograph No. 11.

[Len78] W. Lenzen. *Recent work in epistemic logic*, volume 30 of *Acta Philosophica Fennica*. North-Holland, Amsterdam, 1978.

[Lev84] H. Levesque. A logic of implicit and explicit belief. In *Proceedings of the 6th National Conference of the AAAI*, pages 198–202. Morgan Kaufman, Palo Alto, 1984.

[Lew18] C. I. Lewis. *A Survey of Symbolic Logic*. University of California Press, Berkeley and Los Angeles, 1918.

[Lew69] D. Lewis. *Convention, A Philosophical Study*. Harvard University Press, Cambridge, MA, 1969.

[LHR99] A. Lomuscio, W. van der Hoek, and M. Ryan. A spectrum of modes of knowledge sharing between agents: completeness, decidability and complexity. In preparation, 1999.

[LL59] C. I. Lewis and C. H. Langford. *Symbolic Logic*. Dover Publications, New York, 2 edition, 1959.

[LM94] G. Lakemeyer and S. Meyer. Enhancing the power of a decidable first-order reasoner. In E. S. P. Torasso and J. Doyle, editors, *Proceedings of the 4th International Conference on Principles of Knowledge Representation and Reasoning*, pages 403–414. Morgan Kaufmann, May 1994.

[LMR99] A. Lomuscio, R. van der Meyden, and M. Ryan. Hypercubes: Agents only broadcasting. Submitted to ACM Transactions on Computational Logic (ToCL), 1999.

[Lom95] A. Lomuscio. QLB: Una logica predicativa delle credenze – sintassi, semantica e una procedura di prova automatica dei teoremi. Master's thesis, Politecnico di Milano, Milan, December 1995.

[Lom97] A. Lomuscio. Logic engineering for modelling MAS. Technical report, School of Computer Science, University of Birmingham, March 1997. Thesis Proposal.

[Lom99] A. Lomuscio. Combining logics for multi-agent system theories. In preparation, 1999.

[LR98a] A. Lomuscio and M. Ryan. Ideal agents sharing (some!) information. Technical Report CSR-98-1, School of Computer Science, University of Birmingham, UK, January 1998. ftp://ftp.cs.bham.ac.uk/pub/tech-reports/1998/CSR-98-01.ps.gz.

[LR98b] A. Lomuscio and M. Ryan. Ideal agents sharing (some!) knowledge. In H. Prade, editor, *Proceedings of Proceedings of the 13th European Conference on Artificial Intelligence*, pages 557–561, August 1998.

[LR98c] A. Lomuscio and M. Ryan. Model checking and refinement for $S5_n$. In *Proceedings of the ECAI98-workshop on Practical Reasoning and Rationality (PRR98)*, Brighton, August 1998.

[LR98d] A. Lomuscio and M. Ryan. On the relation between interpreted systems and Kripke models. In M. Pagnucco, W. R. Wobcke, and C. Zhang, editors, *Agent* *and Multi-Agent Systems - Proceedings of the AI97 Workshop on the theoretical and practical foundations of intelligent agents and agent-oriented systems*, volume 1441 of *Lecture Notes in Artificial Intelligence*. Springer Verlag, Berlin, May 1998.

[LR99a] A. Lomuscio and M. Ryan. An algorithmic approach to knowledge evolution. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing (AIEDAM)*, 13(2), 1999.

[LR99b] A. Lomuscio and M. Ryan. A spectrum of modes of knowledge sharing between agents. In M. J. Wooldridge and N. R. Jennings, editors, *Intelligent Agents VI — Proceedings of the Sixth International Workshop on Agent Theories, Architectures, and Languages (ATAL-99)*, Lecture Notes in Artificial Intelligence. Springer-Verlag, Heidelberg, 1999. To appear.

[MA99] M. Marx and C. Areces. Failure of interpolation in combined modal logics. Research Report PP-1999-03, Institute for Logic, Language and Computation (ILLC), Amsterdam, The Netherlands, January 1999.

[Mae94] P. Maes. Agents that reduce work and information overload. *Communications of the ACM*, 37(7):31–40, July 1994.

[Mae95] P. Maes. Artificial life meets enfertainment: Lifelike autonomous agents. *Communications of the ACM*, 38(11):108–114, November 1995.

[Mak66] D. Makinson. On some completeness theorems in modal logic. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 12:379–384, 1966.

[McC78] J. McCarthy. Formalization of two puzzles involving knowledge. Unpublished manuscript, Computer Science Department., Stanford University, 1978.

[McC79] J. McCarthy. Ascribing mental qualities to machines. In M. Ringle, editor, *Philosophical Perspectives in Artificial Intelligence*. Harvester Press, 1979.

[McK41] J.C McKinsey. A solution of the decision problem for the Lewis systems S2 and S4 with an application to topology. *The Journal of Symbolic Logic*, 6:117–134, 1941.

[Men64] E. Mendelson. *Introduction to Mathematical Logic*. Van Nostrand Reinhold, New York, 1964.

[Mey88] B. Meyer. *Object-Oriented Software Construction*. Prentice Hall International - Series in Computer Science, 1988.

[Mey94] R. van der Meyden. Axioms for knowledge and time in distributed systems with perfect recall. In *Proceedings, Ninth Annual IEEE Symposium on Logic in Computer Science*, pages 448–457, Paris, France, 4–7 July 1994. IEEE Computer Society Press.

[Mey96] R. van der Meyden. Knowledge based programs: On the complexity of perfect recall in finite environments. In Y. Shoham, editor, *Proceedings of the Sixth Conference on Theoretical Aspects of Rationality and Knowledge*, pages 31–50, San Francisco, March17–20 1996. Morgan Kaufmann Publishers.

[Mey98] R. van der Meyden. On homogeneous broadcasting systems. Private communication, December 1998.

[MH95] J.-J. Ch. Meyer and W. van der Hoek. *Epistemic Logic for AI and Computer Science*, volume 41 of *Cambridge Tracts in Theoretical Computer Science*. Cambdridge University Press, 1995.

[Mil81] P. Milgrom. An axiomatic characterization of common knowledge. *Econometrica*, 49(1):219–222, 1981.

[Mon60] R. Montague. Logical necessity, physical necessity, ethics, and quantifiers. *Inquiry*, 4:259–269, 1960.

[Mon63] R. Montague. Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizations. *Acta Philosophica Fennica*, 16:153–167, 1963.

[MP91] Z. Manna and A. Pnueli. *The Temporal Logic of Reactive and Concurrent Systems: Specification*. Springer-Verlag, 1991.

[MP92] Z. Manna and A. Pnueli. *The temporal logic of reactive and concurrent systems*, volume 1. Springer-Verlag, Berlin/New York, 1992.

[MP95] Z. Manna and A. Pnueli. *Temporal Verification of Reactive Systems: Safety*. Springer-Verlag, 1995.

[MS85] T. Maibaum and M. Sadler. Axiomatising specification theory. In H.-J. Kreowski, editor, *Recent Trends in Data Type Specification*, number 116 in Informatik-Facberichte, pages 171–177. Springer-Verlag, New York, N.Y., 1985.

[MS93] Y. Moses and Y. Shoham. Belief as defeasible knowledge. *Artificial Intelligence*, 64(2):299–321, December 1993.

[MSHI78] J. McCarthy, J. M. Sato, T. Hayashi, and S. Igarishi. On the model theory of knowledge. Technical Report STAN-CS-79-725, Stanford University, 1978.

[Mye79] G. J. Myers. *The Art of Software Testing*. Wiley - Interscience, New York, 1979.

[Ono77] H. Ono. On some intuitionistic modal logics. *Publ. Inst. Math. Sci. Kyoto*, 13:687–722, 1977.

[Par96] H. Van Dyke Parunak. Applications of DAI in industry. In G. O'Hare and N. Jennings, editors, *Foundations of Distributed Artificial Intelligence*, chapter 4. John Wiley and Sons, 1996.

[Par98] H. Van Dyke Parunak. What can agents do in industry, and why? An overview of industrially-oriented R&D at CEC. *Lecture Notes in Computer Science*, 1435, 1998.

[Pau94] L. C. Paulson. Isabelle: A generic theorem prover. *Lecture Notes in Computer Science*, 828:xvii + 321, 1994.

[Pfa87] J. Pfalzgraf. A note on simplices as geometric configurations. *Archives of Mathematics*, 49:134–140, 1987.

[Pop94] S. Popkorn. *First Steps in Modal Logic*. Cambridge University Press: Cambridge, England, 1994.

[Pra79] V. R. Pratt. Models of program logics. In *20th Annual Symposium on Foundations of Computer Science*, pages 115–122, Long Beach, Ca., USA, October 1979. IEEE Computer Society Press.

[Pra95] V. Pratt. Anatomy of the Pentium Bug. In P. D. Mosses, M. Nielsen, and M. I. Schwartzbach, editors, *TAPSOFT'95: Theory and Practice of Software Development*, number 915 in Lecture Notes in Computer Science, pages 97–107. Springer Verlag, May 1995.

[Pre96] H. Prendinger. Modal logics for modelling actions and agents. *Lecture Notes in Computer Science*, 1085:525–536, 1996.

[Pri62] A. N. Prior. Possible worlds (idea attributed to P. T. Geach). *Philosophical Quarterly*, 12:36–43, 1962.

[PSJ98] S. Parsons, C. Sierra, and N. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3):261–292, June 1998.

[Rao96] A. S. Rao. Decision procedures for propositional linear-time Belief-Desire-Intention logics. In M. Wooldridge, J. P. Müller, and M. Tambe, editors, *Intelligent Agents II (LNAI 1037)*, pages 33–48. Springer-Verlag: Heidelberg, Germany, 1996.

[RG91] A. Rao and M. Georgeff. Deliberation and its role in the formation of intentions. In B. D. D'Ambrosio, P. Smets, and P. Bonissone, editors, *Proceedings of the 7th Conference on Uncertainty in Artificial Intelligence*, pages 300–307, San Mateo, CA, USA, July 1991. Morgan Kaufmann Publishers.

[RG98] A. S. Rao and M. P. Georgeff. Decision procedures for BDI logics. *Journal of Logic and Computation*, 8(3):293–343, June 1998.

[RN95] S. J. Russell and P. Norvig. *Artificial Intelligence. A Modern Approach*. Prentice-Hall, Englewood Cliffs, NJ, 1995.

[Sah75] H. Sahlqvist. Completeness and correspondence in the first and second order semantics for modal logic. In S. Kanger, editor, *Proceedings 3rd Scandinavian Logic Symposium, Uppsala, Sweden, 9–11 Apr 1973*, volume 82 of *Studies in Logic and the Foundations of Mathematics*, pages 110–143. North-Holland, Amsterdam, 1975.

[Sch98] M. Schroeder. *Autonomous, Model-based Diagnosis Agents*. Kluwer Academic Publisher, 1998.

[Sco70] D. Scott. Advice on modal logic. In K. Lambert, editor, *Philosophical Problems in Logic*, pages 143–173. Reidel, Dordrecht, Netherlands, 1970.

[Seg68] K. Segerberg. *Results in Nonclassical Logic*. Berlingska Boktryckeriet, Lund, 1968.

[Sho93] Y. Shoham. Agent-oriented programming. *Artificial Intelligence*, 60:51–92, March 1993.

[Sur98] T. Surendonk. *Canonicity for Intensional Logics*. PhD thesis, The Automated Reasoning Project, Australian National University, Canberra, 1998.

[Thi92] E. Thijsse. *On Partial Logic and Knowledge Representation*. Eburon, Delft, Netherlands, 1992. Ph.D. thesis, Tilburg University.

[Tho80a] R. Thomason. A note on syntactical treatments of modality. *Synthese*, 44:391–395, 1980.

[Tho80b] S. K. Thomason. Independent propositional modal logics. *Studia Logica*, 1980.

[Tur90] R. Turner. *Truth and Modality for Knowledge Representation*. Pitman Publishing: London, 1990.

[Voo92] F. Voorbraak. Generalized Kripke models for epistemic logic. In Y. Moses, editor, *Proceedings of the 4th Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 214–228, Monterey, CA, March 1992. Morgan Kaufmann.

[Wan92] X. Wang. The McKinsey axiom is not compact. *The Journal of Symbolic Logic*, 57(4):1230–1238, December 1992.

[Wel93] M. Wellman. A market-oriented programming environment and its application to distributed multicommodity flow problems. *Journal of Artificial Intelligence Research*, 1:1–23, 1993.

[WG96] P. Wavish and M. Graham. A situated action approach to impleminting characters in computer games. *Journal of Applied artificial intelligence*, 10(1):53–74, 1996.

[Wij90] D. Wijesekera. Constructive modal logics I. *Annals of Pure and Applied Logic*, 50(3):271–301, 14 December 1990.

[WJ95] M. Wooldridge and N. R. Jennings. Intelligent agents: theory and practice. *Knowledge Engineering Review*, 2(10):115–152, 1995.

[WJ98] M. Wooldridge and N. R. Jennings. Pitfalls of agent-oriented development. In Katia P. Sycara and Michael Wooldridge, editors, *Proceedings of the 2nd International Conference on Autonomous Agents (AGENTS-98)*, pages 385–391, New York, May 1998. ACM Press.

[Wol96] F. Wolter. A counterexample in tense logic. *Notre Dame Journal of Formal Logic*, 37(2):167–173, Spring 1996.

[Woo92] M. Wooldridge. *The Logical Modelling of Computational Multi-Agent Systems*. PhD thesis, UMIST, Manchester, UK, 1992.

[Woo97] M. Wooldridge. Agent-based software engineering. *IEE Proceedings Software Engineering*, 144(1):26–37, 1997.