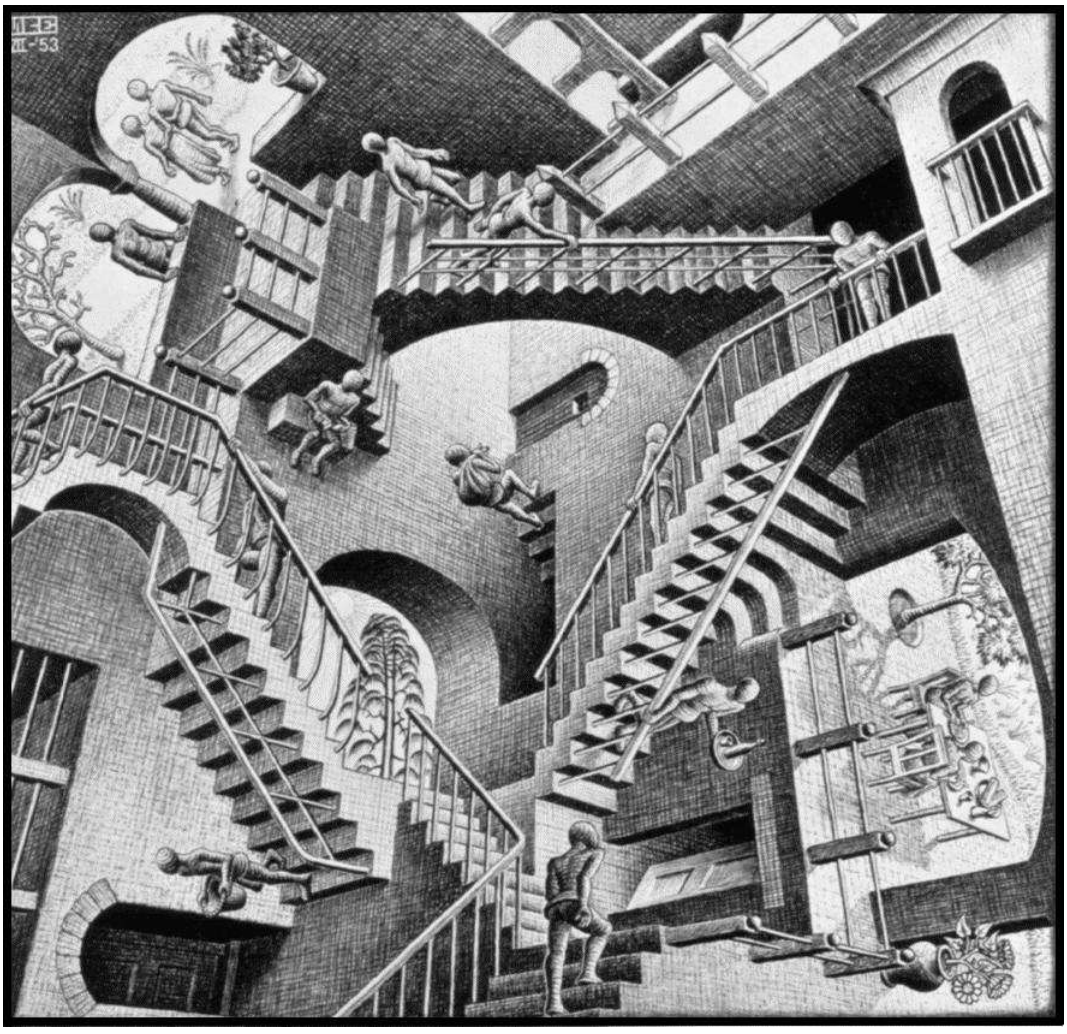
Perspectives on belief and change

Guillaume Aucher

Perspectives on Belief and Change



GUILLAUME AUCHER

University of Otago
Department of Computer Science
.
Université Toulouse III - Paul Sabatier
Ecole Doctorale Mathématiques, Informatique et télécommunications de Toulouse
Institut de Recherche en Informatique de Toulouse

Perspectives on Belief and Change

Thesis presented and defended by

GUILLAUME AUCHER

at Toulouse on the 9th of July 2008 to obtain the degrees of

Doctor of Philosophy of the University of Otago
Docteur de l'Université de Toulouse

Speciality: Computer Science

SUPERVISORS:

Hans van Ditmarsch Andreas Herzig

REVIEWERS:

Wiebe van der Hoek Willem Labuschagne
Pierre Marquis

EXAMINERS:

Johan van Benthem Claudette Cayrol
Jérôme Lang Karl Schlechta

À ma mère

Acknowledgement

First of all, I owe special thanks to my supervisors Hans van Ditmarsch and Andreas Herzig for giving me the opportunity to do this PhD and for arranging this joint supervision between France and New Zealand (both administratively and scientifically). Their high competence combined with their human qualities provided me a very pleasant and valuable supervision. They let me a large amount of liberty and autonomy while at the same time always keeping their door open for discussions. I thank them for their stimulating and numerous advices and comments, and also for their constant human support during these three years of adventure.

Then I want to thank Wiebe van der Hoek, Willem Labuschagne and Pierre Marquis for accepting to review my thesis and for their detailed reviews. I also thank Johan van Benthem, Claudette Cayrol, Jérôme Lang and Karl Schlechta for the honor of having them in my jury and for their questions and comments during my defense.

I thank the members of the computer science department of the university of Otago and the members of the LILaC and RPDMP teams of the Institut de Recherche en Informatique de Toulouse for their comments and questions during my various seminars in Otago and Toulouse. In particular, I thank Jérôme Lang and Philippe Balbiani for their feedback and various comments on my work, collaborations and for professional advices. I thank Didier Dubois for endless and yet very interesting discussions. On the other side of the world, I thank Willem Labuschagne for inspiring discussions (and with whom I would be delighted, too, to have another chat over a cup of tea). I thank Geoff Wyvill for a very interesting email correspondence. Outside Otago and Toulouse, my work also benefited, consciously or not, from discussions and comments of other people. For that, I thank: Jelle Gerbrandy, Fenrong Liu, Paul Harrenstein, Eric Pacuit, Olivier Roy and Yanjing Wang; the anonymous referees of: the Journal of Applied Non-Classical Logic, ECSQARU'07, AAMAS'08, the workshop on dynamics of knowledge and belief; and the audiences of: the PALMYR'06 workshops in Paris and Amsterdam, the Dagstuhl workshops on belief change, the 'Dynamic Logic Montréal' workshop, the 'belief revision and dynamic logic' workshop in ESSLLI and the workshop on dynamics of knowledge and belief in Osnabrueck. I thank Olivier Roy and Isidora Stojanovic for inviting me to PALMYR; Patrick Girard and Mathieu Marion for inviting me to the workshop 'Dynamic Logic Montréal'; Giacomo Bonano, James Delgrande, Jérôme Lang and Hans Rott for inviting me to the Dagstuhl workshops.

I also want to thank the university of Otago for funding my research and the University of Toulouse and the French Ministry of Foreign Affairs for funding my various travels. I thank Gaëlle Fontaine for accommodation during my stay in Amsterdam for PALMYR'06 and Patrick Girard's aunt for accommodation during my stay in Montréal for the workshop 'Dynamic logic Montréal'.

There are also other friends and colleagues who, even if they were not in direct contact with my work, somehow contributed by their presence to the achievement of this thesis and with whom I had good times. I thank (among others): John Askin, Sihem Belabbes, Meghyn Bienvenu, Fahima Cheick, Yannick Chevalier, Marwa Elhouri, Mounira Kourjieh, Olivier Gasquet, Benoit Gaudou, Faisal Hassan, Tiago de Lima, Dominique Longin (my 'latex guru', thanks to whom the thesis looks so nice :-), Emiliano Lorini, Jérôme Louradour, Vanessa Miranville, Philippe Muller, Georgina Pickerell, Marie de Roquemaurel, François Schwarzentruber, Bilal Said, Eduardo Sanchez, Nicolas Troquard, Ivan Varzinczak, and my hiking mates in Otago. I thank the members of the LILaC and RPDMP teams for providing me, lunch after lunch, a very nice environment at IRIT. Finally, I thank my family for its constant support during these three unforgettable years.

July 2008.

Contents

1 Introduction **1**

2 Epistemic logic **5**
2.1 Introduction 5
2.2 State of the art 6
2.2.1 Semantics 6
2.2.2 Axiomatization 10
2.2.3 Common belief 12
2.2.4 An epistemic logic toolkit 13
2.3 A new approach: internal versus external perspectives of the world 16
2.3.1 Intuitions 16
2.3.2 A semantics for the internal approach 18
2.3.3 Some connections between the internal and the external approach . . 25
2.3.4 Axiomatization of the internal semantics 28
2.4 Conclusion 32

3 Dynamic epistemic logic **35**
3.1 Introduction 35
3.2 The BMS system 36
3.2.1 State of the art 36
3.2.2 On seriality preservation 40
3.3 Dynamizing the internal approach 47
3.3.1 Multi-agent possible event and internal event model 47
3.3.2 The update product 49
3.4 Some connections between the external and the internal approach 54
3.4.1 From (external) event model to internal event model 54
3.4.2 Preservation of the update product 54
3.5 Concluding remarks 56

4 Internal approach: the case of private announcements **57**
4.1 Introduction . . . 57
4.2 Generalizing AGM to the multi-agent case . . . 58
4.2.1 Expansion . . . 58
4.2.2 Revision . . . 64
4.3 Multi-agent rationality postulates . . . 73
4.3.1 On the kind of information a formula is about . . . 73
4.3.2 Some rationality postulates specific to our multi-agent approach . . . 75
4.4 A revision operation . . . 77
4.4.1 Mathematical preliminaries . . . 77
4.4.2 Definition of the revision operation . . . 79
4.4.3 Properties of the revision operation . . . 81
4.4.4 Concrete example . . . 86
4.5 Conclusion . . . 90

5 External approach: a general formalism **93**
5.1 Introduction . . . 93
5.2 Mathematical preliminaries . . . 94
5.3 Starting with beliefs as probabilities . . . 98
5.3.1 The static part . . . 98
5.3.2 The dynamic part . . . 102
5.3.3 The update mechanism . . . 105
5.4 Adding knowledge . . . 110
5.4.1 State of the art . . . 110
5.4.2 Our proposal . . . 113
5.5 Adding agents . . . 120
5.5.1 The static part . . . 121
5.5.2 The dynamic part . . . 121
5.5.3 The update mechanism . . . 122
5.6 Comparisons . . . 123
5.7 Conclusion . . . 125

6 Exploring the power of converse events **127**
6.1 Introduction . . . 127
6.2 EDL: Epistemic Dynamic Logic with converse . . . 128
6.2.1 The language $\mathcal{L}_{\mathsf{EDL}}$ of EDL . . . 128
6.2.2 Semantics of EDL . . . 128
6.2.3 Completeness . . . 131
6.3 From BMS to EDL . . . 132
6.4 Conclusion and related work . . . 138

7 Conclusion and further research **141**

Abstract **144**

Résumé **146**

Bibliography **149**

Index **158**

Chapter 1

Introduction

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

The way a fixed situation is perceived by agents can be represented by statements about the agents' beliefs: for example 'agent A believes that the door of the room is open' or 'agent A believes that her colleague is busy this afternoon'. '*Logical* systems' means that agents can reason about the situation and their beliefs about it: if agent A believes that her colleague is busy this afternoon then agent A infers that he will not visit her this afternoon. We moreover often assume that our situations involve several agents which interact between each other. So these agents have beliefs about the situation (such as 'the door is open') but also about the other agents' beliefs: for example agent A might believe that agent B believes that the door is open. These kinds of beliefs are called higher-order beliefs. *Epistemic logic* [Hintikka, 1962; Fagin *et al.*, 1995; Meyer and van der Hoek, 1995], the logic of belief and knowledge, can capture all these phenomena and will be our main starting point to model such fixed ('static') situations. Uncertainty can of course be expressed by beliefs and knowledge: for example agent A being uncertain whether her colleague is busy this afternoon can be expressed by 'agent A does not *know* whether her colleague is busy this afternoon'. But we sometimes need to enrich and refine the representation of uncertainty: for example, even if agent A does not know whether her colleague is busy this afternoon, she might consider it more probable that he is actually busy. So other logics have been developed to deal more adequately with the representation of uncertainty, such as probabilistic logic, fuzzy logic or possibilistic logic, and we will refer to some of them in this thesis (see [Halpern, 2003] for a survey on reasoning about uncertainty).

But things become more complex when we introduce events and change in the picture. Issues arise even if we assume that there is a single agent. Indeed, if the incoming information conveyed by the event is coherent with the agent's beliefs then the agent can

just add it to her beliefs. But if the incoming information contradicts the agent's beliefs then the agent has somehow to revise her beliefs, and as it turns out there is no obvious way to decide what should be her resulting beliefs. Solving this problem was the goal of the logic-based *belief revision theory* developed by Alchourrón, Gärdenfors and Makinson (to which we will refer by the term AGM) [Alchourrón *et al.*, 1985; Gärdenfors, 1988; Gärdenfors and Rott, 1995]. Their idea is to introduce 'rationality postulates' that specify which belief revision operations can be considered as being 'rational' or reasonable, and then to propose specific revision operations that fulfill these postulates. However, AGM does not consider situations where the agent might also have some uncertainty about the incoming information: for example agent A might be uncertain due to some noise whether her colleague told her that he would visit her on Tuesday or on Thursday. In this thesis we also investigate this kind of phenomenon. Things are even more complex in a multi-agent setting because the way agents update their beliefs depends not only on their beliefs about the event itself but also on their beliefs about the way the other agents perceived the event (and so about the other agents' beliefs about the event). For example, during a private announcement of a piece of information to agent A the beliefs of the other agents actually do not change because they believe nothing is actually happening; but during a public announcement all the agents' beliefs might change because they all believe that an announcement has been made. Such kind of subtleties have been dealt with in a field called *dynamic epistemic logic* [Gerbrandy and Groeneveld, 1997; Baltag *et al.*, 1998; van Ditmarsch *et al.*, 2007b]. The idea is to represent by an event model how the event is perceived by the agents and then to define a formal update mechanism that specifies how the agents update their beliefs according to this event model and their previous representation of the situation. Finally, the issues concerning belief revision that we raised in the single agent case are still present in the multi-agent case.

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

Studying and proposing logical models for belief change and belief representation has applications in several areas. First in artificial intelligence, where machines or robots need to have a formal representation of the surrounding world (which might involve other agents), and formal mechanisms to update this representation when they receive incoming information. Such formalisms are crucial if we want to design autonomous agents, able to act autonomously in the real world or in a virtual world (such as on the internet). Indeed, the representation of the surrounding world is essential for a robot in order to reason about the world, plan actions in order to achieve goals…and it must be able to update and revise its representation of the world itself in order to cope autonomously with unexpected events. Second in game theory (and consequently in economics), where we need to model games involving several agents (players) having beliefs about the game and about the other agents' beliefs (such as agent A believes that agent B has the ace of spade, or agent A believes that

agent B believes that agent A has the ace of heart...), and how they update their representation of the game when events (such as showing privately a card or putting a card on the table) occur. Third in cognitive psychology, where we need to model as accurately as possible epistemic state of human agents and the dynamics of belief and knowledge in order to explain and describe cognitive processes.

The thesis is organized as follows. In Chapter 2, we first recall epistemic logic. Then we observe that representing an epistemic situation involving several agents depends very much on the modeling point of view one takes. For example, in a poker game the representation of the game will be different depending on whether the modeler is a poker player playing in the game or the card dealer who knows exactly what the players' cards are. In this thesis, we will carefully distinguish these different modeling approaches and the different kinds of formalisms they give rise to. In fact, the interpretation of a formalism relies quite a lot on the nature of these modeling points of view. Classically, in epistemic logic, the models built are supposed to be correct and represent the situation from an external and objective point of view. We call this modeling approach the perfect external approach. In Chapter 2, we study the modeling point of view of a particular modeler-agent involved in the situation with other agents (and so having a possibly erroneous perception of the situation). We call this modeling approach the internal approach. We propose a logical formalism based on epistemic logic that this agent uses to represent 'for herself' the surrounding world. We then set some formal connections between the internal approach and the (perfect) external approach. Finally we axiomatize our logical formalism and show that the resulting logic is decidable.

In Chapter 3, we first recall dynamic epistemic logic as viewed by Baltag, Moss and Solecki (to which we will refer by the term BMS). Then we study in which case seriality of the accessibility relations of epistemic models is preserved during an update, first for the full updated model and then for generated submodels of the full updated model. Finally, observing that the BMS formalism follows the (perfect) external approach, we propose an internal version of it, just as we proposed an internal version of epistemic logic in Chapter 2.

In Chapter 4, we still follow the internal approach and study the particular case where the event is a private announcement. We first show, thanks to our study in Chapter 3, that in a multi-agent setting, expanding in the AGM style corresponds to performing a private announcement in the BMS style. This indicates that generalizing AGM belief revision theory to a multi-agent setting amounts to study private announcement. We then generalize the AGM representation theorems to the multi-agent case. Afterwards, in the spirit of the AGM approach, we go beyond the AGM postulates and investigate multi-agent rationality postulates specific to our multi-agent setting inspired from the fact that the kind of phenomenon we study is private announcement. Finally we provide an example of revision operation that we apply to a concrete example.

In Chapter 5, we follow the (perfect) external approach and enrich the BMS formalism with probabilities. This enables us to provide a fined-grained account of how human agents interpret events involving uncertainty and how they revise their beliefs. Afterwards, we review different principles for the notion of knowledge that have been proposed in the literature and show how some principles that we argue to be reasonable ones can all be captured

in our rich and expressive formalism. Finally, we extend our general formalism to a multi-agent setting.

In Chapter 6, we still follow the (perfect) external approach and enrich our dynamic epistemic language with converse events. This language is interpreted on structures with accessibility relations for both beliefs and events, unlike the BMS formalism where events and beliefs are not on the same formal level. Then we propose principles relating events and beliefs and provide a complete characterization, which yields a new logic EDL. Finally, we show that BMS can be translated into our new logic EDL thanks to the converse operator: this device enables us to translate the structure of the event model directly within a particular axiomatization of EDL, without having to refer to a particular event model in the language (as done in BMS).

In Chapter 7 we summarize our results and give an overview of remaining technical issues and some desiderata for future directions of research.

Parts of this thesis are based on publication, but we emphasize that they have been entirely rewritten in order to make this thesis an integrated whole. Sections 4.2.2 and 4.3 of Chapter 4 are based on [Aucher, 12 16 May 2008]. Sections 5.2, 5.3 and 5.5 of Chapter 5 are based on [Aucher, 2007]. Chapter 6 is based on [Aucher and Herzig, 2007].

Chapter 2

Epistemic logic

2.1 Introduction

Epistemic logic is a modal logic [Blackburn *et al.*, 2001] that is concerned with the logical study of the notions of knowledge and belief. It is then concerned with understanding the process of *reasoning* about knowledge and belief. As epistemology, it stems from the Greek word $\epsilon\pi\iota\sigma\tau\eta\mu\eta$ or 'episteme' meaning knowledge. But epistemology is more concerned with analyzing the very *nature* of knowledge (addressing questions such as "What is the definition of knowledge?" or "How is knowledge acquired?"). In fact, epistemic logic grew out of epistemology in the middle ages thanks to the efforts of Burley and Ockham [Boh, 1993]. But the formal work, based on modal logic, that inaugurated contemporary research into epistemic logic dates back only to 1962 and is due to Hintikka [Hintikka, 1962]. It then sparked in the 1960's discussions about the inherent properties of knowledge and belief and many axioms for these notions were proposed and discussed [Lenzen, 1978]. More recently, these kind of philosophical theories were taken up by researchers in economics [Battigalli and Bonanno, 1999], artificial intelligence and theoretical computer science [Fagin *et al.*, 1995] [Meyer and van der Hoek, 1995] where reasoning about knowledge is a central topic. Due to the new setting in which epistemic logic was used, new perspectives and new features such as computability issues were then added to the agenda of epistemic logic.

In this chapter, we first give an outline of contemporary epistemic logic. Then we propose a new approach to epistemic logic by stressing the importance of choosing a particular modeling point of view. This leads us to distinguish two main modeling approaches that we call the internal and the external approach. We then focus on the internal approach and study its relationship with the external one.

2.2 State of the art

In this section we briefly describe the modal approach to epistemic logic initiated by Hintikka, focusing on the multi-agent case. We first define the semantics of epistemic logic based on the notion of epistemic model in Section 2.2.1. We axiomatize it in Section 2.2.2 and discuss some axioms. Then we concentrate on the notion of common belief in Section 2.2.3. Finally in Section 2.2.4, we recall results and notions from modal logic, such as bisimulation, which will be used throughout this thesis. We focus mainly on the notion of belief. The notion of knowledge and its relationship with belief will be tackled at more length in Section 5.4.1 of Chapter 5.

2.2.1 Semantics

First some important notations. Throughout the thesis Φ is a set of propositional letters and G is a finite set of agents. We assume that the number of agents N is bigger than 1.

As we said, epistemic logic is a modal logic. So what we call an epistemic model is just a particular kind of Kripke model as used in modal logic. The only difference is that instead of having a single accessibility relation we have a set of accessibility relations, one for each agent.

Definition 2.2.1 (Epistemic model)
An *epistemic model* M is a triple $M = (W, R, V)$ such that

- W is a non-empty set of possible worlds;
- $R : G \to 2^{W \times W}$ assigns an accessibility relation to each agent;
- $V : \Phi \to 2^W$ assigns a set of possible worlds to each propositional letter and is called a valuation.

If $M = (W, R, V)$ is an epistemic model, a pair (M, w_a) with $w_a \in W$ is called a *pointed epistemic model*. We also write $R_j = R(j)$ and $R_j(w) = \{w' \in W \mid wR_jw'\}$, and $w \in M$ for $w \in W$. □

Intuitively, a pointed epistemic model (M, w_a) represents from an external point of view how the actual world w_a is perceived by the agents G. The possible worlds W are the relevant worlds needed to define such a representation and the valuation V specifies which propositional facts (such as 'it is raining') are true in these worlds. Finally the accessibility relations R_j can model either the notion of knowledge or the notion of belief. We set $w' \in R_j(w)$ in case the world w' is compatible with agent j's belief (respectively knowledge) in world w.

Now inspiring ourselves from modal logic, we can define a language for epistemic models. The modal operator is just replaced either by a 'belief' or a 'knowledge' operator, one for each agent. As we said, we focus on the notion of belief, and we write B_j the belief operator; the notion of knowledge, whose operator is written K_j, will be studied in more depth in Section 5.4.1.

Definition 2.2.2 (Language $\mathcal{L}$)
The language $\mathcal{L}$ is defined as follows:

$$\mathcal{L} : \varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \varphi \mid B_j\varphi$$

where p ranges over Φ and j over G. Moreover, $\varphi \vee \varphi'$ is an abbreviation for $\neg(\neg\varphi \wedge \neg\varphi')$; $\varphi \rightarrow \varphi'$ is an abbreviation for $\neg\varphi \vee \varphi'$; $\hat{B}_j\varphi$ is an abbreviation for $\neg B_j \neg\varphi$; and $\bot$ is an abbreviation for $\neg\top$. □

Intuitively, $B_j p$ means that the agent j believes that the propositional fact p is true. But in fact, in this language we can express not only what the agents believe about the world but also what they believe about what the other agents believe about what the other agents believe, and so on. This is exemplified by formulas of the form $B_j B_i p$ or $B_j B_k B_i q \ldots$ These kinds of belief are called 'higher-order' beliefs. In fact we can quantify this nesting of belief operators thanks to the notion of *degree* of a formula.

Definition 2.2.3 (Degree of an epistemic formula)
The *degree* $deg(\varphi)$ of an epistemic formula φ is defined inductively as follows:

- $deg(p) = deg(\top) = 0$;
- $deg(\neg\varphi) = deg(\varphi)$; $deg(\varphi \wedge \psi) = max\{deg(\varphi), deg(\psi)\}$;
- $deg(B_j\varphi) = 1 + deg(\varphi)$.

□

Now we can give meaning to the formulas of this language by defining truth conditions for these formulas on the class of epistemic models.

Definition 2.2.4 (Truth conditions for $\mathcal{L}$)
Let $M = (W, R, V)$ be an epistemic model and $w \in W$. $M, w \models \varphi$ is defined inductively as follows:

$$\begin{array}{lll} M, w \models \top & & \\ M, w \models p & \text{iff} & w \in V(p) \\ M, w \models \neg\varphi & \text{iff} & \text{not } M, w \models \varphi \\ M, w \models \varphi \wedge \varphi' & \text{iff} & M, w \models \varphi \text{ and } M, w \models \varphi' \\ M, w \models B_j\varphi & \text{iff} & \text{for all } v \in R_j(w), M, v \models \varphi \end{array}$$

We write $M \models \varphi$ for $M, w \models \varphi$ for all $w \in M$. □

So the agent j believes φ in world w (formally $M, w \models B_j\varphi$) if φ is true in all the worlds that the agent j considers possible (in world w).

But note that the notion of belief might comply to some constraints (or axioms) such as $B_j\varphi \rightarrow B_j B_j\varphi$: if agent j believes something, she knows that she believes it. These constraints might affect the nature of the accessibility relations R_j which may then comply to some extra properties. So, we are now going to define some particular classes of epistemic models that all add some extra constraints on the accessibility relations R_j. We will see in the next section that these constraints are matched by particular axioms for the belief operator B_j.

Definition 2.2.5 (Properties of accessibility relations)
We list below a list of properties for the accessibility relations R_j that will be used in the sequel.

- *seriality*: for all w, $R_j(w) \neq \emptyset$;
- *transitivity*: for all w, w', w'', if $w' \in R_j(w)$ and $w'' \in R_j(w')$ then $w'' \in R_j(w)$;
- *euclidicity*: for all w, w', w'', if $w' \in R_j(w)$ and $w'' \in R_j(w)$ then $w' \in R_j(w'')$;
- *reflexivity*: for all w, $w \in R_j(w)$;
- *confluence*: for all w, w', w'', if $w' \in R_j(w)$ and $w'' \in R_j(w)$ then there is v such that $v \in R_j(w')$ and $v \in R_j(w'')$;
- *weakly connected*: for all w, w', w'', if ($w' \in R_j(w)$ and $w'' \in R_j(w)$) then ($w' \in R_j(w'')$ or $w' = w''$ or $w'' \in R_j(w')$;
- *.3.2*: for all w, w', w'', if $w'' \in R_j(w)$ and not $w \in R_j(w'')$ then $w' \in R_j(w)$ implies $w'' \in R_j(w')$;
- *.4*: for all w, w', w'', if ($w'' \in R_j(w)$ and $w \neq w''$) then ($w' \in R_j(w)$ implies $w'' \in R_j(w')$).

We list below classes of epistemic models that will be used in the sequel.

- K_G-models: no restriction;
- $\mathsf{KD45}_\mathsf{G}$-models: the accessibility relations are serial, transitive and euclidean;
- $\mathsf{S4}_\mathsf{G}$-models: the accessibility relations are reflexive and transitive;
- $\mathsf{S4.2}_\mathsf{G}$-models: the accessibility relations are reflexive, transitive and confluent;
- $\mathsf{S4.3}_\mathsf{G}$-models: the accessibility relations are reflexive, transitive and weakly connected;
- $\mathsf{S4.3.2}_\mathsf{G}$-models: the accessibility relations are reflexive, transitive and satisfy .3.2;
- $\mathsf{S4.4}_\mathsf{G}$-models: the accessibility relations are reflexive, transitive and satisfy .4;
- $\mathsf{S5}_\mathsf{G}$-models: the accessibility relations are reflexive and euclidean (and thus transitive).

□

Now we define the notions of satisfiability, validity and epistemic consequence with respect to a certain class of epistemic models.

Definition 2.2.6 (Satisfiability, validity and consequence)
Let $\mathsf{L} \in \{\mathsf{K}_\mathsf{G}, \mathsf{KD45}_\mathsf{G}, \mathsf{S4}_\mathsf{G}, \mathsf{S4.2}_\mathsf{G}, \mathsf{S4.3}_\mathsf{G}, \mathsf{S4.3.2}_\mathsf{G}, \mathsf{S4.4}_\mathsf{G}, \mathsf{S5}_\mathsf{G}\}$.

Let $\varphi \in \mathcal{L}$. We say that φ is L-*satisfiable* if there is an L-model M and $w \in M$ such that $M, w \models \varphi$. We say that φ is L-*valid*, written $\models_{\mathsf{L}} \varphi$, if for all L-models M and all $w \in M$, $M, w \models \varphi$.

More generally, if C is a class of epistemic models, we say that φ is C-*valid*, written $\models_{\mathsf{C}} \varphi$, if for all $M \in \mathsf{C}$ $M \models \varphi$.

Let $\Gamma \subseteq \mathcal{L}$. We say that φ is an *epistemic* L-*consequence* of Γ, written $\Gamma \models_{\mathsf{L}} \varphi$, if for every pointed epistemic model (M, w), if $M, w \models \Gamma$ then $M, w \models \varphi$. ($M, w \models \Gamma$ means that for all $\varphi \in \Gamma$, $M, w \models \varphi$.) □

These notions of satisfiability, validity and epistemic consequence have an intuitive import. Satisfiability of φ means that there exists an actual situation in which φ is true. Validity of φ means that in any situation, φ is true. Finally, φ is an epistemic consequence of Γ if in any situation where Γ is true, φ is also true. This notion of epistemic consequence corresponds in modal logic to the notion of local consequence. In fact, there is another notion of logical consequence in modal logic called global consequence, namely: $\Gamma \models \varphi$ if for every epistemic model M if $M, w \models \Gamma$ for all $w \in M$ then $M, w \models \varphi$ for all $w \in M$ [Blackburn *et al.*, 2001]. But we prefer to choose the notion of local consequence because this notion of global consequence does not really have a natural intuitive meaning in epistemic logic. Indeed one could prove that $\varphi \models B_j\varphi$ with the global notion of consequence, which is of course counterintuitive if we want to include agents that are ignorant of some facts.

Example 2.2.7 ('Coin' example)
We take up more or less the coin example of [Baltag and Moss, 2004]:

'Ann and Bob enter a room where a quizmaster holds a coin in his hand. The quizmaster throws the coin in the air which lands in a small box on a table in front of the quizmaster. The quizmaster can see the coin but Ann and Bob cannot. The quizmaster then closes the box.'

This situation is modeled in the pointed epistemic model (M, w_a) of Figure 2.1. The quizmaster is not considered as an agent involved in the situation, so he is not represented in the epistemic model (he might simply be a robot or a computer program). The accessibility relations are represented by arrows indexed by A (standing for Ann) or B (standing for Bob); p stands for 'the coin is heads up' and the boxed world w_a stands for the actual world. In any world, Bob and Ann consider the world where the coin is heads up and the world where the coin is tails up as being possible. So any world is accessible to any other world for A and B. Now, thanks to the language $\mathcal{L}$ we can express formally what is actually true

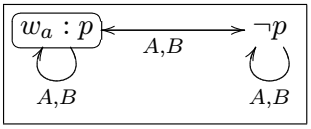


Figure 2.1: 'Coin' example.

in this situation. For example $M, w_a \models p \wedge (\hat{B}_A p \wedge \hat{B}_A \neg p) \wedge (\hat{B}_B p \wedge \hat{B}_B \neg p)$ means that the coin is heads up but both Ann and Bob do not know wether it is heads or tails up. $M, w_a \models B_B(\hat{B}_A p \wedge \hat{B}_A \neg p) \wedge B_A(\hat{B}_B p \wedge \hat{B}_B \neg p)$ means that Bob believes that Ann does not know wether the coin is heads or tails up, and that Ann does so about Bob as well. □

Epistemic models are one way to model multi-agent epistemic states. In fact, other semantic frameworks have been proposed in the literature: for example 'interpreted systems' [Fagin *et al.*, 1995] used in distributed systems, Cantwell's 'N-agent frame' [Cantwell, 2005], Lomuscio's 'hypercubes' [Lomuscio, 1999] or 'type spaces' [Harsanyi, 1967 1968] used in game theory and economics and axiomatized in [Heifetz and Mongin, 2001] (but note that this formalism uses also probability). Nevertheless it has been showed that all these formalisms can be mapped equivalently to (certain types of) epistemic models.

2.2.2 Axiomatization

Now we are going to axiomatize the semantics just defined with the help of particular (modal) logics. Generally speaking, a modal logic L is built from a set of axiom schemes and inference rules, called a *proof system*. Then a formula φ belongs to this logic either if it is an axiom or if it is derived by applying successively some inference rules to some axioms. In that case we say that φ is L-*provable* or that φ is a *theorem* of L and we write it $\vdash_{\mathsf{L}} \varphi$. A formula is L-*consistent* if its negation is not L-*provable*, formally $\nvdash_{\mathsf{L}} \neg\varphi$. (see [Blackburn *et al.*, 2001] for more details.)

An epistemic logic is obtained by joining together N modal logics (we recall that N is the cardinality of G). For sake of simplicity, it is often assumed that the axioms are the same for all the agents, meaning that they all reason with the same principles. Below is defined the simplest epistemic logic $\mathsf{K_G}$ obtained by putting together N modal logics K (which is the simplest modal logic).

Definition 2.2.8 (Proof system of $\mathsf{K_G}$)
The logic $\mathsf{K_G}$ is defined by the following axiom schemes and inference rules:

Taut	$\vdash_{K_G} \varphi$ for all propositional tautologies φ	
K	$\vdash_{K_G} B_j(\varphi \to \psi) \to (B_j\varphi \to B_j\psi)$ for all $j \in G$	(Distribution)
Nec	If $\vdash_{K_G} \varphi$ then $\vdash_{K_G} B_j\varphi$ for all $j \in G$	(Necessitation)
MP	If $\vdash_{K_G} \varphi$ and $\vdash_{K_G} \varphi \to \psi$ then $\vdash_{K_G} \psi$	(Modus Ponens).

□

The axioms of an epistemic logic obviously display the way the agents reason. For example the axiom K together with the rule of inference MP entail that if you believe φ ($B_j\varphi$) and you believe that φ implies ψ ($B_j(\varphi \to \psi)$) then you believe that ψ ($B_j\psi$). Stronger constraints

can be added. The following are often used in the literature.

D	$B_j\varphi \to \hat{B}_j\varphi$	(Consistency)
4	$B_j\varphi \to B_jB_j\varphi$	(Positive introspection)
5	$\neg B_j\varphi \to B_j\neg B_j\varphi$	(Negative introspection)
T	$B_j\varphi \to \varphi$	(Knowledge property)
.2	$\neg B_j\neg B_j\varphi \to B_j\neg B_j\neg B_j\varphi$	(Confluence)
.3	$\hat{B}_j\varphi \wedge \hat{B}_j\psi \to \hat{B}_j(\varphi \wedge \hat{B}_j\psi) \vee \hat{B}_j(\varphi \wedge \psi) \vee \hat{B}_j(\psi \wedge \hat{B}_j\varphi)$	(Weakly connected)
.3.2	$(\hat{B}_j\varphi \wedge \hat{B}_jB_j\psi) \to B_j(\hat{B}_j\varphi \vee \psi)$	
.4	$(\varphi \wedge \hat{B}_jB_j\psi) \to B_j(\varphi \vee \psi)$	(True belief)

Axiom D intuitively means that the agents' beliefs cannot be inconsistent: they do not believe both a formula and its negation. Axioms 4 and 5 intuitively mean that our agents know what they believe and disbelieve. The other axioms are more suitable for the notion of knowledge studied in Chapter 5. For example, axiom T intuitively means that everything an agent knows is true (which is not generally the case for the notion of belief). The commonly used logics are specified as follows:

$$\begin{array}{ll} \mathsf{KD45_G}: & \mathsf{K_G} + \mathsf{D} + 4 + 5; \\ \mathsf{S4_G}: & \mathsf{K_G} + \mathsf{T} + 4; \\ \mathsf{S4.2_G}: & \mathsf{S4_G} + .2; \\ \mathsf{S4.3_G}: & \mathsf{S4_G} + .3; \\ \mathsf{S4.3.2_G}: & \mathsf{S4_G} + .3.2; \\ \mathsf{S4.4_G}: & \mathsf{S4_G} + .4; \\ \mathsf{S5_G}: & \mathsf{S4_G} + 5. \end{array}$$

The relative strength of the logics for knowledge is as follows: $\mathsf{S4_G}, \mathsf{S4.2_G}, \mathsf{S4.3_G}, \mathsf{S4.3.2_G}$, $\mathsf{S4.4_G}, \mathsf{S5_G}$. The logics are in increasing order, so for instance all the theorems of $\mathsf{S4.2_G}$ are also theorems of $\mathsf{S4.3_G}, \mathsf{S4.3.2_G}, \mathsf{S4.4_G}$ and $\mathsf{S5_G}$.

An interesting feature of epistemic (and modal) logic is that we can somehow match the constraints imposed by the axioms on the belief operator B_j with constraints on the accessibility relations R_j. So for example any logic containing the axiom schemes 4 and T will be valid on the class of reflexive and transitive models, and any formula valid on the class of reflexive and transitive models is provable in $\mathsf{S4_G}$. In other words, the notions of validity and provability coincide.

Theorem 2.2.9 (Soundness and completeness)
For all $\varphi \in \mathcal{L}$ and $\mathsf{L} \in \{\mathsf{K_G}, \mathsf{KD45_G}, \mathsf{S4_G}, \mathsf{S4.2_G}, \mathsf{S4.3_G}, \mathsf{S4.3.2_G}, \mathsf{S4.4_G}, \mathsf{S5_G}\}$,

$$\vdash_{\mathsf{L}} \varphi \textit{ iff } \models_{\mathsf{L}} \varphi$$

The 'if' direction is called completeness and the 'only if' direction is called soundness. Soundness is often easily proved by induction on the length of the proof. Completeness is often proved by contraposition, building a L-model (usually called canonical model) satisfying a given L-consistent formula.

Finally, all the logics introduced are decidable, which intuitively means that we can run an algorithm that decides wether or not a given formula is satisfiable (see [Blackburn *et al.*, 2001] for details). Below, we list the complexity of the satisfiability problem for each of them. All these results are due to Halpern and Moses [Halpern and Moses, 1992].

- NP-complete for N=1 in $\mathsf{KD45}_\mathsf{G}$ and $\mathsf{S5}_\mathsf{G}$;
- PSPACE-complete for $N \geq 2$ in $\mathsf{KD45}_\mathsf{G}$ and $\mathsf{S5}_\mathsf{G}$;
- PSPACE-complete for any N in K_G and $\mathsf{S4}_\mathsf{G}$.

So far we have not really exploited the fact that we are in a multi-agent setting. That is what we are going to do now.

2.2.3 Common belief

In a multi-agent setting there are two important concepts: general belief (or knowledge) and common belief (or knowledge). The notion of common belief (or knowledge) was first studied by Lewis in the context of conventions [Lewis, 1969]. It was then applied to distributed systems [Fagin *et al.*, 1995] and to game theory [Aumann, 1976], where it allows to express that the rationality of the players, the rules of the game and the set of players are commonly known [Aumann, 1977].

General belief of φ means that everybody in the group of agents G believes that φ. Formally this corresponds to $\bigwedge_{j \in G} B_j \varphi$ and it is written $E_G \varphi$. Common belief of φ means that everybody believes φ but also that everybody believes that everybody believes φ, that everybody believes that everybody believes that everybody believes φ, and so on *ad infinitum*. Formally, this corresponds to $E_G\varphi \wedge E_G E_G \varphi \wedge E_G E_G E_G \varphi \wedge \ldots$ As we do not allow infinite conjunction the notion of common knowledge will have to be introduced as a primitive in our language.

Before defining the language with this new operator, we are going to give an example introduced by Lewis [Lewis, 1969] that illustrates the difference between these two notions (here we exceptionally use the notion of knowledge instead of belief to make things clearer). Lewis wanted to know what kind of knowledge is needed so that the statement p: "every driver must drive on the right" be a convention among a group of agents. In other words he wanted to know what kind of knowledge is needed so that everybody feels safe to drive on the right. Suppose there are only two agents i and j. Then everybody knowing p (formally $E_G p$) is not enough. Indeed, it might still be possible that the agent i considers possible that the agent j does not know p (formally $\neg K_i K_j p$). In that case the agent i will not feel safe to drive on the right because he might consider that the agent j, not knowing p, could drive on the left. To avoid this problem, we could then assume that everybody knows that everybody knows that p (formally $E_G E_G p$). This is again not enough to ensure that everybody feels safe to drive on the right. Indeed, it might still be possible that agent i considers possible that agent j considers possible that agent i does not know p (formally $\neg K_i K_j K_i p$). In that case and from i's point of view, j considers possible that i, not knowing p, will drive on the

left. So from i's point of view, j might drive on the left as well (by the same argument as above). So i will not feel safe to drive on the right. Reasoning by induction, Lewis showed that for any $k \in \mathbb{N}$, $E_G p \wedge E_G^1 p \wedge \ldots \wedge E_G^k p$ is not enough for the drivers to feel safe to drive on the right. In fact what we need is an infinite conjunction. In other words, we need common knowledge of p: $C_G p$.

Definition 2.2.10 (Language $\mathcal{L}^C$)

- The language $\mathcal{L}^C$ is defined inductively as follows:

$$\mathcal{L}^C : \varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \varphi \mid B_j\varphi \mid E_{G_1}\varphi \mid C_{G_1}\varphi$$

  where p ranges over Φ, j over G and G_1 ranges over subsets of G. Moreover, $E_{G_1}\varphi$ is an abbreviation for $\bigwedge_{j \in G_1} B_j\varphi$.

- For every epistemic model M and $w \in M$,

$$M, w \models C_{G_1}\varphi \text{ iff for all } v \in (\bigcup_{j \in G_1} R_j)^+(w),\ M, v \models \varphi;$$

  where $(\bigcup_{j \in G_1} R_j)^+$ is the transitive closure of $\bigcup_{j \in G_1} R_j$.

□

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

Finally, we can define logics with the common knowledge operator that extend the existing logics without it. For all $\mathsf{L} \in \{\mathsf{K_G}, \mathsf{KD45_G}, \mathsf{S4_G}, \mathsf{S4.2_G}, \mathsf{S5_G}\}$, the logic $\mathsf{L^C}$ is defined by adding the following axiom schemes and inference rule to those of L.

E	$\vdash_{\mathsf{L^C}} E_{G_1}\varphi \leftrightarrow \bigwedge_{j \in G_1} B_j\varphi$
Mix	$\vdash_{\mathsf{L^C}} C_{G_1}\varphi \rightarrow E_{G_1}(\varphi \wedge C_{G_1}\varphi)$
Ind	if $\vdash_{\mathsf{L^C}} \varphi \rightarrow E_{G_1}(\psi \wedge \varphi)$ then $\vdash_{\mathsf{L^C}} \varphi \rightarrow C_{G_1}\psi$ (induction rule)

We can then show that the logic $\mathsf{L^C}$ is sound and complete with respect to the class of L-models. Note that other axiomatizations exist: for instance by Lismont and Mongin [Lismont and Mongin, 1994] and Bonanno [Bonanno, 1996] (without induction rule). All these logics are still decidable but their complexity is higher, which is the price to pay for more expressivity.

2.2.4 An epistemic logic toolkit

In this section, we will list techniques stemming from modal logic that will be used throughout this thesis.

Bisimulation

Definition 2.2.11 (Bisimulation)
Let $M = (W, R, V)$ and $M' = (W', R', V')$ be two epistemic models, and let $w \in M$, $w' \in M'$. A non-empty binary relation $Z \subseteq W \times W'$ (with wZw') is called a *bisimulation* between M and M', written $Z : M, w \leftrightarroweq M', w'$, if the following conditions are satisfied.

1. If wZw' then for all $p \in \Phi$, $w \in V(p)$ iff $w' \in V'(p)$;
2. if wZw' and $v \in R_j(w)$ then there exists $v' \in R_j(w')$ such that vZv';
3. if wZw' and $v' \in R_j(w')$ then there exists $v \in R_j(w)$ such that vZv'.

We can define bisimilarity between M, w and M', w', written $M, w \leftrightarroweq M', w'$ as follows. $M, w \leftrightarroweq M', w'$ iff there is a relation Z such that $Z : M, w \leftrightarroweq M', w'$. □

The main theorem about bisimulation is the following.

Theorem 2.2.12 [Blackburn *et al.*, 2001]
Let M, M' be two epistemic models and $w \in M$, $w' \in M'$. If $M, w \leftrightarroweq M', w'$ then for all $\varphi \in \mathcal{L}^C$, $M, w \models \varphi$ iff $M', w' \models \varphi$.

So intuitively, if two epistemic models are bisimilar then they contain the same information. It can be shown that the converse also holds in case the epistemic models are finite.

Generated submodel and height

Definition 2.2.13 (Generated submodel)
Let $M = (W, R, V)$ and $M' = (W', R', V')$ be two epistemic models and $W_a \subseteq W$.

- We say that M' is a *submodel* of M if $W' \subseteq W$; for all $j \in G$, $R'_j = R_j \cap (W' \times W')$ and for all $p \in \Phi$, $V'(p) = V(p) \cap W'$. We also say that M' is the *restriction* of M to W'.
- The *submodel of M generated by W_a* is the restriction of M to $\{(\bigcup_{j \in G} R_j)^*(w) \mid w \in W_a\}$[1].
  In case the submodel of M generated by W_a is M itself, we say that M is generated by W_a.

□

Note that in the above definition, W_a could be a singleton $\{w\}$. In that case w is called the *root* of the generated submodel. The main property about generated submodels is the following.

Proposition 2.2.14 [Blackburn *et al.*, 2001]
Let $M = (W, R, V)$ be an epistemic model and M' a submodel of M generated by some $W_a \subseteq W$. Then for all $w \in M'$ and all $\varphi \in \mathcal{L}^C$, $M, w \models \varphi$ iff $M', w \models \varphi$.

[1] if R is a relation, the reflexive transitive closure of R, written R^*, is defined by $R^*(w) = \{w\} \cup \{v |$ there is $w = w_1, \ldots, w_n = v$ such that $w_i R w_{i+1}\}$, see [Blackburn *et al.*, 2001]

This property then means in particular that if we are interested only in the worlds W_a in M, then the submodel of M generated by these worlds contains all the relevant information in M about these worlds.

Now we define the related notion of height of worlds and models.

Definition 2.2.15 (Height)
Let (M, w) be an epistemic model generated by w. The notion of *height* of worlds in M is defined by induction. The only world of height 0 is the root w; the worlds of height $n+1$ are those immediate successors of worlds of height n that have not yet been assigned a height smaller than $n+1$.

The *height of a (generated) model* (M, w) is the maximum n such that there is a world of height n in (M, w), if such a maximum exists; otherwise the height of (M, w) is infinite. □

Syntactic characterization of finite epistemic models

Proposition 2.2.16 [Barwise and Moss, 1997; van Benthem, 2006]
Let Φ be finite, and let M be a finite epistemic model and $w \in M$. Then there is an epistemic formula $\delta_M(w) \in \mathcal{L}^C$ (involving common knowledge) such that

1. $M, w \models \delta_M(w)$
2. *For every finite epistemic model M' and world $w' \in M'$, if $M', w' \models \delta_M(w)$ then $M, w \leftrightarrows M', w'$.*

This proposition intuitively means that any *finite* epistemic model can be completely characterized by an epistemic formula.

The universal modality

Definition 2.2.17 (Language $\mathcal{L}^U$)
We define the language $\mathcal{L}^U$ inductively as follows.

$$\mathcal{L}^U : \varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \varphi \mid B_j\varphi \mid U\varphi,$$

where p ranges over Φ and j over G. Moreover $O\varphi$ is an abbreviation for $\neg U \neg \varphi$.The truth condition for the universal modality U is defined as follows.

$$M, w \models U\varphi \text{ iff for all } v \in M, M, v \models \varphi.$$

□

So the universal modality is something stronger than common belief. Its axiomatization is made of the axioms for S5 [Blackburn *et al.*, 2001].

2.3 A new approach: internal versus external perspectives of the world

2.3.1 Intuitions

In the literature about epistemic logic, when it comes to model epistemic situations involving several agents $j_1, \ldots, j_N$, not much is said explicitly about which modeling point of view is considered. However, modeling an epistemic situation depends very much on the modeling point of view. Indeed, the models built will be quite different whether the modeler is one of the agents $j_1, \ldots, j_N$ or not. Let us take up the Example 2.2.7. Now, assume that the quizmaster somehow manages to privately announce to Bob that the coin is heads up (by showing him the coin for example), Ann suspecting nothing about it (she might be inattentive or out of the room for a while). On the one hand, if the modeler is somebody external (different from Ann and Bob) knowing everything that has happened, then in the model that this modeler builds to represent this resulting situation Bob knows whether the coin is heads or tails up. On the other hand, if the modeler is Ann herself then in the model that Ann builds to represent this resulting situation Bob does not know whether the coin is heads or tails up. As we see in this example, specifying the modeling point of view is also quite essential to interpret the formal models.[2]

But what kinds of modeling points of view are there? For a start, we can distinguish whether the modeler is one of the agents $j_1, \ldots, j_N$ or not.

1. First, consider the case where the modeler is *one* of the agents $j_1, \ldots, j_N$. In the rest of the thesis we call this modeler-agent agent Y (like You). The models she builds could be seen as models she has 'in her mind'. They represent the way she perceives the surrounding world. In that case, agent Y is involved in the situation, she is considered on a par by the other agents and interacts with them. So she should be represented in the formalism and her models should deal not only with the other agents' beliefs but also with the other agents' beliefs about her own beliefs. This is an internal and subjective point of view, the situation is modeled from the inside. Therefore, for this very reason her beliefs might be erroneous. Hence the models she builds might also be erroneous. We call this agent point of view the *internal* point of view.

2. Second, consider the case where the modeler is not one of the agents $j_1, \ldots, j_N$. The modeler is thus somebody external to the situation. She is not involved in the situation and she does not exist for the agents, or at least she is not taken into consideration in their representation of the world. So she should not be represented in the formalism and particularly the agents' beliefs about her own beliefs should also not be represented because they simply do not exist. The models that this modeler builds

[2] This is somehow similar to what happens in Newtonian mechanics in physics where we always have to specify which (Galilean) referential we consider when we want to model a phenomenon because the perception of this phenomenon depends on this referential. For example, assume somebody drops a ball from the top of a ship's high mast sailing rapidly nearby a harbor. Then, viewed from the referential of the ship, the trajectory of the ball will be a straight line. But viewed from the referential of the harbor, the trajectory will be a parabola.

are supposed to represent the situation from an external and objective point of view. Typically, as in the internal point of view, her models deal with the epistemic states of all the agents $j_1, \ldots, j_N$ and also the actual state of the world. There are then two other possibilities depending on whether or not the modeler has a perfect knowledge of the situation.

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

(b) In case the modeler does not have a perfect knowledge of the situation then, unlike the perfect external point of view, the models built might be erroneous. The models could also be correct but then, typically, the modeler would be uncertain about which is the actual world (in that sense, she would not have a perfect knowledge of the situation). What the modeler knows can be obtained for example by observing what the agents say and do, by asking them questions ... We call this point of view the *imperfect external* point of view.

Because we proceeded by successive dichotomies, we claim that the internal, the perfect external and the imperfect external points of view are the only three possible points of view when we want to model epistemic situations. From now on we will call them the internal, the external and the imperfect external approaches. In the literature, these three approaches are sometimes mixed leading to technical or philosophical problems. We will give an example of such problems in Section 3.5. However, note that if in the external approach the object of study is not the epistemic states of all the agents $j_1, \ldots, j_N$ (and the actual state of the world) but rather the epistemic state of only one of these agents, then the perfect external approach focused on this agent boils down to the internal approach (for this agent). Note also that in the imperfect external approach, in practice, the modeler could perfectly be one of the agents who wants to reason about the other agents from an external point of view, as if she was not present.

The fields of application of these three approaches are different. The internal and imperfect external approaches have rather applications in artificial intelligence where agents/robots acting in the world need to have a formal representation of the surrounding world and to cope with uncertain information. The internal approach has also applications in cognitive psychology where the aim is to model the cognition of one agent (possibly in a multi-agent setting). The perfect external approach has rather applications in game theory [Battigalli and Bonanno, 1999], cognitive psychology or distributed systems [Fagin *et al.*, 1995] for example. Indeed, in these fields we need to model situations accurately from an external point of view in order to explain and predict what happens in these situations.

In this thesis we will focus only on the perfect external and the internal approach. That is why from now on we omit the term 'perfect' in 'perfect external'. For a work considering

similar questions as ours using an imperfect external approach, see [Nittka, 2008; Booth and Nittka, 2007b; Booth and Nittka, 2007a].

Standard epistemic logic described in Section 2.2 and all the papers cited there rather follow the (perfect) external approach. On the other hand, AGM belief revision theory [Alchourrón *et al.*, 1985] rather follows the internal approach. But AGM is designed for a single agent. In fact there is no logical formalism for the internal approach in a multi-agent setting. That is what we are going to propose in this chapter.

2.3.2 A semantics for the internal approach

To define a semantics for the internal approach in a multi-agent setting, we will start from the AGM approach, based on the notion of possible world, and then extend it to the multi-agent case. Then we will propose an equivalent formalism which will be used in the rest of this thesis.

But first we have to make some assumption. As we said in the previous section, the internal approach has applications in artificial intelligence and in cognitive psychology. So the objects we introduce should be essentially finite. Indeed, computers cannot easily deal with infinite structures and a human cognition is by nature finite. So the set Φ of propositional letters is assumed to be finite in the rest of this chapter.

Multi-agent possible world and internal model

In the AGM framework, one considers a single agent Y. The possible worlds are supposed to represent how the agent Y perceives the surrounding world. As she is the only agent, these possible worlds deal only with propositional facts about the surrounding world. Now, if we suppose that there are other agents than agent Y, a possible world for Y in that case should also deal with how the other agents perceive the surrounding world. These "multi-agent" possible worlds should then not only deal with propositional facts but also with epistemic facts. So to represent a multi-agent possible world we need to add a modal structure to our (single agent) possible worlds. We do so as follows.

Definition 2.3.1 (Multi-agent possible world)
A *multi-agent possible world* (M, w) is a *finite* pointed epistemic model $M = (W, R, V, w)$ generated by $w \in W$ such that R_j is serial, transitive and euclidean for all $j \in G$, and

1. $R_Y(w) = \{w\}$;
2. there is no v and $j \neq Y$ such that $w \in R_j(v)$.

□

Let us have a closer look at the definition. Condition 2 will be motivated later, but note that any pointed epistemic model satisfying the conditions of a multi-agent possible world except condition 2 is bisimilar to a multi-agent possible world. Condition 1 ensures that in case Y is the only agent then a multi-agent possible world boils down to a possible world, as in the AGM theory. Condition 1 also ensures that in case Y assumes that the situation is

correctly represented by the multi-agent possible world (M, w) then for her w is the (only) actual world. In fact the other possible worlds of a multi-agent possible world are just present for technical reasons: they express the other agents' beliefs (in world w). One could get rid of the condition that a multi-agent possible world (M, w) is generated by w but the worlds which do not belong to the submodel generated by w would have neither philosophical nor technical motivation. Besides, for the same reason that Φ is finite, a multi-agent possible world is also assumed to be finite. Finally, notice that we assume that accessibility relations are serial, transitive and euclidean. This means that the agents' beliefs are consistent and that agents know what they believe and disbelieve (see Section 2.2.2). These seem to be very natural constraints to impose on the notion of belief. Intuitively, this notion of belief corresponds for example to the kind of belief in a theorem that you have after having proved this theorem and checked the proof several times. In the literature, this notion of belief corresponds to Lenzen's notion of conviction [Lenzen, 1978] or to Gärdenfors' notion of acceptance [Gärdenfors, 1988] or to Voorbraak's notion of rational introspective belief [Voorbraak, 1993]. In fact, in all the agent theories the notion of belief satisfies these constraints: in Cohen and Levesque's theory of intention [Cohen and Levesque, 1990] or in Rao and Georgeff BDI architecture [Georgeff and Rao, 1991] [Rao and Georgeff, 1991] or in Meyer et. al. KARO architecture [van Linder *et al.*, 1998] [Meyer *et al.*, 2001] or in Wooldridge BDI logic LORA [Wooldridge, 2000]. However, one should note that all these agent theories follow the external approach and thus use standard epistemic models (defined in Definition 2.2.1) to represent the situation. This is of course at odds with their intention to implement their theories in machines.[3]

Remark 2.3.2 In this chapter we deal only with the notion of belief but one could also add the notion of knowledge. Indeed, it might be interesting to express things such as 'the agent Y believes that agent j does not *know* p' (even if this could be rephrased in terms of beliefs). We refrain to do so in order to highlight the main new ideas and because in most applications of the internal approach the notion of knowledge is not essential. □

Example 2.3.3 We see in Figure 2.2 that a multi-agent possible world is really a generalization of a possible world. □

In the single agent case (in AGM belief revision theory), the epistemic state of the agent Y is represented by a finite set of possible worlds. In a multi agent setting, this is very similar: the epistemic state of the agent Y is represented by a (disjoint and) finite set of *multi-agent* possible worlds. We call this an internal model of type 1.

[3] One could argue that the epistemic models of these theories could be somehow used by the agent at stake (the machine/robot) to reason about the other agents' beliefs from an external point of view, as if she was not present. But in that case, because this agent would have some uncertainty about the situation, a single epistemic model would not be enough for the agent to represent the situation. Instead, the agent would need a *set* of epistemic models, each epistemic model representing a possible determination of the situation (for the agent). So, even in that case, the formalism would still be very different because the primitive semantical object to consider should not be a single epistemic model but rather a *set* of epistemic models. In fact, this boils down to follow the imperfect external approach.

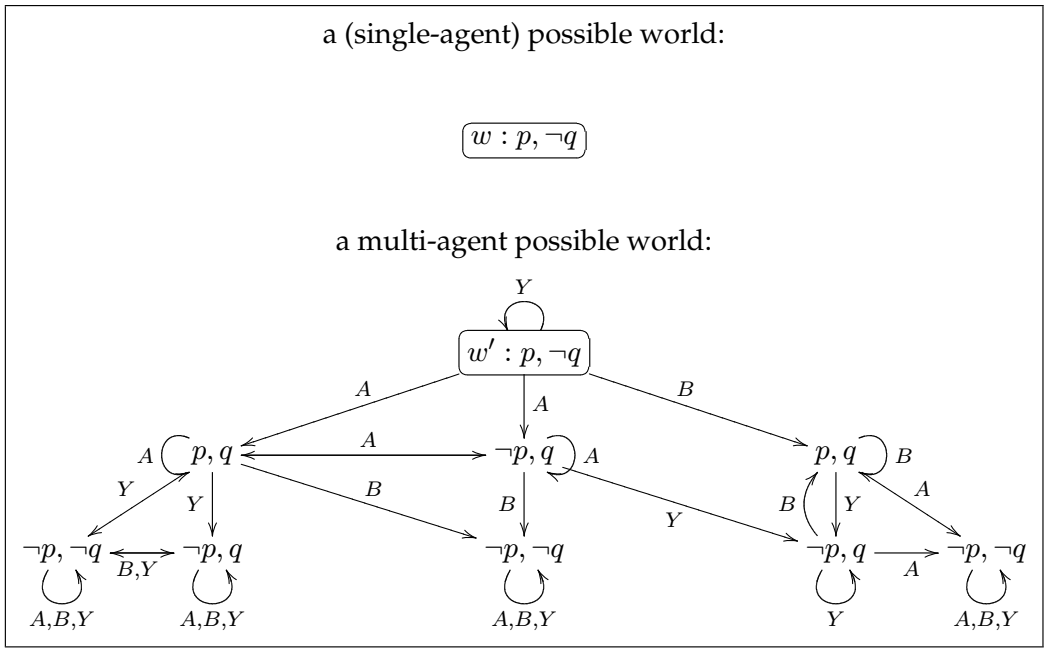


Figure 2.2: From possible world to multi-agent possible world

Definition 2.3.4 (Internal model of type 1)
An *internal model of type 1* is a disjoint and finite union of multi-agent possible worlds. □

An internal model of type 1 will sometimes be written $(\mathcal{M}, W_a)$ where W_a are the roots of its multi-agent possible worlds.

Example 2.3.5 ('Coin' example)
Let us take up the 'coin example' of Example 2.2.7 before the private announcement of the quizmaster and let us consider Bob's internal point of view. So in this example, Bob stands for the designated agent Y. Bob's internal model of type 1 is depicted in Figure 2.3. There p stands for 'the coin is heads up', A for Ann and B for Bob. In this internal model, Bob does not know wether the coin is heads or tails up (formally $\neg B_B p \wedge \neg B_B \neg p$). Indeed, in one multi-agent possible world (on the left) p is true at the root and in the other (on the right) p is false at the root. Bob also believes that Ann does not know whether the coin is heads or tails up (formally $B_B(\neg B_A p \wedge \neg B_A \neg p)$. Indeed, in both multi-agent possible worlds, $\neg B_A p \wedge \neg B_A \neg p$ is true (at the roots). Finally, Bob believes that Ann believes that she does not know whether the coin is heads or tails up (formally $B_B B_A(\neg B_B p \wedge \neg B_B \neg p)$) since $B_A(\neg B_B p \wedge \neg B_B \neg p)$ is true at the roots of both multi-agent possible worlds. □

Thanks to condition 2 in the definition of a multi-agent possible world, we could define the notion of internal model differently. Indeed, we could perfectly set an accessibility relation between the roots of the multi-agent possible worlds. Figure 2.4 gives an example of such a process, starting from the example of Figure 2.3. Condition 2 ensures us that by

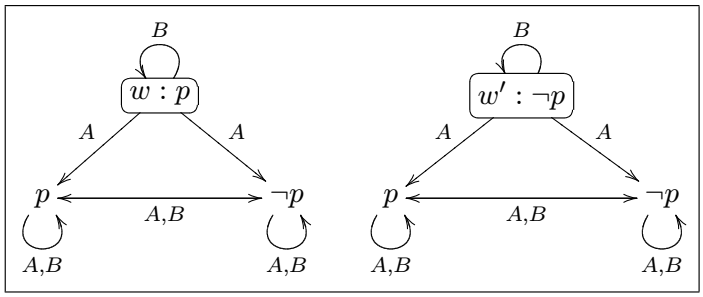


Figure 2.3: Bob's internal model of type 1 before the private announcement

doing so we do not modify the information present in the original internal model. Indeed, if condition 2 was not fulfilled then it might be possible that j's beliefs about Y's beliefs (for some $j \neq Y$) might be different between the original internal model and the new one, due to the creation of these new accessibility relations between the multi-agent possible worlds. This phenomenon will become explicit when we define the language for the internal models of type 1. (Condition 2 will turn out to be useful in Chapter 4 as well.)

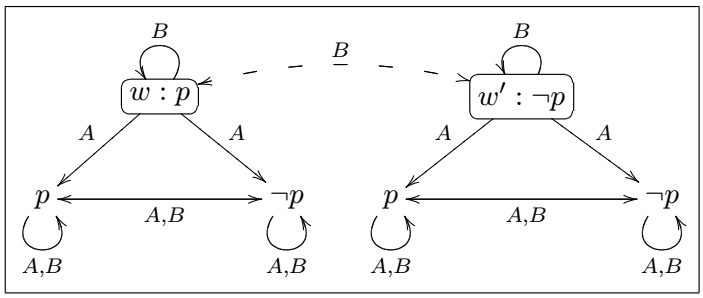


Figure 2.4: A new definition of internal model

Then in this new formalism, one can notice that the former roots of the multi-agent possible worlds form an equivalence class for the accessibility relation indexed by B, which stands in this example for the agent Y. Note also that the accessibility relations are still serial, transitive and euclidean. This leads us to the following new definition of an internal model.

Definition 2.3.6 (Internal model of type 2)
An *internal model of type 2* is a couple $(\mathcal{M}, W_a)$ where $\mathcal{M}$ is a finite epistemic model $\mathcal{M} = (W, R, V)$ generated by $W_a \subseteq W$ such that R_j is serial, transitive and euclidean for all $j \in G$, and $R_Y(w_a) = W_a$ for all $w_a \in W_a$. W_a is called the *actual equivalence class*. □

Definition 2.3.7 (Internal model of type 2 associated to an internal model of type 1)
Let $(\mathcal{M}, W_a) = \{(M^1, w^1), \ldots, (M^n, w^n)\}$ be an internal model of type 1 (with $M^k = (W^k, R^k, V^k)$).

The *internal model of type 2 associated to* $(\mathcal{M}, W_a)$ is the internal model of type 2, $S_2(\mathcal{M}, W_a) = (W', R', V', W_a)$, defined as follows.

- $W' = \bigcup_k W^k$;
- $R'_j = \bigcup_k R^k_j$ for $j \neq Y$, and $R'_Y = \{(w^k, w^{k'}) \mid w^k, w^{k'} \in W_a\} \cup \bigcup_k R^k_j$;
- $V'(p) = \bigcup_k V^k(p)$.

□

So from an internal model of type 1, one can easily define an equivalent internal model of type 2. But of course, the other way around, from an internal model of type 2 one can also define an equivalent internal model of type 1. This will be done in Proposition 2.3.12.

Example 2.3.8 ('Coin' example)
In Figure 2.5 Bob's internal model of type 2 before the private announcement is depicted. We recall that in this example Bob stands for the designated agent Y. The worlds of the actual equivalence class are within boxes. It turns out that this internal model is bisimilar to the one depicted in Figure 2.4, which is itself an equivalent representation of Bob's internal model of type 1 depicted in Figure 2.3. So Bob's internal model of type 2 depicted in Figure 2.5 is an equivalent representation of the Bob's internal model of type 1 depicted in Figure 2.3. One can indeed check for example that the formulas $\neg B_B p \wedge \neg B_B \neg p$, $B_B(\neg B_A p \wedge \neg B_A \neg p)$ and $B_B B_A(\neg B_B p \wedge \neg B_B \neg p)$ are indeed true. Note that this second representation is much more compact. □

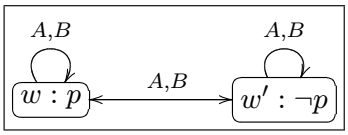


Figure 2.5: Bob's internal model of type 2 before the private announcement

As we said in Section 2.3.1, the internal approach can be applied in artificial intelligence. In this case, the agent Y is an artificial agent (such as a robot) that has an internal model 'in her mind'. But to stick with a more standard approach (used in the single agent case), we could perfectly consider that the agent Y has sentences from a particular language 'in her mind' and draws inferences from them. In that respect, this language could also be used by the agent Y in the former approach to perform some model checking in her internal model in order to reason about the situation or to answer queries. So in any case we do need to define a language.

Language for the internal approach

Definition 2.3.9 (Language $\mathcal{L}$)

$$\mathcal{L} : \varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \varphi \mid B_j\varphi$$

where p ranges over Φ and j over G. □

For sake of simplicity and in order to highlight the new results, we do not introduce a common knowledge operator, but this could be done easily. In fact all the results of this chapter still hold if we add the common knowledge operator to the language. Note that the language is identical to the usual language of epistemic logic. If we consider the class of internal models of type 2 then its truth conditions are also the same and are spelled out in Definition 2.2.4. But if we consider the class of internal models of type 1 then its truth conditions are a bit different and are set out below.

Definition 2.3.10 (Truth conditions for $\mathcal{L}$)
Let $(\mathcal{M}, \{w^1, \ldots, w^n\}) = \{(M^1, w^1), \ldots, (M^n, w^n)\}$ be an internal model of type 1 and let $w \in \mathcal{M}$. Then $w \in M^k$ for some k, with $M^k = (W^k, R^k, V^k)$. $\mathcal{M}, w \models \varphi$ is defined inductively as follows:

$$\begin{array}{lll}
\mathcal{M}, w \models \top & & \\
\mathcal{M}, w \models p & \text{iff} & w \in V^k(p) \\
\mathcal{M}, w \models \neg\varphi & \text{iff} & \text{not } \mathcal{M}, w \models \varphi \\
\mathcal{M}, w \models \varphi \wedge \varphi' & \text{iff} & \mathcal{M}, w \models \varphi \text{ and } \mathcal{M}, w \models \varphi' \\
\mathcal{M}, w \models B_Y\varphi & \text{iff} & \begin{cases} \text{for all } w^i \in W_a, \mathcal{M}, w^i \models \varphi & \text{if } w \in W_a \\ \text{for all } w' \in R^k_Y(w), \mathcal{M}, w' \models \varphi & \text{if } w \in W^k - W_a \end{cases} \\
\mathcal{M}, w \models B_j\varphi & \text{iff} & \text{for all } w' \in R^k_j(w), \mathcal{M}, w' \models \varphi \quad \text{if } j \neq Y
\end{array}$$

□

Note that the truth condition for the operator B_Y is defined as if there were accessibility relations indexed by Y between the roots of the multi-agent possible worlds. Condition 2 of Definition 2.3.1 then ensures that the agents j's beliefs about agent Y's beliefs (with $j \neq Y$) of a given multi-agent possible world stay the same whatever other multi-agent possible world we add to this multi-agent possible world. This would of course be a problem if it was not the case. Condition 2 thus provides a kind of modularity of the multi-agent possible worlds in an internal model (of type 1).

This truth condition for the operator B_Y is of course completely in line with the truth conditions for the internal models of type 2. In fact, thanks to the definition of this language, we can show that the two types of internal models are somehow equivalent. Both will be useful in this thesis, but in the rest of this chapter we consider that all internal models will be internal models of type 2.

Definition 2.3.11 (Equivalence between internal models of type 1 and 2)
Let $(\mathcal{M}, W_a)$ be an internal model of type 1 and $(\mathcal{M}', W'_a)$ be an internal model of type 2. $(\mathcal{M}, W_a)$ and $(\mathcal{M}', W'_a)$ are equivalent if and only if

- for all $w \in W_a$ there is $w' \in W'_a$ such that for all $\varphi \in \mathcal{L}$, $\mathcal{M}, w \models \varphi$ iff $\mathcal{M}', w' \models \varphi$;
- for all $w \in W'_a$ there is $w' \in W_a$ such that for all $\varphi \in \mathcal{L}$, $\mathcal{M}, w \models \varphi$ iff $\mathcal{M}', w' \models \varphi$.

□

Proposition 2.3.12 (Internal models of type 1 and 2 are equivalent)
Let $(\mathcal{M}, W_a)$ be an internal model of type 2. Then there is an internal model of type 1 $(\mathcal{M}', W'_a)$ which is equivalent to $(\mathcal{M}, W_a)$.

Let $(\mathcal{M}, W_a)$ be an internal model of type 1. Then there is an internal model of type 2 $(\mathcal{M}', W'_a)$ which is equivalent to $(\mathcal{M}, W_a)$.

PROOF. We only prove the first part. For the second part, it suffices to take the internal model of type 2 associated to $(\mathcal{M}, W_a)$.

Let $(\mathcal{M}, W_a) = (W, R, V)$ be an internal model of type 2. For each $w \in W_a$ we define a corresponding multi-agent possible world (M^w, w) as follows: for all $k \neq Y$, let $M^k = (W^k, R^k, V^k)$ be the submodel of M generated by $R_k(w)$. The multi-agent possible world $(M^w, w) = (W^w, R^w, V^w, w)$ is then defined as follows.

- $W^w = \{w\} \sqcup \bigsqcup_{k \neq Y} W^k$;
- $R^w_j = \left(R_j \cup \bigcup_{k \neq Y} R^k_j \right) \cap W^w \times W^w$ for all $j \in G$;
- $V^w(p) = \left(V(p) \cup \bigcup_{k \neq Y} V^k(p) \right) \cap W^w$.

Then one can easily show that (M^w, w) is a multi-agent possible world and that $\mathcal{M}' = \{(M^w, w) \mid w \in W_a\}$ is an internal model of type 1 which is equivalent to $(\mathcal{M}, W_a) = (W, R, V)$. QED

Thanks to the truth conditions we can now define the notions of satisfiability and validity of a formula. The truth conditions are defined for any world of the internal model. However, the satisfiability and the validity should not be defined relatively to any possible world of the internal model (as it is usually done in epistemic logic) but only to the possible worlds of the actual equivalence class. Indeed, these are the worlds that do count for the agent Y in an internal model: they are the worlds that agent Y actually considers possible. The other possible worlds are just here for technical reasons in order to express the other agents' beliefs (in these worlds). This leads us to the following definition of satisfiability and validity.

Definition 2.3.13 (Internal satisfiability and validity)
Let $\varphi \in \mathcal{L}$. The formula φ is *internally satisfiable* if there is an internal model $(\mathcal{M}, W_a)$ and there is $w \in W_a$ such that $\mathcal{M}, w \models \varphi$. The formula φ is *internally valid* if for all internal models $(\mathcal{M}, W_a)$, $\mathcal{M}, w \models \varphi$ for all $w \in W_a$. In this last case we write $\models_{\mathsf{Int}} \varphi$. □

Remark 2.3.14 One could define the notions of internal satisfiability and internal validity differently. One could say that $\varphi \in \mathcal{L}$ is satisfiable if there is an internal model $(\mathcal{M}, W_a)$ such that $\mathcal{M}, w \models \varphi$ for all $w \in W_a$. Then, following this new definition, $\varphi \in \mathcal{L}$ is valid if for every internal model $(\mathcal{M}, W_a)$, there is $w \in W_a$ such that $\mathcal{M}, w \models \varphi$.

This second notion of internal validity corresponds to Gärdenfors' notion of validity [Gärdenfors, 1988]. In fact these two notions of internal validity correspond to the two notions of validity introduced by Levi and Arlo Costa [Arló Costa and Levi, 1996]: they call the first one "positive validity" and the second one "negative validity".

These two notions coincide in the single agent case but not in the multi-agent case. Indeed, the Moore sentence $p \wedge \neg B_Y p$ is positively satisfiable but not negatively satisfiable. Nevertheless there are some connections between them. We can indeed prove that $\varphi \in \mathcal{L}$ is positively valid if and only if $B_Y \varphi$ is negatively valid. Moreover, both have advantages and drawbacks. On the one hand, positive validity is intuitive because it says that a formula φ is valid if in every possible situation, the agent Y believes φ. However positive satisfiability is less intuitive because φ is positively satisfiable if there exists a situation in which the agent Y does not reject φ. On the other hand, negative satisfiability is also intuitive because it says that φ is negatively satisfiable if there exists a situation in which agent Y believes φ. However negative validity is less intuitive because it says that φ is negatively valid if in every situation agent Y does not reject φ. □

As we saw in Section 2.2.1, in modal logic [Blackburn *et al.*, 2001] there are two notions of semantic consequence. In the internal approach we can also define two notions of semantic consequence.

Definition 2.3.15 (Local and global internal consequence)
Let C be a class of internal models; let Σ be a set of formulas of $\mathcal{L}$ and let $\varphi \in \mathcal{L}$.

- We say that φ is a *local internal consequence* of Σ over C, written $\Sigma \models_{\mathsf{C}} \varphi$, if for all internal models $(\mathcal{M}, W_a) \in \mathsf{C}$ and all $w \in W_a$, if $\mathcal{M}, w \models \Sigma$ then $\mathcal{M}, w \models \varphi$.
- We say that φ is a *global internal consequence* of Σ over C, written $\Sigma \models^g_{\mathsf{C}} \varphi$, if and only if for all internal models $(\mathcal{M}, W_a) \in \mathsf{C}$, if $\mathcal{M}, w \models \Sigma$ for all $w \in W_a$ then $\mathcal{M}, w \models \varphi$ for all $w \in W_a$.

□

For example, if we take any class C of internal models then it is not necessarily the case that $\varphi \models_{\mathsf{C}} B_Y \varphi$, whereas we do have that $\varphi \models^g_{\mathsf{C}} B_Y \varphi$. Moreover, these two notions can be informally associated to the two notions of satisfiability mentioned in Remark 2.3.14: local internal consequence can be associated to positive satisfiability and the global internal consequence can be associated to negative satisfiability.

2.3.3 Some connections between the internal and the external approach

Intuitively, there are some connections between the internal and the external approach. Indeed, in the external approach the modeler is supposed to perfectly know how the agents

perceive the surrounding world. So from the model she builds we should be able to extract the internal model of each agent. Likewise, it seems natural to claim that for the agent Y a formula is true if and only if, externally speaking, the agent Y believes this formula. In this subsection we are going to formalize these intuitions.

From external model to internal model and vice versa

First we define the notion of external model. An external model is a pointed epistemic model $(M, w_a) = (W, R, V, w_a)$ where $w_a \in W$ and the accessibility relations R_j are serial, transitive and euclidean. So what we call an external model is just a standard pointed epistemic model used in epistemic logic. An external model is supposed to model truthfully and from an external point of view how all the agents involved in the same situation perceive the actual world (represented formally by w_a). This is thus simply the type of model built by the modeler in the external approach spelled out in Section 2.3.1. The language and truth conditions for these external models are the same as in epistemic logic and are spelled out in Definitions 2.3.9 and 2.2.4. The notion of external validity is also the same as in epistemic logic and we say that $\varphi \in \mathcal{L}$ is *externally valid*, written $\models_{\mathsf{Ext}} \varphi$, if for all external model (M, w), $M, w \models \varphi$ (and similarly for *external satisfiability*).

Now for a given external model representing truthfully how a situation is perceived by the agents, we can extract for each agent her internal model pertaining to this situation.

Definition 2.3.16 (Model associated to an agent in an external model)
Let (M, w_a) be an external model and let $j \in G$. The *model associated to agent j in* (M, w_a) is the submodel of M generated by $R_j(w_a)$. Besides $R_j(w_a)$ is its actual equivalence class. □

Because the external model is supposed to model truthfully the situation, w_a does correspond formally to the actual world. So $R_j(w_a)$ are the worlds that the agent j actually considers possible in reality. In agent j's internal model pertaining to this situation, these worlds should then be the worlds of the actual equivalence class. Finally, taking the submodel generated by these worlds ensures that the piece of information encoded in the worlds $R_j(w_a)$ in the external model is kept unchanged in the associated internal model. This notion of model associated to agent j in (M, w_a) corresponds to the notion of belief horizon of agent j of Tallon, Vergnaud and Zamir [Tallon *et al.*, 2004].

Proposition 2.3.17 *Let (M, w_a) be an external model. The model associated to agent j in (M, w_a) is an internal model (of type 2).*

PROOF. Let $(\mathcal{M}', W'_a)$ be the internal model associated to agent j in (M, w_a) (with $\mathcal{M}' = (W', R', V')$).

Obviously, $\mathcal{M}'$ is generated by W_a. By the generated submodel property, R'_j is serial, transitive and euclidean for all j. Finally, because R_j is euclidean, for all $w \in W_a (= R_j(w_a))$, $R_j(w) = R_j(w_a) = W_a$.

So $(\mathcal{M}', W_a)$ is indeed an internal model. QED

Example 2.3.18 ('Coin' example)
In Figure 2.6 is depicted the 'coin example' after the private announcement to Bob (see Section 2.3.1). We can check that in the external model, Ann does not know whether the coin is heads or tails up and moreover believes that Bob does not know either. This is also true in the internal model associated to Ann. However, in the external model, Bob knows that the coin is heads up but this is false in the internal model associated to Ann and true in Bob's internal model. Note finally that the internal model associated to Bob is the same as the external model. This is because we assumed that Bob perceived correctly the situation and what happened. □

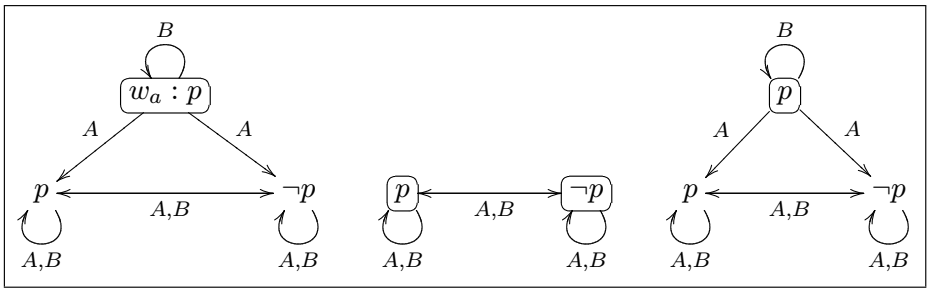


Figure 2.6: External model (M, w_a) (*left*); Internal model associated to Ann (*center*), Internal model associated to Bob (*right*).

So we know from an external model how to obtain the internal model of each agent. But the other way round, we could wonder how to get the external model (of a particular situation) if we suppose given the internal models of each agent. In that case we must moreover assume that the modeler knows the real state of the world, more precisely she knows what propositional facts are true in the actual world. We can then introduce a single world w_a whose valuation V_a satisfies these propositional facts. The external model is built by setting accessibility relations indexed by j from w_a to the actual equivalence class of j's internal model, and so for each agent j.

Definition 2.3.19 (External model associated to a set of internal models and an actual world)
Let $\{(\mathcal{M}^j, W^j) \mid j \in G\}$ be a set of internal models of type 2 for each agent j ($\mathcal{M}^j = (W^j, R^j, V^j)$) and (w_a, V_a) a possible world together with a valuation.[4]

The external model associated to $\{(\mathcal{M}^j, W^j) \mid j \in G\}$ and (w_a, V_a), written $Ext\left[\{(\mathcal{M}^j, W^j) \mid j \in G\}, (w_a, V_a)\right] = (W, R, V, w_a)$, is defined as follows.

- $W = \{w_a\} \cup \bigcup_{k \in G} W^k$;
- $R_j = \{(w_a, w^j) \mid w^j \in W^j\} \cup \bigcup_{k \in G} R^k_j$ for each $j \in G$;

[4] An internal model *for agent* j is an internal model where the designated agent is j instead of Y.

- $V(p) = V_a(p) \cup \bigcup_{k \in G} V^k(p)$.

□

We can then easily check that this model is indeed an external model.

A semantic correspondence

As we said in Section 2.3.2, the language of the internal approach is the same as that of the external approach. This enables us to draw easily some connections between the two approaches.

Proposition 2.3.20 *For all $\varphi \in \mathcal{L}$, $\models_{\mathit{Int}} \varphi$ iff $\models_{\mathit{Ext}} B_Y \varphi$.*

PROOF. For all $\varphi \in \mathcal{L}$, $\models_{\mathsf{Int}} \varphi$ iff $\models_{\mathsf{Ext}} B_Y \varphi$ amounts to prove that for all $\varphi \in \mathcal{L}$, φ is internally satisfiable iff $\hat{B}_Y \varphi$ is externally satisfiable.

Assume that φ is internally satisfiable. Then there is an internal model $(\mathcal{M}, W_a)$ and $w \in W_a$ such that $\mathcal{M}, w \models \varphi$. But $w \in R_Y(w)$, so $\mathcal{M}, w \models \hat{B}_Y \varphi$. Besides, $(\mathcal{M}, w)$ can be viewed as an external model. So $\hat{B}_Y \varphi$ is externally satisfiable.

Assume that $\hat{B}_Y \varphi$ is externally satisfiable. Then there is an external model (M, w_a) such that $M, w_a \models \hat{B}_Y \varphi$. Then there is $w \in R_Y(w_a)$ such that $M, w \models \varphi$. Let $(\mathcal{M}', W_a)$ be the internal model associated to (M, w_a) and agent Y. Then $w \in W_a$ and $\mathcal{M}', w \models \varphi$ by the generated submodel property. So φ is internally satisfiable. QED

Intuitively, this result is correct: for you φ is true if and only if from an external point of view you believe that φ is the case. (Note that this result does not hold for the notion of negative validity.)

As we said earlier, instead of internal models, agent Y might have formulas 'in her mind' in order to represent the surrounding world. But to draw inferences from them she needs a proof system. In other words, we still need to axiomatize the internal semantics. That is what we are going to do now.

2.3.4 Axiomatization of the internal semantics

First some notation. Let Ext designate from now on the logic $\mathsf{KD45}_\mathsf{G}$. So for all $\varphi \in \mathcal{L}$, $\vdash_{\mathsf{Ext}} \varphi$ iff $\varphi \in \mathsf{KD45}_\mathsf{G}$.

Definition 2.3.21 (Proof system of Int)
The *internal logic* Int is defined by the following axiom schemes and inference rules:

T	$\vdash_{\mathsf{Int}} B_Y \varphi \to \varphi$;
I-E	$\vdash_{\mathsf{Int}} \varphi$ for all $\varphi \in \mathcal{L}$ such that $\vdash_{\mathsf{Ext}} \varphi$;
MP	if $\vdash_{\mathsf{Int}} \varphi$ and $\vdash_{\mathsf{Int}} \varphi \to \psi$ then $\vdash_{\mathsf{Int}} \psi$.

□

Let us have a closer look at the axiom schemes. The first one tells us that for you, everything you believe is true. This is coherent if we construe the notion of belief as conviction. The second one tells us that you should believe everything which is objectively true, i.e. which is true independently of your own subjectivity. Finally note that the necessitation rule ($\vdash_{\mathsf{Int}} \varphi$ implies $\vdash_{\mathsf{Int}} B_j\varphi$ for all j) is not present, which is intuitively correct. Indeed, if for you φ is true (i.e. you believe φ) then in general there is no reason that you should believe that the other agents believe φ as well. For example, $B_Y\varphi \to \varphi$ is internally valid but $B_j(B_Y\varphi \to \varphi)$ (for $j \neq Y$) should not be internally valid.

As we announced it in Section 2.3.2, if we add a common knowledge operator to our language then the axiomatization for the language with common knowledge is identical to the one of the above definition.

Remark 2.3.22 In Remark 2.3.14, we proposed an alternative definition of validity for the internal semantics, called negative validity. We do not have a complete axiomatization for the negative validity. However we know that the axiom scheme $\varphi \to B_j\varphi$ is valid but Modus Ponens does not hold anymore. □

Theorem 2.3.23 (Soundness and completeness)
For all $\varphi \in \mathcal{L}$, $\models_{\mathsf{Int}} \varphi$ iff $\vdash_{\mathsf{Int}} \varphi$.

PROOF. Proving the soundness of the axiomatic system is straightforward. We only focus on the completeness proof.

Let φ be a Int-consistent formula. We need to prove that there is an internal model $(\mathcal{M}_{\mathsf{Int}}, W_a)$, there is $w \in W_a$ such that $\mathcal{M}_{\mathsf{Int}}, w \models \varphi$.

Let $Sub^+(\varphi)$ be all the subformulas of φ with their negations. Let W_{Int} be the set of maximal Int-consistent subsets of $Sub^+(\varphi)$. Let W_{Ext} be the set of maximal Ext-consistent subsets of $Sub^+(\varphi)$. For all $\Gamma, \Gamma' \in W_{\mathsf{Int}} \cup W_{\mathsf{Ext}}$, let $\Gamma/B_j = \{\psi \mid B_j\psi \in \Gamma\}$ and $B_j\Gamma = \{B_j\psi \mid B_j\psi \in \Gamma\} \cup \{\neg B_j\psi \mid \neg B_j\psi \in \Gamma\}$.

We define the epistemic model $M = (W, R, V)$ as follows:

- $W = W_{\mathsf{Int}} \cup W_{\mathsf{Ext}}$;
- for all $j \in G$ and $\Gamma, \Gamma' \in W$, $\Gamma' \in R_j(\Gamma)$ iff $\Gamma/B_j = \Gamma'/B_j$ and $\Gamma/B_j \subseteq \Gamma'$;
- $\Gamma \in V(p)$ iff $p \in \Gamma$.

We are going to prove the *truth lemma*, i.e. for all $\psi \in Sub^+(\varphi)$, all $\Gamma \in W$

$$M, \Gamma \models \psi \text{ iff } \psi \in \Gamma$$

We prove it by induction on ψ. The case $\psi = p$ is fulfilled by the definition of the valuation. The cases $\psi = \neg\chi, \psi = \psi_1 \wedge \psi_2$ are fulfilled by the induction hypothesis. It remains to prove the case $\psi = B_j\chi$.

- Assume $\psi \in \Gamma$. Then $\chi \in \Gamma/B_j$. So for all Γ' such that $\Gamma' \in R_j(\Gamma)$, $\chi \in \Gamma'$. So for all Γ' such that $\Gamma' \in R_j(\Gamma)$ $M, \Gamma' \models \chi$ by induction hypothesis. So $M, \Gamma \models B_j\chi$, i.e. $M, \Gamma \models \chi$.

- Assume $M, \Gamma \models B_j\psi$. Then $B_j\Gamma \cup \Gamma/B_j \cup \{\neg\psi\}$ is not Ext-consistent.

  Assume on the contrary that $B_j\Gamma \cup \Gamma/B_j \cup \{\neg\psi\}$ is Ext-consistent. Then there is $\Gamma' \in W_{\mathsf{Ext}}$ such that $B_j\Gamma \cup \Gamma/B_j \cup \{\neg\psi\} \subseteq \Gamma'$. So $\Gamma/B_j = \Gamma'/B_j$ and $\Gamma/B_j \subseteq \Gamma'$. Then $\Gamma' \in R_j(\Gamma)$ and $\neg\psi \in \Gamma'$, i.e. $\Gamma' \in R_j(\Gamma)$ and $M, \Gamma' \models \neg\psi$ by induction hypothesis. So $M, \Gamma \models \neg B_j\psi$, which is impossible by assumption.

  So $B_j\Gamma \cup \Gamma/B_j \cup \{\neg\psi\}$ is not Ext-consistent. Now we consider two cases: first $\Gamma \in W_{\mathsf{Int}}$ and then $\Gamma \in W_{\mathsf{Ext}}$.

  1. $\Gamma \in W_{\mathsf{Int}}$. Then there are $\varphi_1, \ldots, \varphi_n \in \Gamma/B_j$, $\varphi'_1, \ldots, \varphi'_m \in B_j\Gamma$ such that
     $\vdash_{\mathsf{Ext}} \varphi_1 \to (\varphi_2 \to \ldots \to (\varphi_n \to (\varphi'_1 \to (\varphi'_2 \to \ldots \to (\varphi'_m \to \psi))))))$. So
     $\vdash_{\mathsf{Ext}} B_j[\varphi_1 \to (\varphi_2 \to \ldots \to (\varphi_n \to (\varphi'_1 \to (\varphi'_2 \to \ldots \to (\varphi'_m \to \psi))))))]$ by the necessitation rule of Ext. So
     $\vdash_{\mathsf{Ext}} B_j\varphi_1 \to (B_j\varphi_2 \to \ldots \to (B_j\varphi_n \to (B_j\varphi'_1 \to (B_j\varphi'_2 \to \ldots \to (B_j\varphi'_m \to B_j\psi))))))$. i.e. $\vdash_{\mathsf{Ext}} B_j\varphi_1 \to (B_j\varphi_2 \to \ldots \to (B_j\varphi_n \to (\varphi'_1 \to (\varphi'_2 \to \ldots \to (\varphi'_m \to B_j\psi))))))$ because for all i $\vdash_{\mathsf{Ext}} \varphi'_i \leftrightarrow B_j\varphi'_i$. So
     $\vdash_{\mathsf{Int}} B_j\varphi_1 \to (B_j\varphi_2 \to \ldots \to (B_j\varphi_n \to (\varphi'_1 \to \ldots \to (\varphi'_m \to B_j\psi))))$ by axiom scheme (I-E).
     But $B_j\varphi_1, \ldots, B_j\varphi_n, \varphi'_1, \ldots, \varphi'_m \in \Gamma$. So $B_j\psi \in \Gamma$.
  2. $\Gamma \in W_{\mathsf{Ext}}$. Then there are $\varphi_1, \ldots, \varphi_n \in \Gamma/B_j$ and $\varphi'_1, \ldots, \varphi'_m \in B_j\Gamma$ such that
     $\vdash_{\mathsf{Ext}} \varphi_1 \to (\varphi_2 \to \ldots \to (\varphi_n \to (\varphi'_1 \to (\varphi'_2 \to \ldots \to (\varphi'_m \to \psi))))))$. So
     $\vdash_{\mathsf{Ext}} B_j[\varphi_1 \to (\varphi_2 \to \ldots \to (\varphi_n \to (\varphi'_1 \to (\varphi'_2 \to \ldots \to (\varphi'_m \to \psi))))))]$ by the necessitation rule of Ext. So
     $\vdash_{\mathsf{Ext}} B_j\varphi_1 \to (B_j\varphi_2 \to \ldots \to (B_j\varphi_n \to (B_j\varphi'_1 \to (B_j\varphi'_2 \to \ldots \to (B_j\varphi'_m \to B_j\psi))))))$. i.e. $\vdash_{\mathsf{Ext}} B_j\varphi_1 \to (B_j\varphi_2 \to \ldots \to (B_j\varphi_n \to (\varphi'_1 \to (\varphi'_2 \to \ldots \to (\varphi'_m \to B_j\psi))))))$ because for all i $\vdash_{\mathsf{Ext}} \varphi'_i \leftrightarrow B_j\varphi'_i$.
     But $B_j\varphi_1, \ldots, B_j\varphi_n, \varphi'_1, \ldots, \varphi'_m \in \Gamma$. So $B_j\psi \in \Gamma$.

  Finally we have shown that in all cases $B_j\psi \in \Gamma$.

So we have proved the truth lemma. Now we need to prove that the accessibility relations R_j are serial, transitive and euclidean.

- Transitivity. Assume that $\Gamma' \in R_j(\Gamma)$ and $\Gamma'' \in R_j(\Gamma')$. i.e. $\Gamma'/B_j = \Gamma''/B_j$ and $\Gamma'/B_j \subseteq \Gamma''$; and $\Gamma/B_j = \Gamma'/B_j$ and $\Gamma/B_j \subseteq \Gamma'$. Then clearly $\Gamma/B_j = \Gamma''/B_j$ and $\Gamma/B_j \subseteq \Gamma''$. i.e. $\Gamma'' \in R_j(\Gamma)$.

- Euclidicity. Assume that $\Gamma' \in R_j(\Gamma)$ and $\Gamma'' \in R_j(\Gamma)$. i.e. $\Gamma/B_j = \Gamma'/B_j$ and $\Gamma/B_j \subseteq \Gamma'$; and $\Gamma/B_j = \Gamma''/B_j$ and $\Gamma/B_j \subseteq \Gamma''$. Then clearly $\Gamma'/B_j = \Gamma''/B_j$ and $\Gamma'/B_j \subseteq \Gamma''$. i.e. $\Gamma'' \in R_j(\Gamma')$.

- Seriality. We only prove the case $\Gamma \in W_{\mathsf{Int}}$. The case $\Gamma \in W_{\mathsf{Ext}}$ is similar. We are going to show that $B_j\Gamma \cup \Gamma/B_j$ is Ext-consistent.

  Assume the contrary. Then there are $\varphi_1, \ldots, \varphi_n \in \Gamma/B_j$ and $\varphi'_1, \ldots, \varphi'_m \in B_j\Gamma$ such that

  $\vdash_{\mathsf{Ext}} \varphi_1 \to (\varphi_2 \to \ldots \to (\varphi_n \to (\varphi'_1 \to \ldots \to (\varphi'_{m-1} \to \neg\varphi'_m))))$. So

  $\vdash_{\mathsf{Ext}} B_j[\varphi_1 \to (\varphi_2 \to \ldots \to (\varphi_n \to (\varphi'_1 \to \ldots \to (\varphi'_{m-1} \to \neg\varphi'_m))))]$. So

  $\vdash_{\mathsf{Ext}} B_j\varphi_1 \to (B_j\varphi_2 \to \ldots \to (B_j\varphi_n \to (B_j\varphi'_1 \to \ldots \to (B_j\varphi'_{m-1} \to B_j\neg\varphi'_m))))$. So

  $\vdash_{\mathsf{Ext}} B_j\varphi_1 \to (B_j\varphi_2 \to \ldots \to (B_j\varphi_n \to (\varphi'_1 \to \ldots \to (\varphi'_{m-1} \to \neg\varphi'_m))))$.

  But $B_j\varphi_1, \ldots, B_j\varphi_n, \varphi'_1, \ldots, \varphi'_m \in \Gamma$. So $\neg\varphi'_m \in \Gamma$ which is impossible because $\varphi'_m \in \Gamma$.

  Finally $B_j\Gamma \cup \Gamma/B_j$ is Ext-consistent. So there is $\Gamma' \in W_{\mathsf{Ext}}$ such that $B_j\Gamma \cup \Gamma/B_j \subseteq \Gamma'$. i.e. there is $\Gamma' \in W$ such that $\Gamma' \in R_j(\Gamma)$

Finally we prove that for all $\Gamma \in W_{\mathsf{Int}}$, $\Gamma \in R_Y(\Gamma)$ (*).
Let $\Gamma \in W_{\mathsf{Int}}$. For all $B_Y\varphi \in \Gamma$, $\varphi \in \Gamma$ by axiom scheme (T). So $\Gamma/B_j \subseteq \Gamma$. So $\Gamma \in R_Y(\Gamma)$.

φ is a Int-consistent formula so there is $\Gamma \in W_{\mathsf{Int}}$ such that $\varphi \in \Gamma$, i.e. $M, \Gamma \models \varphi$. Let $\mathcal{M}_{\mathsf{Int}}$ be the submodel generated by $R_Y(\Gamma)$. Then clearly $(\mathcal{M}, W_a)$ with $W_a = R_Y(\Gamma)$ is an internal model. Finally, because $\Gamma \in R_Y(\Gamma)$ by (*), there is $\Gamma \in W_a$ such that $\mathcal{M}_{\mathsf{Int}}, \Gamma \models \varphi$. QED

From this axiomatization we can prove other nice properties.

Proposition 2.3.24 *For all $\varphi \in \mathcal{L}$,*

1. $\vdash_{\mathsf{Int}} \varphi$ *iff* $\vdash_{\mathsf{Ext}} B_Y\varphi$;
2. $\vdash_{\mathsf{Int}} \varphi$ *iff* $\vdash_{\mathsf{Int}} B_Y\varphi$.

PROOF. Item 1 comes from Proposition 2.3.20. For item 2, let $\varphi \in \mathcal{L}$. If $\vdash_{\mathsf{Int}} B_Y\varphi$ then $\vdash_{\mathsf{Int}} \varphi$ by axiom (T). Assume now that $\vdash_{\mathsf{Int}} \varphi$. Then $\vdash_{\mathsf{Ext}} B_Y\varphi$ by item 1. Then $\vdash_{\mathsf{Int}} B_Y\varphi$ by axiom (I-E). QED

Finally the internal logic Int has also nice computational properties. Its complexity is the same as in the external approach (see end of Section 2.2.2).

Theorem 2.3.25 (Decidability and complexity of Int)
The internal logic Int is decidable and its validity problem is $PSPACE$-complete for $N \geq 3$.

PROOF.

- The decidability of Int can be proved in two ways. First, because Ext is decidable, Int is also decidable by Proposition 2.3.20. Second, because Int has the finite model property (see proof of Theorem 2.3.23), Int is decidable.

- Because the validity problem is PSPACE-complete for Ext if $N \geq 2$ then the validity problem for Int is in PSPACE by Proposition 2.3.20.

  Besides, as a corollary of the lemma below, we get that the validity problem for Int is PSPACE-complete if $N \geq 3$ because the validity problem for Ext is PSPACE-complete if $N = 2$.

  Lemma 2.3.26 *Assume $\{Y, i, j\} \subseteq G$ and let $\varphi \in \mathcal{L}$ dealing only with agents Y and j. Then,*

  $$\models_{Ext} \varphi \text{ iff } \models_{Int} t(\varphi)$$

  where $t(\varphi)$ is the formula obtained by replacing every occurence of Y by i.

  PROOF. Assume $\varphi \in \mathcal{L}$ dealing only with agents Y and j is externally satisfiable. Then clearly $t(\varphi)$ is also externally satisfiable. Let $M = (W, R, V)$ be an external model generated by $w \in M$ such that $M, w \models t(\varphi)$. Let M' be the epistemic model obtained from M by replacing the accessibility relation R_Y by $R'_Y = \{(v, v); v \in W\}$. Then clearly $M', w \models t(\varphi)$ and (M', w) is a multi-agent possible world. So $t(\varphi)$ is internally satisfiable.

  Finally, if $\models_{Ext} \varphi$ then clearly $\models_{Ext} t(\varphi)$. So $\models_{Int} t(\varphi)$ by axiom I-E. QED

  QED

Remark 2.3.27 Soon after Hintikka's seminal book was published [Hintikka, 1962], an issue now known as the logical omniscience problem was raised by Castañeda about Hintikka's epistemic logic: his "senses of 'knowledge' and 'belief' are much too strong [...] since most people do not even understand all deductions from premises they know to be true" [Castañeda, 1964]. It sparked a lot of work aimed at avoiding this problem (such as [Levesque, 1984], [Fagin and Halpern, 1988] or [Duc, 2001]).

While we believe that it is indeed a problem when we want to model or describe human-like agents, we nevertheless believe that it is not really a problem for artificial agents. Indeed, these agents are supposed to reason according to our internal logic and because of its decidability, artificial agents are in fact logically omniscient (even if it will take them some time to compute all the deductions). □

2.4 Conclusion

In this chapter, we have first identified what we claim to be the only three possible modeling approaches by proceeding by successive dichotomies. Afterwards, we have focused on the internal approach for which a logical formalism is missing in a multi-agent setting, although such a formalism is crucial if we want to design autonomous agents. We have proposed one by generalizing the possible world semantics of AGM belief revision theory. This formalism enabled us to draw some formal links between the external and the internal approach which

are in line with our intuitions of these two approaches. Finally, we have provided an axiomatization of our formalism whose axioms are also in line with our intuitions of the internal approach.

So far we have described epistemic situations from a static point of view. In the next chapter we are going to add dynamics to the picture and study belief change from a logical point of view.

Chapter 3

Dynamic epistemic logic

3.1 Introduction

Dynamic epistemic logic is concerned with the logical study in a multi-agent setting of knowledge and belief change, and more generally about information change. These changes can be due to events that change factual properties of the actual world [van Ditmarsch *et al.*, 2005; Kooi, 2007]: for example a coin is publicly (or privately) flipped over. But what is mostly studied in dynamic epistemic logic are events that do not change factual properties of the world (they are called epistemic events) but that nevertheless bring about changes of (higher-order) beliefs: for example a coin is revealed publicly (or privately) to be heads up.

Dynamic epistemic logic is a young field of research. Some of its predecessors are van Benthem [van Benthem, 1989] and Moore in artificial intelligence [Moore, 1985]. But it really started with Plaza's logic of public announcement [Plaza, 1989]. Independently, Gerbrandy and Groeneveld proposed a system dealing moreover with private announcement [Gerbrandy and Groeneveld, 1997; Gerbrandy, 1999] and that was inspired by the work of Veltman [Veltman, 1996]. Another system was proposed by van Ditmarsch whose main inspiration was the Cluedo game [van Ditmarsch, 2000; van Ditmarsch, 2002] and which is axiomatized in [van Ditmarsch *et al.*, 2003]. It was specifically developed for the logic S5 supposed to model the notion of knowledge. But the most influential and original system was the BMS system proposed by Baltag, Moss and Solecki [Baltag *et al.*, 1998], [Baltag and Moss, 2004]. This system can deal with all the types of situations studied in the works above and provides a general approach to the topic. So we will focus on this system in this chapter.

First we will recall the BMS system, together with new results concerning the preservation of seriality of the accessibility relations for belief during the update. Afterwards, we will present an 'internalization' of this system done in the same spirit as in Chapter 2. Finally, we will set some connections between the external and the internal approach in this dynamic setting, still very much in line with what we did in Chapter 2.

3.2 The BMS system

3.2.1 State of the art

Event model

Just as all the systems proposed in dynamic epistemic logic, the BMS system takes epistemic logic as a starting point. Epistemic logic is used to model how the agents perceive the actual world in terms of beliefs about the world and about the other agents' beliefs. The insight of the BMS approach is that one can describe how an event is perceived by the agents in a very similar way. Indeed, the agents' perception of an event can also be described in terms of beliefs: for example, while the quizmaster tells Bob that the coin is heads up (event a_a) Ann *believes* that nothing happens (event b). This leads them to define the notion of event model whose definition is very similar to that of an epistemic model.

Definition 3.2.1 (Event model)
An *event model* A is a triple $A = (E, R, Pre)$ such that

- E is a finite and non-empty set of possible events;
- $R : G \to 2^{E \times E}$ assigns an accessibility relation to each agent;
- $Pre : E \to \mathcal{L}^C$ assigns an epistemic formula to each possible event (where $\mathcal{L}^C$ is defined in Definition 2.2.10).

If $A = (E, R, Pre)$ is an event model, a pair (A, a_a) where $a_a \in E$ is called a *pointed event model*. We also write $R_j = R(j)$ and $R_j(a) = \{b \in E \mid aR_jb\}$, and $a \in A$ for $a \in E$. □

The main difference with the definition of an epistemic model is that we no longer have a valuation V but instead a function Pre. This function is supposed to specify under which condition an event can physically take place in a possible world.

Example 3.2.2 ('Coin' example)

1. Assume that the quizmaster announces privately to Bob that the coin is heads up. This event is depicted in Figure 3.1. a_a stands for the event 'the quizmaster truthfully announces that the coin is heads up' and b stands for the event 'nothing happens'. The boxed event corresponds to the actual event. So while the quizmaster announces to Bob that the coin is heads up (a_a), Ann believes that nothing happens (b): this explains the accessibility relation indexed by A between a_a and b. The precondition for a_a is that the coin is indeed heads up (p) while the precondition for b is any tautology (like $\top$) because the event where nothing happens can take place in any world.
2. Assume that the quizmaster announces publicly that the coin is heads up. This event is depicted in Figure 3.2. There, a_a stands for 'the quizmaster truthfully announces that the coin is heads up'. Because this event is correctly perceived by Ann and Bob, a_a is the only event considered possible by them. Finally, for this announcement to be made in a possible world, the coin has to be heads up in this world (p).

□

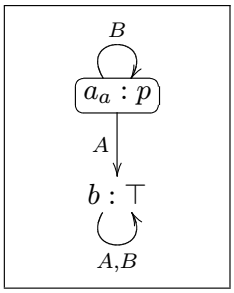


Figure 3.1: Private announcement of p to Bob

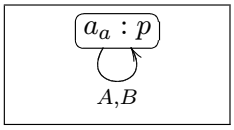


Figure 3.2: Public announcement of p

Product Update

Now, in reality after (or during) this event e, the agents update their beliefs by taking into account these two pieces of information: the event e and the initial situation s. This gives rise to a new situation $s \times e$. This actual update is rendered formally by the following mathematical update product between a pointed epistemic model and a pointed event model.

Definition 3.2.3 (Update product)
Let $M = (W, R, V, w_a)$ be a pointed epistemic model and $A = (E, R, Pre, a_a)$ a pointed event model such that $M, w_a \models Pre(a_a)$. We define their *update product* to be the pointed epistemic model $M \otimes A = (W \otimes E, R', V', w'_a)$ where

1. $W \otimes E = \{(w, a) \in W \times E \mid M, w \models Pre(a)\}$;
2. $(v, b) \in R'_j(w, a)$ iff $v \in R_j(w)$ and $b \in R_j(a)$;
3. $V'(p) = \{(w, a) \in W \otimes E \mid w \in V(p)\}$;
4. $w'_a = (w_a, a_a)$.

□

***Intuitive interpretation*:**

1. The possible worlds that we consider after the update are all the ones resulting from the performance of one of the events in one of the worlds, under the assumption that the event can physically take place in the corresponding world (assumption expressed by the function Pre).

2. The components of our event model are 'simple' events (in the sense of BMS, see [Baltag and Moss, 2004] for more details). In particular this means that the uncertainty about the situation is independent from the uncertainty about the event. This independence allows us to 'multiply' these uncertainties to compute the new accessibility (or uncertainty) relation.

3. The definition of the valuation exemplifies the fact that our events do not change facts. (That is why we call them *epistemic* events, as already said above.)

4. Finally, we naturally assume that the actual event can indeed take place in the actual world.

Example 3.2.4 ('Coin' example)

1. Assume that the quizmaster has privately announced to Bob that the coin is heads up. The resulting situation is depicted in the pointed epistemic model of Figure 3.3. This pointed epistemic model is obtained by updating the original situation depicted in Figure 2.1 with the pointed event model depicted in Figure 3.1. (Note that this pointed epistemic model is the same as the external model of Figure 2.6 in Example 2.3.18.) As we said in Example 2.3.18, in this resulting epistemic model, Bob knows that the coin is heads up but Ann does not know whether it is either heads or tails up and believes Bob does not know either.

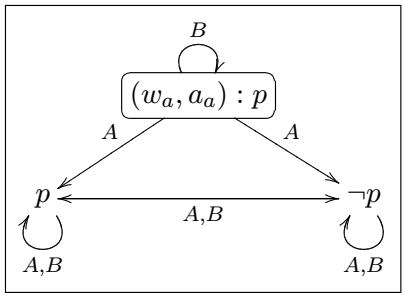


Figure 3.3: Situation after the private announcement to B that p is true

2. Assume that the quizmaster has publicly announced that the coin is heads up. The resulting situation is depicted in Figure 3.4. This pointed epistemic model is obtained by updating the original situation depicted in Figure 2.1 with the pointed event model depicted in Figure 3.2. In this resulting epistemic model, it is common knowledge that the coin is heads up.

□

Language and axiomatization

Of course, it seems natural to extend the language of epistemic logic to incorporate this new dynamic feature. To do so, we inspire ourselves with the programs of Propositional Dynamic Logic [Pratt, 1976] and introduce a modality $[a]$.

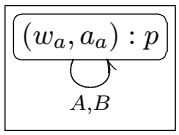


Figure 3.4: Situation after the public announcement that p is true

Definition 3.2.5 (Language $\mathcal{L}_A$)
Let A be an event model. The language $\mathcal{L}_A$ is defined inductively as follows.

$$\mathcal{L}_A : \varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \varphi \mid B_j\varphi \mid [a]\varphi$$

where p ranges over the set of propositional letters Φ, j over G and a over the possible events of A. □

Note that the event model A, which a priori is part of the semantics, is given in the very definition of the syntax of the language. In fact, in the BMS formalism the formula $[a]\varphi$ is even written $[A, a]\varphi$. The truth condition for $[a]\varphi$ is defined as follows.

Definition 3.2.6 (Truth conditions for $\mathcal{L}_A$)
Let $(M, w) = (W, R, V, w)$ be a pointed epistemic model.

$$M, w \models [a]\varphi \text{ iff (if } M, w \models Pre(a) \text{ then } M \otimes A, (w, a) \models \varphi).$$

□

So, according to this truth condition, $[a]\varphi$ should be read 'after every execution of the possible event a, φ holds'. Let C be a class of epistemic models and $\varphi \in \mathcal{L}_A$. We say that φ is C-valid, written $\models_{\mathsf{C}} \varphi$, iff for all epistemic model $M \in$ C and all $w \in M$, $M, w \models \varphi$. We also write $\models \varphi$ for $\models_{K_G} \varphi$.

Now we can axiomatize the semantics just defined by the following logic called BMS.

Definition 3.2.7 (Proof system of BMS)
The logic BMS is defined by the proof system of $\mathsf{K_G}$ together with the following axiom schemes and inference rules:

R1	$\vdash_{\mathsf{BMS}} [a]p \leftrightarrow (Pre(a) \to p)$	
R2	$\vdash_{\mathsf{BMS}} [a]\neg\varphi \leftrightarrow (Pre(a) \to \neg[a]\varphi)$	
R3	$\vdash_{\mathsf{BMS}} Pre(a) \to ([a]B_j\varphi \leftrightarrow B_j[a_1]\varphi \wedge \ldots \wedge B_j[a_n]\varphi)$ where $a_1, \ldots, a_n$ is the list of b such that $b \in R_j(a)$	
Nec	If $\vdash_{\mathsf{BMS}} \varphi$ then $\vdash_{\mathsf{BMS}} [a]\varphi$	(Necessitation)
Distr	$\vdash_{\mathsf{BMS}} [a](\varphi \to \psi) \to ([a]\varphi \to [a]\psi)$	(Distribution)

□

R1, R2, R3 are called reduction axioms. They enable to prove that any formula of $\mathcal{L}_A$ is logically equivalent to a formula of $\mathcal{L}$. Then we get that BMS is complete with respect to the class of $\mathsf{K_G}$-models (for $\mathcal{L}_A$) thanks to the fact that $\mathsf{K_G}$ is complete with respect to the class of $\mathsf{K_G}$-models (for $\mathcal{L}$).

Theorem 3.2.8 [Baltag and Moss, 2004](Soundness and completeness)
Let A be a fixed event model. For all $\varphi \in \mathcal{L}_A$,

$$\vdash_{BMS} \varphi \textit{ iff } \models_{K_G} \varphi.$$

3.2.2 On seriality preservation

We did not assume any particular property for the accessibility relations of event models and epistemic models, such as seriality or transitivity. But we could perfectly add them. Then we could wonder which properties are preserved by the update product, i.e. in case the accessibility relations of the epistemic model and the event model satisfy a property, do the accessibility relations of the updated model satisfy the same property? We know that reflexivity, transitivity and euclidicity are preserved [Baltag and Moss, 2004]. However, seriality is not preserved. For example, if we update the epistemic model depicted in Figure 3.3 by the public announcement that Bob believes that the coin is heads up (formally $B_B p$) then we get the epistemic model depicted on the right of Figure 3.5 where Ann's accessibility relation is not serial.

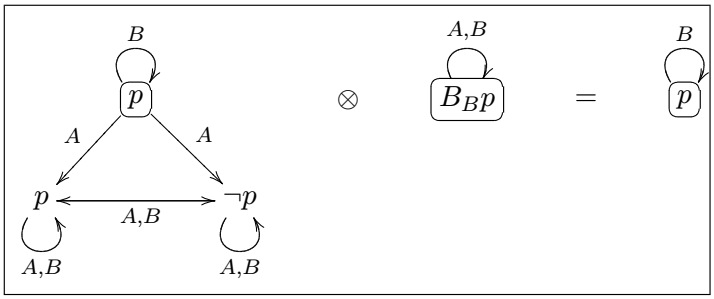


Figure 3.5: Failure of seriality preservation

We are now going to study under which conditions seriality is preserved. We will split our account in two parts. First, we will investigate under which conditions the entire updated model is serial. Second, we will investigate under which conditions a generated submodel of the entire updated model is serial.

Seriality preservation for the entire BMS product

First of all, for a given epistemic model M and a given event model A, we say that the update product $M \otimes A$ is *defined* if there is $w \in M$ and $a \in A$ such that $M, w \models Pre(a)$. We introduce this definition because seriality of updated models makes sense only for defined updated models.

Proposition 3.2.9 *Let A be a serial event model[1] and let M be an epistemic model. Then*

$M \otimes A$ is defined and serial iff $M \models O\left(\bigvee_{a\in A} Pre(a)\right) \wedge U \bigwedge_{a\in A}\left(Pre(a) \rightarrow \bigwedge_{j\in G} \hat{B}_j \bigvee_{b\in R_j(a)} Pre(b)\right)$.[2]

PROOF. $M \models O\left(\bigvee_{a\in A} Pre(a)\right)$ clearly means that the model $M\otimes A$ is defined. Now it remains to prove that $M \otimes A$ is serial iff $M \models U \bigwedge_{a\in A}\left(Pre(a) \rightarrow \bigwedge_{j\in G} \hat{B}_j \bigvee_{b\in R_j(a)} Pre(b)\right)$.

- Assume that $M \models U \bigwedge_{a\in A}\left(Pre(a) \rightarrow \bigwedge_{j\in G} \hat{B}_j \bigvee_{b\in R_j(a)} Pre(b)\right)$ (*). Let $(w, a) \in M \otimes A$ and $j \in G$. Then $M, w \models Pre(a)$. So $M, w \models \bigwedge_{j\in G} \hat{B}_j \bigvee_{b\in R_j(a)} Pre(b)$ by (*). Then $M, w \models \hat{B}_j \bigvee_{b\in R_j(a)} Pre(b)$. So there is $v \in R_j(w)$ and $b \in R_j(a)$ such that $M, v \models Pre(b)$. Then there is $(v, b) \in M \otimes A$ such that $(v, b) \in R_j(w, a)$ by definition of $M \otimes A$. So $M \otimes A$ is serial.
- Assume that $M \not\models U \bigwedge_{a\in A}\left(Pre(a) \rightarrow \bigwedge_{j\in G} \hat{B}_j \bigvee_{b\in R_j(a)} Pre(b)\right)$. Then there is $w \in M$ and $a \in A$ such that $M, w \models Pre(a) \wedge \left(\bigvee_{j\in G} B_j \bigwedge_{b\in R_j(a)} \neg Pre(b)\right)$. Then there is $j \in G$ such that $M, w \models B_j \bigwedge_{b\in R_j(a)} \neg Pre(b)$ (**). So $(w, a) \in M \otimes A$ but there is no $v \in R_j(w)$ and $b \in R_j(a)$ such that $(v, b) \in R_j(w, a)$. Indeed, otherwise we would have $M, w \models \hat{B}_j \bigvee_{b\in R_j(a)} Pre(b)$, which contradicts (**). So $M \otimes A$ is not serial.

QED

From now on, we write $\mathcal{P}(A) = O\left(\bigvee_{a\in A} Pre(a)\right) \wedge U \bigwedge_{a\in A}\left(Pre(a) \rightarrow \bigwedge_{j\in G} \hat{B}_j \bigvee_{b\in R_j(a)} Pre(b)\right)$.

$O\left(\bigvee_{a\in A} Pre(a)\right)$ expresses that the updated model $M \otimes A$ is defined.

$U \bigwedge_{a\in A}\left(Pre(a) \rightarrow \bigwedge_{j\in G} \hat{B}_j \bigvee_{b\in R_j(a)} Pre(b)\right)$ expresses that the updated model $M \otimes A$ is serial. Note that M does not need to be serial for the proposition to hold. But of course if M is serial then the proposition still holds. From this proposition we can easily prove the following corollary.

Corollary 3.2.10 *Let C be a class of epistemic models and A a serial event model.*

[1] i.e. all its accessibility relations are serial

[2] The existential modality O and the universal modality U were defined in Section 2.2.4.

$$\models_{\mathsf{C}} \neg\mathcal{P}(A)$$
iff there is no epistemic model $M \in \mathsf{C}$ such that $M \otimes A$ is defined and serial.

In other words this corollary tells us under which condition, for a given event model A, whatever epistemic model M we chose, $M \otimes A$ will not be defined or not serial. If this condition is fulfilled that would mean intuitively that in any epistemic situation, if the event (corresponding to this event model) is performed, then afterwards in any case (some of) the agents' beliefs are inconsistent. This is of course counter intuitive and we should then avoid such kinds of event (models).

Now we are going to give an example of a family of event models A where such a phenomenon occurs (i.e. there is no M such that $M \otimes A$ is defined and also serial).

Proposition 3.2.11 *Let $\varphi = \psi \wedge B_i\neg\psi$ for some $\psi \in \mathcal{L}$, and let A be the event model corresponding to the public announcement of φ. Then there is no $\mathsf{KD45}_G$-model M such that $M \otimes A$ is defined and serial.*

PROOF. Thanks to Corollary 3.2.10 it suffices to prove that $\models_{\mathsf{KD45}_G} \neg\mathcal{P}(A)$ i.e. $\models_{\mathsf{KD45}_G} O\varphi \rightarrow O\left(\varphi \wedge \bigvee_{j\in G} B_j\neg\varphi\right)$ because $\mathcal{P}(A) = O\varphi \wedge U\left(\varphi \rightarrow \bigwedge_{j\in G} \hat{B}_j\varphi\right)$. Let M be a $\mathsf{KD45}_G$-model such that $M \models O\varphi$. Let $w \in M$ such that $M, w \models \varphi$. One can easily prove that $\models_{\mathsf{KD45}_G} \varphi \rightarrow B_i\neg\varphi$. So $M, w \models \varphi \wedge B_i\neg\varphi$. Then $M, w \models \varphi \wedge \bigvee_{j\in G} B_j\neg\varphi$ i.e. $M \models O\left(\varphi \wedge \bigvee_{j\in G} B_j\neg\varphi\right)$. QED

We can compare this result with the notion of selfrefuting formula studied in [van Ditmarsch and Kooi, 2006]. Selfrefuting formulas are formulas that are no longer true after they are publicly announced. An example of such formulas is Moore's sentence $p \wedge \neg B_j p$: if it is announced then p becomes common knowledge and in particular $B_j p$ becomes true. Here our formulas are a bit different: after they are announced agent j's beliefs become inconsistent.

Seriality preservation for generated submodels

The results above have certainly a logical interest. But, in practice, the updated models we are really interested in are generated submodels of the entire updated model. Indeed, by definition, multi-agent possible worlds (M, w) are *generated* by w and internal models $(\mathcal{M}, W_a)$ are *generated* by W_a. So, in an updated model (which could be composed of several disjoint generated submodels), we would like to know under which conditions a particular generated submodel of the entire updated model is serial, and not necessarily the entire updated model. That is what we are going to investigate now. We start with a formal definition.

Definition 3.2.12 Let A be an event model, $a \in A$ and $n \in \mathbb{N}$. We define $\delta^n(a)$ inductively as follows.

- $\delta^0(a) = Pre(a)$;

- $\delta^{n+1}(a) = \delta^0(a) \wedge \bigwedge_{j\in G} \hat{B}_j \bigvee_{b\in R_j(a)} \delta^n(b) \wedge \bigwedge_{j\in G} B_j \bigwedge_{b\in R_j(a)} (Pre(b) \to \delta^n(b))$.

□

Intuitively, $M, w \models \delta^n(a)$ means that the submodel of $M \otimes A$ generated by (w, a) is defined and serial up to modal depth n. This interpretation is endorsed by the following two lemmas which will be used to prove the main proposition.

Lemma 3.2.13 *Let M be an epistemic model and let A be an event model. For all $w \in M$, $a \in A$, $n \in \mathbb{N}$,*

$M, w \models \delta^{n+1}(a)$ iff
$M, w \models \delta^1(a)$ and for all $v \in M$ such that $w = w_0 R_{j_1} w_1 R_{j_2} \dots R_{j_n} w_n = v$ such that there are $a = a_0 R_{j_1} a_1 R_{j_2} \dots R_{j_n} a_n = b$ such that for all $i \in \{0, \dots, n\}$, $M, w_i \models Pre(a_i)$,

$$M, v \models \bigwedge_{j\in G} \hat{B}_j \bigvee_{c\in R_j(b)} Pre(c)$$

PROOF. We prove it by induction on n. The case $n = 0$ is clear. We prove the induction step. Assume the property is true for n.

- Assume $M, w \models \delta^{n+2}(a)$. Then $M, w \models \delta^1(a)$ because $\delta^{n+1}(b) \to Pre(b)$ and $\delta^1(a) = Pre(a) \wedge \bigwedge_{j\in G} \hat{B}_j \bigvee_{b\in R_j(a)} Pre(b)$. Let $v \in M$ such that $w = w_0 R_{j_1} w_1 R_{j_2} \dots R_{j_{n+1}} w_{n+1} = v$ and such that there are $a = a_0 R_{j_1} a_1 R_{j_2} \dots R_{j_{n+1}} a_{n+1} = b$ such that for all $i \in \{0, \dots, n+1\}$, $M, w_i \models Pre(a_i)$.

  By assumption, $M, w \models \bigwedge_{j\in G} B_j \bigwedge_{b\in R_j(a)} (Pre(b) \to \delta^{n+1}(b))$. So $M, w_1 \models \bigwedge_{b\in R_{j_1}(a)} (Pre(b) \to \delta^{n+1}(b))$. Besides $a_1 \in R_{j_1}(a)$ and $M, w_1 \models Pre(a_1)$. So $M, w_1 \models \delta^{n+1}(a_1)$.

  Then, by induction hypothesis, for all v' such that $w_1 = w'_1 R_{j_2} \dots R_{j_{n+1}} w'_{n+1} = v'$ such that there are $a_1 = a'_1 R_{j_2} \dots R_{j_{n+1}} a'_{n+1} = a'$ such that for all i, $M, w'_i \models Pre(a'_i)$,

  $$M, v' \models \bigwedge_{j\in G} \hat{B}_j \bigvee_{b'\in R_j(a')} Pre(b').$$

  So $M, v \models \bigwedge_{j\in G} \hat{B}_j \bigvee_{c\in R_j(b)} Pre(c)$

- Assume $M, w \models \delta^1(a)$ and assume that for all $v \in M$ such that $w = w_0 R_{j_1} \dots R_{j_n} w_{n+1} = v$ such that there are $a = a_0 R_{j_1} \dots R_{j_n} a_{n+1} = b$ such that for all i, $M, w_i \models Pre(a_i)$,

$$M, v \models \bigwedge_{j \in G} \hat{B}_j \bigvee_{c \in R_j(b)} Pre(c).$$

Now, assume $M, w \not\models \delta^{n+2}(a)$.

Then $M, w \models \neg Pre(a) \vee \left(\bigvee_{j \in G} B_j \bigwedge_{b \in R_j(a)} \neg \delta^{n+1}(b) \right) \vee \left(\bigvee_{j \in G} \hat{B}_j \bigvee_{b \in R_j(a)} (Pre(b) \wedge \neg \delta^{n+1}(b)) \right)$.

- $M, w \models \neg Pre(a)$ is impossible by assumption.
- Assume $M, w \models \bigvee_{j \in G} B_j \bigwedge_{b \in R_j(a)} \neg \delta^{n+1}(b)$. Then for some $i \in G$, $M, w \models B_i \bigwedge_{b \in R_i(a)} \neg \delta^{n+1}(b)$.
  Then for all $v \in R_i(w)$ and all $b \in R_i(a)$, $M, v \models \neg \delta^{n+1}(b)$ (*).
  But by assumption $M, w \models \delta^1(a)$, i.e. $M, w \models Pre(a) \wedge \bigwedge_{j \in G} \hat{B}_j \bigvee_{b \in R_j(a)} Pre(b)$. Then $M, w \models \hat{B}_i \bigvee_{b \in R_i(a)} Pre(b)$, i.e. there is $v \in R_i(w)$ and $b \in R_i(a)$ such that $M, v \models Pre(b)$ (1).
  So $M, v \models \bigwedge_{j \in G} \hat{B}_j \bigvee_{b \in R_j(a)} Pre(b)$ (2) by assumption (take $w_1 = \ldots = w_n = v$ and $a_1 = \ldots = a_n = b$).
  Then by (1) and (2) we get $M, v \models \delta^1(b)$.
  Besides, by assumption and because wR_iv and aR_ib, for all u such that $v = v_0 R_{j_1} \ldots R_{j_n} u$ such that there are $b = b_0 R_{j_1} \ldots R_{j_n} b_n = c$ such that for all i $M, v_i \models Pre(b_i)$

  $$M, u \models \bigwedge_{j \in G} \hat{B}_j \bigvee_{d \in R_j(c)} Pre(d).$$

  So $M, v \models \delta^{n+1}(b)$ by induction hypothesis. This is impossible by (*).
- Assume $M, w \models \bigvee_{j \in G} \hat{B}_j \bigvee_{b \in R_j(a)} (Pre(b) \wedge \neg \delta^{n+1}(b))$.
  Then there is $i \in G$, $v \in R_i(w)$ and $b \in R_i(a)$ such that $M, v \models Pre(b) \wedge \neg \delta^{n+1}(b)$.
  By the same argument as above we get to a contradiction.

So finally $M, w \models \delta^{n+2}(a)$.

QED

Lemma 3.2.14 *Let M be a finite epistemic model and A be a finite serial event model. Let $n = |M| \cdot |A|$.*[3] *For all $w \in M$ and $a \in A$ such that $M, w \models Pre(a)$,*

1. *$R_j(w, a) \neq \emptyset$ for all $j \in G$ iff $M, w \models \bigwedge_{j \in G} \hat{B}_j \bigvee_{b \in R_j(a)} Pre(b)$;*

[3] $|M|$ is the number of possible worlds of M and $|A|$ is the number of possible events of A.

2. $(v,b) \in \left(\bigcup_{j\in G} R_j\right)^+ (w,a)$ *iff there are* $w = w_0 R_{j_1} w_1 R_{j_2} \dots R_{j_n} w_{n-1} = v$ *and* $a = a_0 R_{j_1} a_1 R_{j_2} \dots R_{j_n} a_{n-1} = b$ *such that for all* i, $M, w_i \models Pre(a_i)$.

PROOF.

1. Assume $M, w \models \bigwedge_{j\in G} \hat{B}_j \bigvee_{b\in R_j(a)} Pre(b)$. Then for all $j \in G$, there is $v \in R_j(w)$ and $b \in R_j(a)$ such that $M, v \models Pre(b)$. Then, by definition of the product update, for all j, there is $(v,b) \in M \otimes A$ such that $(v,b) \in R_j(w,a)$. So for all $j \in G$, $R_j(w,a) \neq \emptyset$.

   Assume $M, w \not\models \bigwedge_{j\in G} \hat{B}_j \bigvee_{b\in R_j(a)} Pre(b)$. Then there is $j \in G$ such that for all $v \in R_j(w)$ and for all $b \in R_j(a)$, $M, v \not\models Pre(b)$. Then, by definition of the product update, there is no $(v,b) \in M \otimes A$ such that $(v,b) \in R_j(w,a)$. So $R_j(w,a) = \emptyset$ for some $j \in G$.

2. $M \otimes A$ is of cardinality at most n due to our hypothesis that $n = |M| \cdot |A|$. So every world $(v,b) \in M \otimes A$ such that $(v,b) \in \left(\bigcup_{j\in G} R_j\right)^+ (w,a)$ is accessible from (w,a) in at most $n-1$ steps. So,

   $(v,b) \in \left(\bigcup_{j\in G} R_j\right)^+ (w,a)$ iff

   there are $j_1, \dots, j_{n-1}$ and $(w_1,a_1), \dots, (w_{n-1}, a_{n-1}) \in M \otimes A$ such that

   $(w,a) R_{j_1} (w_1,a_1) R_{j_2} \dots R_{j_{n-1}} (w_{n-1}, a_{n-1}) = (v,b)$ iff

   there are $w = w_0 R_{j_1} w_1 R_{j_2} \dots R_{j_{n-1}} w_{n-1} = v$ and $a = a_0 R_{j_1} a_1 R_{j_2} \dots R_{j_{n-1}} a_{n-1} = b$ such that for all i, $M, w_i \models Pre(a_i)$.

QED

Proposition 3.2.15 *Let* M *be a finite epistemic model and let* A *be a finite serial event model. Let* $w \in M$, $a \in A$ *and* $n = |M| \cdot |A|$.

The submodel of $M \otimes A$ *generated by* (w,a) *is defined and serial iff* $M, w \models \delta^n(a)$.

PROOF. First, note that the submodel of $M \otimes A$ generated by (w,a) is defined and serial iff

- (w,a) is defined;
- $R_j(w,a) \neq \emptyset$ for all $j \in G$;
- $R_j(v,b) \neq \emptyset$ for all $(v,b) \in \left(\bigcup_{j\in G} R_j\right)^+ (w,a)$ and for all $j \in G$.

Then we get easily the expected result by Lemma 3.2.14 and Lemma 3.2.13. Indeed, (w,a) is defined and $R_j(w,a) \neq \emptyset$ for all $j \in G$ amounts to say that $M, w \models \delta^1(a)$. And $R_j(v,b) \neq \emptyset$ for all $(v,b) \in \left(\bigcup_{j\in G} R_j\right)^+ (w,a)$ and for all $j \in G$ amounts to say that for all $v \in M$ such that $w = w_0 R_{j_1} w_1 R_{j_2} \dots R_{j_n} w_n = v$ such that there are $a = a_0 R_{j_1} a_1 R_{j_2} \dots R_{j_n} a_n = b$ such that for all $i \in \{0,\dots,n\}$, $M, w_i \models Pre(a_i)$, $M, v \models \bigwedge_{j\in G} \hat{B}_j \bigvee_{c\in R_j(b)} Pre(c)$. QED

This proposition is coherent with our interpretation of $M, w \models \delta^n(a)$. As we said, intuitively, $M, w \models \delta^n(a)$ means that the submodel of $M \otimes A$ generated by (w,a) is (defined and) serial up to modal depth n. So, if n is larger than the modal depth of the submodel $M \otimes A$ generated by (w,a) (which is the case if $n = |M| \cdot |A|$) then all the worlds accessible from (w,a) are serial. So this generated submodel is indeed serial. Accordingly, this also entails that it should be serial for any given modal depth. That is what the following property expresses.

Proposition 3.2.16 *Let M be a finite epistemic model and let A be a finite and serial event model. Let $w \in M, a \in A$ and $n = |M| \cdot |A|$.*

If $M, w \models \delta^n(a)$ then for all $m \geq n$, $M, w \models \delta^m(a)$.

PROOF. The proof follows from Lemma 3.2.13 and the fact that for all $v \in M$ there are $w_1, \dots, w_{n-1}$ such that $w = w_0 R_{j_1} w_1 R_{j_2} \dots R_{j_n} w_n = v$ iff there are $w_1, \dots, w_{m-1}$ such that $w = w_0 R_{j_1} w_1 R_{j_2} \dots R_{j_m} w_m = v$. QED

Similarly, if a submodel of $M \otimes A$ generated by (w,a) is serial up to a given modal depths d then it should also be serial up to all modal depth smaller than d. The following proposition proves that it is indeed the case.

Proposition 3.2.17 *For all event models A and $a \in A$, if $n \geq n'$ then $\models \delta^n(a) \to \delta^{n'}(a)$.*

PROOF. Let A be an event model and $a \in A$. We prove it by induction on n. If $n = 0$ or $n = 1$ then the result trivially holds. Assume it is true for a given $n \geq 1$. Assume $\models \delta^{n+1}(a)$, i.e. $\models \delta^0(a) \wedge \bigwedge_{j\in G} \hat{B}_j \bigvee_{b\in R_j(a)} \delta^n(b) \wedge \bigwedge_{j\in G} B_j \bigwedge_{b\in R_j(a)} (Pre(b) \to \delta^n(b))$.

By induction hypothesis, for all $b \in A$, $\models \delta^n(b) \to \delta^{n-1}(b)$. So

$$\models \left(\delta^0(a) \wedge \bigwedge_{j\in G} \hat{B}_j \bigvee_{b\in R_j(a)} \delta^n(b) \wedge \bigwedge_{j\in G} B_j \bigwedge_{b\in R_j(a)} (Pre(b) \to \delta^n(b))\right) \to$$

$$\left(\delta^0(a) \wedge \bigwedge_{j\in G} \hat{B}_j \bigvee_{b\in R_j(a)} \delta^{n-1}(b) \wedge \bigwedge_{j\in G} B_j \bigwedge_{b\in R_j(a)} (Pre(b) \to \delta^{n-1}(b))\right).$$

i.e. $\models \delta^{n+1}(a) \to \delta^n(a)$. So for all $n' \leq n+1$, $\models \delta^{n+1}(a) \to \delta^{n'}(a)$ by induction hypothesis. QED

Finally, we can strike some relationship between the seriality conditions for the entire updated model and for the generated submodels of the entire updated model. Indeed, one can notice that the entire updated model is serial if and only if all its generated submodels are serial. But in fact, because we consider *all* the generated submodels, it suffices that these generated submodels be serial only up to modal depth 1. That is actually the intuition that led to the definition of $\mathcal{P}(A)$.

Proposition 3.2.18 *Let M be an epistemic model and let A be a serial event model. Then,*

$$M \models \mathcal{P}(A) \leftrightarrow O\left(\bigvee_{a\in A} Pre(a)\right) \wedge U \bigwedge_{a\in A} \left(Pre(a) \rightarrow \delta^1(a)\right).$$

$O\left(\bigvee_{a\in A} Pre(a)\right)$ expresses that the updated model is defined. The rest of the formula expresses its seriality. Note that $\delta^1(a) = Pre(a) \wedge \bigwedge_{j\in G} \hat{B}_j \bigvee_{b\in R_j(a)} Pre(b)$, so we have rediscovered the definition of $\mathcal{P}(A)$.

3.3 Dynamizing the internal approach

One can easily notice that the BMS system, just as standard epistemic logic, is designed for the external approach. So to complete the work of Chapter 2, it would be nice to propose an internal version of the BMS system. We would then get a dynamic epistemic logic for the internal approach. That is what we are going to do in this section.

First, we will propose an internal version of event models. Then we will propose two update products, one for each type of models and we will show that these products are in fact equivalent in some sense. Finally, we will give conditions under which an updated model is an internal model (on the basis of the results of the preceding section).

3.3.1 Multi-agent possible event and internal event model

As we said, in the BMS system, events are represented very similarly to epistemic situations. This similarity of the formalism in the external approach can naturally be transferred to the internal approach as well. So the way we represented agent Y's perception of the surrounding world can easily be adapted to represent her perception of events as well. This leads us to the following definitions.

Definition 3.3.1 (Multi-agent possible event)
A *multi-agent possible event* (A, a) is a *finite* pointed event model $A = (E, R, Pre, w)$ generated[4] by $a \in E$ such that R_j is serial, transitive and euclidean for all $j \in G$, and

1. $R_Y(a) = \{a\}$;

[4]An event model A is generated from S if the restriction of A to $\{(\bigcup_{j\in G} R_j)^*(a) \mid a \in S\}$ is A itself. This definition is completely in line with Definition 2.2.13

2. there is no b and $j \neq Y$ such that $a \in R_j(b)$.

□

The motivations for this definition are completely similar to those for the notion of multi-agent possible world, so we do not spell them out here. Likewise, we can also define the notions of internal event model of type 1, of type 2, and the notion of internal event model of type 2 associated to an internal model of type 1. The motivations are still very similar to that of the static case.

Definition 3.3.2 (Internal event model of type 1)
An *internal event model of type 1* is a disjoint and finite union of multi-agent possible events. □

Definition 3.3.3 (Internal event model of type 2)
An *internal event model of type 2* $(\mathcal{A}, A_a)$ is a finite event model $\mathcal{A} = (E, R, Pre, A_a)$ generated by $A_a \subseteq E$ such that R_j is serial, transitive and euclidean for all $j \in G$, and $R_Y(a_a) = A_a$ for all $a_a \in A_a$. A_a is called the *actual equivalence class*. □

Definition 3.3.4 Let $(\mathcal{A}, A_a) = \{(A^1, a^1), \ldots, (A^m, a^m)\}$ be an internal event model of type 1 (with $A^k = (E^k, R^k, Pre^k)$). The *internal event model of type 2 associated to* $(\mathcal{A}, A_a)$ is the internal event model of type 2, $S_2(\mathcal{A}, A_a) = (E', R', Pre', A_a)$, defined as follows.

- $E' = \bigcup_k E^k$;
- $R'_j = \bigcup_k R^k_j$ for $j \neq Y$, and $R'_Y = \{(a_a, a'_a) \mid a_a, a'_a \in A_a\} \cup \bigcup_k R^k_Y$;
- for all $a' \in E'$, $Pre'(a') = Pre^k(a')$ if $a' \in A^k$.

□

Just as in the static case, one could say that an internal event model of type 1 and its associated internal event model of type 2 are in a sense equivalent (although we did not define a notion of validity for event models).

Example 3.3.5 ('Coin' example)
In Figure 3.6 how the private announcement to Bob is perceived by Ann and Bob is depicted. As in the BMS system, the formulas in the possible events are their preconditions and the boxed events are the events of the actual equivalence class. Because Bob perceived correctly the private announcement, his internal event model does correspond to a private announcement. On the other hand, for Ann nothing happened because she did not notice this private announcement to Bob. So her internal event model consists of a single possible event with a tautology as precondition. □

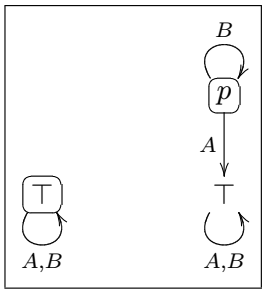


Figure 3.6: Internal event model for Ann (*left*) and Bob (*right*) corresponding to the private announcement to Bob that the coin is heads up

3.3.2 The update product

Definitions

After (or during) the event takes place, the agent Y updates her representation of the world according to her perception of this event. Formally, this amounts to define a product update between an internal model and an internal event model. That is what we are going to do now. But because we have two types of models, we define two types of update products. We will show afterwards that these two types of update are equivalent.

Definition 3.3.6 (Update product of type 1)
Let $(\mathcal{M}, W_a) = \{(M^1, w^1); \ldots; (M^n, w^n)\}$ be an internal model of type 1 (with $M^i = (W^i, R^i, V^i)$). Let $(\mathcal{A}, A_a) = \{(A^1, a^1); \ldots; (A^m, a^m)\}$ be an internal event model of type 1 (with $A^k = (E^k, R^k, Pre^k)$).

If $\mathcal{M}, w^i \not\models Pre^k(a^k)$ then $M^i \otimes_1 A^k$ is not defined. Otherwise, $M^i \otimes_1 A^k$ is defined and it is the submodel of $M = (W, R, V)$ generated by (w^i, a^k), where

- $W = \{(w, a) \mid w \in W^i, a \in A^k, \mathcal{M}, w \models Pre^k(a)\}$;
- $(w', a') \in R_j(w, a)$ iff $w' \in R^i_j(w)$ and $a' \in R^k_j(a)$;
- $(w, a) \in V(p)$ iff $w \in V^i(p)$ for all $p \in \Phi$.

Then the *updated model of type 1* $(\mathcal{M} \otimes_1 \mathcal{A})$ is defined as follows. If for all $(M^i, w^i) \in \mathcal{M}$ and $(A^k, a^k) \in \mathcal{A}$, $M^i \otimes_1 A^k$ is not defined then $\mathcal{M} \otimes_1 \mathcal{A}$ is not defined. Otherwise,

$$\mathcal{M} \otimes_1 \mathcal{A} = \{(M^i \otimes_1 A^k, (w^i, a^k)) \mid (M^i, w^i) \in \mathcal{M}, (A^k, a^k) \in \mathcal{A} \text{ and } M^i \otimes_1 A^k \text{ is defined}\}$$

The updated model of type 1 is written $(\mathcal{M} \otimes_1 \mathcal{A}, W_a \otimes_1 A_a)$. □

Note that in the definition of W, the preconditions $Pre(a^k)$ are evaluated in $\mathcal{M}$ and not in M^i. Indeed, in case $Pre^k(a^k)$ is a formula of the form $B_Y\varphi$, then we need the whole internal

model to evaluate it (see the truth conditions for $B_Y\varphi$ in Definition 2.3.10). That is why we need to redefine the product update completely. This is not the case for the internal model of type 2 where the definition is much more compact.

Definition 3.3.7 (Update product of type 2)
Let $(\mathcal{M}, W_a) = (W, R, V, W_a)$ be an internal model of type 2 and $(\mathcal{A}, A_a) = (E, R, Pre, A_a)$ be an internal event model of type 2. If $W_a \otimes A_a = \{(w, a) \in W_a \times A_a \mid \mathcal{M}, w \models Pre(a)\}$ is empty then the *updated model of type 2* is not defined. Otherwise, the updated model of type 2 is defined and it is the submodel of $\mathcal{M} \otimes \mathcal{A}$ (see Definition 3.2.3) generated by $W_a \otimes A_a$. It is abusively written $(\mathcal{M} \otimes \mathcal{A}, W_a \otimes A_a)$. □

In Proposition 2.3.12, we saw that the notions of internal (event) models of type 1 and type 2 are equivalent. The following proposition shows that the notions of update product of type 1 and type 2 are also equivalent.

Proposition 3.3.8 (Update products of type 1 and 2 are equivalent)
Let $(\mathcal{M}, W_a)$ be an internal model of type 1 and let $(\mathcal{A}, A_a)$ be an internal event model of type 1 such that $\mathcal{M} \otimes_1 \mathcal{A}$ is defined. Let S_2 be the mappings defined in Definition 2.3.7 and Definition 3.3.4. Then for all $(w_a, a_a) \in W_a \otimes_1 A_a$,

$$S_2(\mathcal{M} \otimes_1 \mathcal{A}, W_a \otimes_1 A_a), (w_a, a_a) \leftrightarrows S_2(\mathcal{M}, W_a) \otimes S_2(\mathcal{A}, A_a), (w_a, a_a).^{5}$$

PROOF. Let $S_2(\mathcal{M} \otimes_1 \mathcal{A}, W_a \otimes_1 A_a) = (W^1, R^1, V^1, W_a \otimes_1 A_a)$, $S_2(\mathcal{M}, W_a) \otimes S_2(\mathcal{A}, A_a) = (W^2, R^2, V^2, W_a \otimes A_a)$ and $\mathcal{A} = \{(A^1, a^1), \ldots, (A^m, a^m)\}$.
First we show that $W^1 = W^2$.
$(w, a) \in W_1$
iff there is some k such that $w \in \mathcal{M}$, $a \in A^k$ and $\mathcal{M}, w \models Pre^k(a)$,
iff there is some k such that $w \in S_2(\mathcal{M}, W_a)$, $a \in A^k$ and $S_2(\mathcal{M}, W_a), w \models Pre^k(a)$,
iff $w \in S_2(\mathcal{M}, W_a)$, $a \in S_2(\mathcal{A}, A_a)$ and $S_2(\mathcal{M}, W_a), w \models Pre(a)$,
iff $(w, a) \in W^2$.

For all $(w, a) \in S_2(\mathcal{M} \otimes_1 \mathcal{A}, W_a \otimes_1 A_a)$ and $(w', a') \in S_2(\mathcal{M}, W_a) \otimes S_2(\mathcal{A}, A_a)$, we set

$$(w, a) Z (w', a') \text{ iff } w = w' \text{ and } a = a'.$$

One can then easily show that Z is a bisimulation between $S_2(\mathcal{M} \otimes_1 \mathcal{A}, W_a \otimes A_a)$ and $S_2(\mathcal{M}, W_a) \otimes S_2(\mathcal{A}, A_a)$. QED

Example 3.3.9 ('Coin' example)
In Figure 3.7 and Figure 3.8 Ann's and Bob's scenarios of the 'coin' example are depicted. The first model corresponds to how they perceived the initial situation, the second model corresponds to how they perceived the private announcement and the last model corresponds to how they perceive the resulting situation after this private announcement. We use internal (event) model of type 2. □

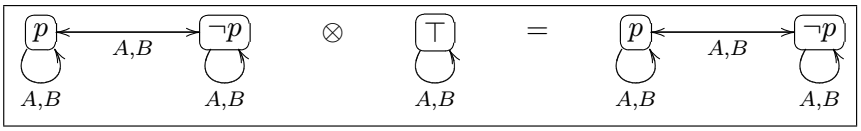


Figure 3.7: Ann's scenario

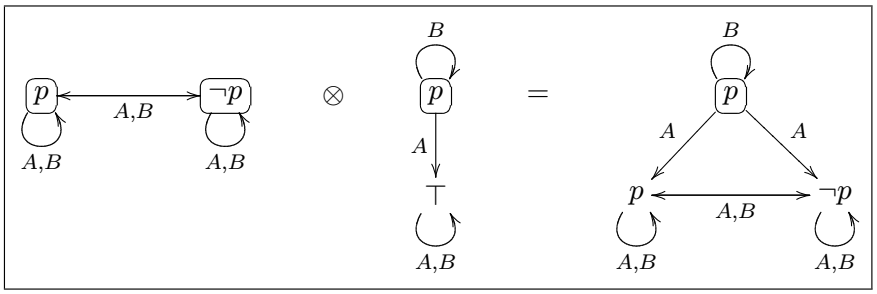


Figure 3.8: Bob's scenario

As you can see, the mechanisms used in the update products of both definitions are essentially the same as in the external approach. This is how it should be. Indeed, the external update product is supposed to render how the agents actually update their beliefs. So the mechanisms used in the update product of the external approach should be the same as the ones used in the update product of the internal approach. This connection will be made more precise in the next section.

Because these update mechanisms are essentially the same, it also entails that, as in the external approach, the updated models of type 2 or of type 1 are not necessarily serial. So the updated model is not necessarily an internal model. We are now going to study under which conditions it is an internal model.

Under which conditions updated models are internal models?

We start with the update product of type 2.

Proposition 3.3.10 *Let $(\mathcal{M}, W_a)$ be an internal model of type 2 and let $(\mathcal{A}, A_a)$ be an internal event model of type 2. Let $n = |\mathcal{M}| \cdot |\mathcal{A}|$ and assume that $(\mathcal{M} \otimes \mathcal{A}, W_a \otimes A_a)$ is defined. Then,*

$$(\mathcal{M} \otimes \mathcal{A}, W_a \otimes A_a) \text{ is an internal model iff } \mathcal{M}, w \models \delta^n(a) \text{ for some } w \in W_a \text{ and } a \in A_a.$$

PROOF. The proof follows directly from Proposition 3.2.15 because the update product $\otimes$ used in $\mathcal{M} \otimes \mathcal{A}$ is exactly the same as the BMS product. QED

[5]Let $(\mathcal{M}, W_a)$ and $(\mathcal{M}', W_a')$ be internal models of type 2 and $w \in \mathcal{M}$, $w' \in \mathcal{M}'$. We write $(\mathcal{M}, W_a), w \leftrightarrows (\mathcal{M}', W_a'), w'$ for $\mathcal{M}, w \leftrightarrows \mathcal{M}', w'$.

As we are now going to see, the seriality condition for updated models of type 1 is a bit more involved. This is because we do not use exactly the BMS product but employ a slightly different one, which is nevertheless based on the BMS one. This was not the case for the definition of updated models of type 2 where the BMS update product is used without any modification.

Proposition 3.3.11 *Let $(\mathcal{M}, W_a) = \{(M^1, w^1); \dots; (M^n, w^n)\}$ be an internal model of type 1 (with $M^i = (W^i, R^i, V^i)$). Let $(\mathcal{A}, A_a) = \{(A^1, a^1); \dots; (A^m, a^m)\}$ be an internal event model of type 1 (with $A^k = (E^k, R^k, Pre^k)$). Assume that $(\mathcal{M} \otimes_1 \mathcal{A}, W_a \otimes_1 A_a)$ is defined. Then,*

$$(\mathcal{M} \otimes_1 \mathcal{A}, W_a \otimes_1 A_a) \text{ is an internal model iff for all } w^i, \mathcal{M}, w^i \models \bigwedge_{(A^k, a^k) \in \mathcal{A}} (Pre^k(a^k) \to \delta^d(a^k))$$

where $d = max\{|M^i| \cdot |A^k| \mid i \in \{1, \dots, n\}, k \in \{1, \dots, m\}\}$ and δ is defined in Definition 3.2.12.

PROOF. First we prove a lemma.

> **Lemma 3.3.12** *Let $d_{i,k} = |M^i| \cdot |A^k|$. Then $\mathcal{M}, w^i \models \delta^{d_{i,k}}(a^k)$ iff $M^i \otimes_1 A^k$ is defined and serial.*
>
> PROOF. Because the update product $M^i \otimes_1 A^k$ is not the same as the BMS update product, we cannot apply directly Proposition 3.2.15.
>
> – Assume $\mathcal{M}, w^i \models \delta^{d_{i,k}}(a^k)$. Then $\mathcal{M}, w^i \models Pre^k(a^k)$, so $M^i \otimes_1 A^k$ is defined.
> Besides, note that $M^i \otimes_1 A^k$ is serial iff (O)
>
> 1. for all $j \in G$, $R_j(w^i, a^k) \neq \emptyset$
> 2. for all $j \neq Y$, for all $v^j \in R_j(w^i)$ and $b^j \in R_j(a^k)$ such that $\mathcal{M}, v^j \models Pre^k(b^j)$, the submodel of $M^i(v^j) \otimes A^k(b^j)$ generated by (v^j, b^j) is serial, where
> $M^i(v^j)$ is the submodel of M^i generated by v^j,
> $A^k(b^j)$ is the submodel of A^k generated by b^j.
>
> $M^i(v^j) \otimes A^k(b^j)$ is the usual BMS update product. Indeed, for all $\varphi \in \mathcal{L}$, all $u^j \in M^i(v^j)$, $M^i(v^j), u^j \models \varphi$ iff $\mathcal{M}, u^j \models \varphi$. So for the worlds of $M^i(v^j)$, the update product $\otimes_1$ is the same as the BMS update product $\otimes$. This will allow us to apply Proposition 3.2.15.
>
> 1. $\mathcal{M}, w^i \models \delta^1(a^k)$ because $d_{i,k} \geq 1$ and Proposition 3.2.17. So $\mathcal{M}, w^i \models \bigwedge_{j \in G} \hat{B}_j \bigvee_{b \in R_j(a^k)} Pre^k(b)$. So $R_j(w^i, a^k) \neq \emptyset$ for all $j \in G$.
> 2. Let $v^j \in R_j(w^i)$ and $b^j \in R_j(a^k)$ such that $\mathcal{M}, v^j \models Pre(b^j)$, i.e. $M^i, v^j \models Pre^k(b^j)$.
> Then $M^i, v^j \models \delta^{d_{i,k}-1}(b^j)$ by definition of $\delta^{d_{i,k}}(a^k)$. Besides, $|M^i(v^j)| \leq |M^i| - 1$ and $|A^k(b^j)| \leq |A^k| - 1$. So $n = |M^i(v^j)| \cdot |A^k(b^j)| \leq d_{i,k} - 1$.
> So $M^i, v^j \models \delta^n(b^j)$ by Proposition 3.2.17.
> So the submodel of $M^i(v^j) \otimes A^k(b^j)$ generated by (v^j, b^j) is defined and serial by Proposition 3.2.15.

– We prove by induction on $m \leq d_{i,k}$ that if $\mathcal{M}, w^i \models \neg\delta^m(a^k)$ then $M^i \otimes_1 A^k$ is either not defined or not serial.

m=0 $\delta^0(a^k) = Pre^k(a^k)$. So $\mathcal{M}, w^i \models \neg Pre^k(a^k)$. So $M^i \otimes_1 A^k$ is not defined.

m+1 $\mathcal{M}, w_i \models \neg\delta^m(a^k)$ iff $\mathcal{M}, w_i \models \neg Pre^k(a^k)$ or $\mathcal{M}, w^i \models \bigvee_{j\in G} B_j \bigwedge_{b\in R_j(a^k)} \neg\delta^{m-1}(b)$ or $\mathcal{M}, w^i \models \bigvee_{j\in G} \hat{B}_j \bigvee_{b\in R_j(a^k)} \big(Pre(b) \wedge \neg\delta^{m-1}(b)\big)$.

1. If $\mathcal{M}, w^i \models \neg Pre^k(a^k)$ then $M^i \otimes_1 A^k$ is not defined.
2. If $\mathcal{M}, w^i \models \bigvee_{j\in G} B_j \bigwedge_{b\in R_j(a^k)} \neg\delta^{m-1}(b)$ then there is $j \in G$ such that $\mathcal{M}, w^i \models B_j \bigwedge_{b\in R_j(a^k)} \neg\delta^{m-1}(b)$
   (a) If $j = Y$ then $\mathcal{M}, w^i \models \neg\delta^{m-1}(a^k)$. So $M^i \otimes A^k$ is not defined or not serial by Induction Hypothesis.
   (b) If $j \neq Y$ then for all $v^j \in R_j(w^i)$, all $b^j \in R_j(a_k)$, $M^i, v^j \models \neg\delta^{m-1}(b^j)$. So $M^i(v^j) \otimes A^k(b^j)$ is not defined or not serial by Proposition 3.2.15. So, because M^i and A^k are serial, $M^i \otimes_1 A^k$ is not serial by observation (O).
3. If $\mathcal{M}, w^i \models \bigvee_{j\in G} \hat{B}_j \bigvee_{b\in R_j(a^k)} (Pre(b) \wedge \neg\delta^{m-1}(b)$ then there is $j \in G$ and $v^j \in R_j(w^i)$ and $b^j \in R_j(a^k)$ such that $\mathcal{M}, v^j \models Pre^k(b^j) \wedge \neg\delta^{m-1}(b^j)$.
   (a) If $j = Y$ then $\mathcal{M}, w^i \models \neg\delta^{m-1}(a^k)$. So the result holds by Induction Hypothesis.
   (b) If $j \neq Y$ then by the same reasoning as in 2)b), we get that $M^i \otimes_1 A^k$ is not defined or not serial.

QED

- Assume for all w^i that $\mathcal{M}, w^i \models \bigwedge_{(A^k,a^k)\in\mathcal{A}} (Pre^k(a^k) \to \delta^d(a^k))$ (*). Let $(M^i, w^i) \in \mathcal{M}$ such that there is $(A^k, a^k) \in \mathcal{A}$ such that $\mathcal{M}, w^i \models Pre^k(a^k)$. Then $\mathcal{M}, w^i \models \delta^d(a^k)$ by (*).

  But $\models \delta^d(a^k) \to \delta^{d_{i,k}}(a^k)$ because $d_{i,k} \leq d$ and because of Proposition 3.2.17.

  So $\mathcal{M}, w^i \models \delta^{d_{i,k}}(a^k)$. Then $M^i \otimes_1 A^k$ is serial by Lemma 3.3.12, and so for all $(M^i, w^i) \in \mathcal{M}$ and $(A^k, a^k) \in \mathcal{A}$ such that $M^i \otimes_1 A^k$ is defined.

  So finally, $\mathcal{M} \otimes_1 \mathcal{A}$ is an internal model.

- Assume that there is w^i and $(A^k, a^k) \in \mathcal{A}$ such that $\mathcal{M}, w^i \models Pre^k(a^k) \wedge \neg\delta^d(a^k)$.

  $d \geq d_{i,k} = |M^i| \cdot |A^k|$ by assumption. So $\mathcal{M}, w^i \models \neg\delta^{d_{i,k}}(a^k)$ by Proposition 3.2.16. So $M^i \otimes_1 A^k$ is defined but not serial by Lemma 3.3.12. So $\mathcal{M} \otimes_1 \mathcal{A}$ is not an internal model.

QED

3.4 Some connections between the external and the internal approach

Just as in Chapter 2, we can perfectly set some connections between the external and the internal approach, and so even in a dynamic setting. We saw in Chapter 2 that from an external model, we could extract the internal model of each agent. We are going to see that we can do the same for the event models: from an external event model, we can extract the internal event model of each agent. Besides, in reality each agent updates her internal model with her respective internal event model, leading to a new situation. It seems natural to wonder whether this new situation corresponds formally to the one obtained by updating classically the two external models. We are going to see that this is indeed the case.

3.4.1 From (external) event model to internal event model

First we define the notion of external event model. An external event model is a pointed event model which is serial, transitive and euclidean. This notion is of course very similar to the notion of external model in the static case. Besides, still similarly, we can easily get from an external event model the event model associated to any agent just as we get from the external model the model associated to any agent.

Definition 3.4.1 (Event model associated to an agent in an external event model)
Let (A, a_a) be an external event model and let $j \in G$. The *event model associated to the agent j in (A, a_a)* is the submodel of A generated by $R_j(a_a)$, and $R_j(a_a)$ is its actual equivalence class. □

Proposition 3.4.2 *Let (A, a_a) be an external event model. The event model associated to agent j in (A, a_a) is an internal event model (of type 2).*

The proof is identical to the static case.

Example 3.4.3 ('Coin' example)
In Figure 3.9 is depicted the external event model corresponding to the private announcement to Bob that the coin is heads up and the event models associated to Ann and Bob. Note that these associated event models are the same as in Figure 3.6. □

Finally, as in the static case, we could perfectly define the external event model of a particular event if we suppose given the internal event models of each agent and the actual event. The definition is completely similar to Definition 2.3.19 so we do not spell it out here.

3.4.2 Preservation of the update product

The BMS system is made up of three main notions: the static models, the event models and the update product. So far, we have set some connections between the external and internal approach for the first two notions. It remains to set some connections between the external and the internal approach for the update product. This is what we are going to do now.

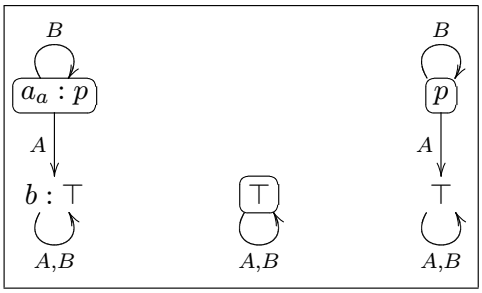


Figure 3.9: External event model (A, a_a) (*left*); Event model associated to Ann (*center*); Event model associated to Bob (*right*)

Theorem 3.4.4 *Let (M, w_a) be an external model and let (A, a_a) be an external event model. Let V_a be the restriction of the valuation of M to w_a. We assume that $M, w_a \models Pre(a_a)$.*

Let $\{(\mathcal{M}^j, W^j) \mid j \in G\}$ be the models associated to (M, w_a) and the agents $j \in G$. Let $\{(\mathcal{A}^j, A^j) \mid j \in G\}$ be the event models associated to (A, a_a) and the agents $j \in G$. Then,

$$M \otimes A, (w_a, a_a) \leftrightarrows Ext\left(\{(\mathcal{M}^j \otimes \mathcal{A}^j, W^j \otimes A^j) \mid j \in G\}, (w_a, V_a)\right), w_a.$$

where Ext is the mapping defined in Definition 2.3.19.

This theorem tells us that the way the agents update their beliefs from an internal point of view coincides with the way they update their beliefs from an external point of view. This is what we should expect.

PROOF. As in [Baltag *et al.*, 1999], one can define a notion of bisimulation between event models by replacing condition 1 of Definition 2.2.11 by 'if aZa' then $Pre(a)$ and $Pre(a')$ are logically equivalent'. [6] Then if $M, w \leftrightarrows M', w'$ and $A, a \leftrightarrows A', a'$ and $M, w \models Pre(a)$ and $M', w' \models Pre(a')$ then $M \otimes A, (w, a) \leftrightarrows M' \otimes A', (w', a')$.

Then for all $j \in G$, if there is $w^j \in R_j(w_a)$ and $a^j \in R_j(a_a)$ such that $M, w^j \models Pre(a^j)$ then $\mathcal{M}^j, w^j \leftrightarrows M, w^j$ and $\mathcal{A}^j, a^j \leftrightarrows A, a^j$ by Proposition 2.2.14. Hence $\mathcal{M}^j \otimes \mathcal{A}, (w^j, a^j) \leftrightarrows M \otimes A, (w^j, a^j)$.

So for all $(w^j, a^j) \in R_j(w_a, a_a)$ there is $(w^j, a^j)' \in R_j(w_a)$ such that
$Ext\left[\{(\mathcal{M}^j \otimes \mathcal{A}^j, W^j \otimes A^j) \mid j \in G\}, (w_a, V_a)\right], (w^j, a^j) \leftrightarrows M \otimes A, (w_a, a_a)$.
Vice versa, for all $(w^j, a^j) \in R_j(w_a)$ there is $(w^j, a^j)' \in R_j(w_a, a_a)$ such that
$Ext\left[\{(\mathcal{M}^j \otimes \mathcal{A}^j, W^j \otimes A^j) \mid j \in G\}, (w_a, V_a)\right], (w^j, a^j) \leftrightarrows M \otimes A, (w^j, a^j)$.
Finally, in $Ext\left[\{(\mathcal{M}^j \otimes \mathcal{A}^j, W^j \otimes A^j) \mid j \in G\}, (w_a, V_a)\right]$, the actual world w_a satisfies the same propositional letters as (w_a, a_a) in $M \otimes A$.
So $M \otimes A, (w_a, a_a) \leftrightarrows Ext\left[\{(\mathcal{M}^j \otimes \mathcal{A}^j, W^j \otimes A^j) \mid j \in G\}, (w_a, V_a)\right], w_a$. QED

[6] Instead of bisimulation between event models we could also use instead the weaker notion of event *emulation* between event models introduced by Ruan in [Ruan, 2004]. Then all that follows would still hold.

3.5 Concluding remarks

As we said in Section 2.3.1, when we want to deal with epistemic situations, specifying which modeling approach one considers and sticking to this approach is quite important. As we also said, mixing the different approaches leads to technical or philosophical problems. For example, in [Herzig *et al.*, 2004], to model an epistemic situation the authors start by considering pointed epistemic models whose pointed world is supposed to be the actual world. So they apparently follow the external approach. Then they consider what they call "purely ontic events" and try to model them. These events are events that do not "bring any feedback to any of the agents" such as "sending an email to somebody without knowing whether it will be received by the addressee". These events are supposed to transform a given world w in a non-empty set of worlds, each being a possible outcome of the execution of this event in world w. In case this set contains two or more worlds, this event is said to be nondeterministic. So, if such an event is performed in the actual world, this yields a set of "potential actual worlds". They represent such a situation by a multi-pointed epistemic model and throughout the paper they deal with both pointed epistemic model and multi-pointed epistemic models. This introduction of multi-pointed models seems a bit ad hoc but the real problem with this approach is that it is difficult to give intuitive meaning to this set of "potential actual worlds", at least in the external approach. It seems to us that this problem stems from a confusion in the different modeling approaches. Indeed, these "purely ontic events" are not truly nondeterministic. "Sending an email to somebody *without knowing* whether it will be received by the addressee" is not per se a nondeterministic event but rather reveals a lack of knowledge about the event from the modeler. In fact, modeling such an event, defined as it is, makes sense only in the imperfectly external approach. Multi-pointed epistemic models are also a way to model epistemic situations in the imperfectly external approach: the pointed worlds are the worlds that the imperfect and external modeler considers as being possibly the actual world. In fact, if we stick from the beginning to this imperfectly external approach, then this set of pointed worlds in the resulting multi-pointed epistemic model can be given a natural meaning: these are the worlds that the modeler considers as being possibly the actual world given his knowledge of the situation and what has happened.

In this chapter we have proposed an internal version of the BMS system, very much in line with what we did for epistemic logic in the preceding chapter. Besides, we have set some connections between the internal and the external approach for all the notions introduced in both approaches: (internal/external) model, (internal/external) event model and product update. So, now we do have a dynamic epistemic logic for the internal approach.

However, we do not have a way to ensure that the updated model is serial (i.e. is an internal model). Intuitively, an updated model which is not serial means that for agent Y there is an agent i (possibly Y herself) whose beliefs are inconsistent. More precisely, it means that for agent Y, it is not common belief that the agents' beliefs are consistent. Of course we would like to avoid it and in that case agent Y needs to revise her beliefs. In the next chapter we are going to propose ways to cope with this issue for the case of private announcement by generalizing AGM belief revision theory to a multi-agent setting.

Chapter 4

Internal approach: the case of private announcements

4.1 Introduction

As we said, dynamic epistemic logic is about the logical study of belief change. But actually, there is another formal approach to belief change that is also based on logic, namely AGM belief revision theory. Unlike dynamic epistemic logic, AGM belief revision theory is designed for a single agent and is not a genuine logical system. It typically deals with changes that the agent's representation of the surrounding world must undergo after receiving *conflicting* information. This differs as well from the systems of dynamic epistemic logic presented so far because in these systems the incoming information is assumed to be consistent with the agents' beliefs.

Belief revision theory was developed before the beginning of dynamic epistemic logic. It really started with Alchourrón, Gärdenfors and Makinson's seminal paper [Alchourrón *et al.*, 1985]. The original motivations for these authors were a bit different from how the theory is used nowadays. Alchourrón's motivation was to model the revision of norms in legal systems whereas Gärdenfors' motivation was to model the revision of scientific theories. Soon after, Grove [Grove, 1988] provided a semantic account of revision based on the possible world semantics. His system was inspired by the sphere semantics that Lewis gave for counterfactuals [Lewis, 1973]. It was then followed by a stream of publications on fine-tuning the notion of epistemic entrenchment [Meyer *et al.*, 2000], on revising belief bases [Benferhat *et al.*, 2002], on the difference between belief revision and belief update [Katsuno and Mendelzon, 1991] and on the problem of iterated belief revision [Darwiche and Pearl, 1997].

The semantics of AGM belief revision theory was the starting point to define our semantics for the internal approach in a multi-agent setting. In fact, our semantics is a generalization of the AGM semantics to the multi-agent case. So we would expect that results about AGM theory can be generalized to the multi-agent case too. In Section 4.2 we are going to

see that it is indeed the case. Then in Section 4.3, we will propose new rationality postulates which are specific to our multi-agent setting. Finally, in Section 4.4 we will give a concrete example of revision operation together with a concrete example.

4.2 Generalizing AGM to the multi-agent case

In AGM belief revision theory, the epistemic state of the agent is often represented by a belief set. A *belief set* K is a set of propositional formulas that is closed under logical consequence. These propositional formulas represent the beliefs of the agent. AGM distinguishes three types of belief change: expansion, revision and contraction. The expansion of K with a propositional formula φ, written $K + \varphi$, consists of adding φ to K and taking all the logical consequences. Note that this might yield inconsistency. The revision of K with φ, written $K * \varphi$, consists of adding φ to K, but in order that the resulting set be consistent, some formulas are removed from K. Finally, the contraction of K with φ, written here $K \dot{-} \varphi$, consists in removing φ from K, but in order that the resulting set be consistent, some other formulas are also removed. Of course there are some connections between these operations. From a contraction operation, one can define a revision operation thanks to the Levi identity:

$$K * \varphi = (K \dot{-} \neg\varphi) + \varphi.$$

And from a revision operation, one can define a contraction operation thanks to the Harper identity:

$$K \dot{-} \varphi = K \cap (K * \neg\varphi).$$

In this chapter, we will focus on the revision and the expansion operation. We will show how these operations can be generalized to a multi-agent setting.

4.2.1 Expansion

State of the art

In this chapter, we assume that the set of propositional letters Φ is finite, and in this paragraph, all the formulas belong to the propositional language $\mathcal{L}_0$ defined over Φ.

Let $Cn(.)$ be the classical consequence operation, i.e. for a set of propositional formulas Σ, $Cn(\Sigma) = \{\chi \mid \Sigma \vdash \chi\}$. We can now define formally a belief set.

Definition 4.2.1 (Belief set)
A *belief set* K is a set of propositional formulas in $\mathcal{L}_0$ such that $Cn(K) = K$. We denote by $K_\perp$ the unique inconsistent belief set consisting of all propositional formulas. □

Classically, in AGM theory, we start by proposing rationality postulates that belief change operations must fulfill. These postulates make precise our intuitions about these operations and what we mean by rational change. Below are the rationality postulates for the expansion operation $+$ proposed by Gärdenfors [Gärdenfors, 1988].

K+1 $K + \varphi$ is a belief set

K+2 $\varphi \in K + \varphi$

K+3 $K \subseteq K + \varphi$

K+4 If $\varphi \in K$ then $K = K + \varphi$

K+5 $K + \varphi$ is the smallest set satisfying $K + 1$-$K + 4$.

$K + 1$ tells us that the expansion operation $+$ is a function from pairs of belief set and formula to belief sets. This entails that we can iterate the expansion operation. $K + 2$ tells us that when the agent expands her belief set by φ then as a result φ is one of her beliefs. All the other postulates refer to some kind of minimal change. $K + 3$ tells us that when the agent expands by φ she does not throw away any of her former beliefs. $K + 4$ tells us that if the agent already believes φ then expanding by φ should not change her beliefs: the change made to add φ to the belief set is minimal.

The following (representation) theorem tells us that these postulates actually determine a *unique* expansion operation on belief sets.

Theorem 4.2.2 [Gärdenfors, 1988]
A function $+$ satisfies $K+1-K+5$ iff for each belief set K and formula φ, $K+\varphi = Cn(K \cup \{\varphi\})$.

So from now on, we define the expansion operation $+$ by $K + \varphi = Cn(K \cup \{\varphi\})$.

So far our approach to expansion was syntactically driven. Now we are going to give a semantical approach to expansion and set some links between these two approaches.

We use the possible world semantics. First we consider the set $\mathcal{W}$ consisting of all the (logically) possible worlds. A possible world w can be viewed as an interpretation, i.e. a function from Φ to $\{\top, \bot\}$ which specifies which propositional letters (such as 'it is raining') are true in this world w. For a propositional formula χ, we write $w \models \chi$ when χ is true at w in the usual sense[1]. Then a formula χ is true in a set W of possible worlds, written $W \models \chi$, if and only if for all $w \in W$, $w \models \chi$. Besides, because Φ is finite, $\mathcal{W}$ is also finite. We can then represent the agent's epistemic state by a subset W of $\mathcal{W}$ (which is consequently finite as well). Intuitively, W is the smallest set of possible worlds in which the agent believes that the actual world is located.

There is actually a very close correspondence between belief sets and sets of possible worlds.

Definition 4.2.3

Let W be a finite set of possible worlds. We define the belief set K_W associated to W by $K_W = \{\chi \mid W \models \chi\}$.

Let K be a belief set. We define the set of possible worlds W_K associated to K by $W_K = \{w \mid w \models \chi \text{ for all } \chi \in K\}$. Then,

[1] $w \models \chi$ is defined inductively by: $w \models p$ iff $w(p) = \top$; $w \models \neg\chi$ iff not $w \models \chi$; and $w \models \chi \wedge \chi'$ iff $w \models \chi$ and $w \models \chi'$.

$$W \models \chi \text{ iff } \chi \in K_W, \text{ and } \chi \in K \text{ iff } W_K \models \chi.$$

□

Now we can define the semantic counterpart of the expansion operation defined previously.

Definition 4.2.4 ((Semantic) expansion)
Let W be a finite set of possible worlds and φ of formula. The *expansion* of W by φ, written $W + \varphi$, is defined as follows.

$$W + \varphi = \{w \in W \mid w \models \varphi\}.$$

□

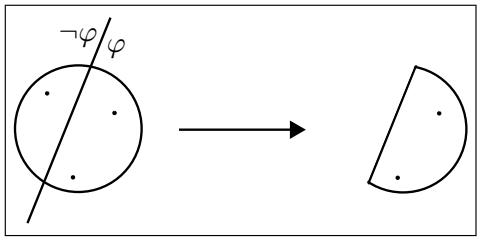


Figure 4.1: AGM expansion by φ

This semantic counterpart of the expansion is described graphically in Figure 4.1. The initial model W is on the left of the arrow and the expanded model $W + \varphi$ is on the right of the arrow. The dots represent possible worlds and the straight line separates the worlds satisfying φ from the worlds satisfying $\neg\varphi$.

Finally, we show that these two definitions of expansion, syntactic and semantic, are in fact equivalent.

Theorem 4.2.5 *For all belief sets K and all finite set of possible worlds W,*

$$\chi \in K + \varphi \textit{ iff } W_K + \varphi \models \chi, \textit{ and } W + \varphi \models \chi \textit{ iff } \chi \in K_W + \varphi.$$

PROOF. $\chi \in K + \varphi$
iff $\chi \in Cn(K \cup \{\varphi\})$
iff $K \cup \{\varphi\} \vdash \chi$
iff for all w such that $w \models \psi$ for all $\psi \in K \cup \{\varphi\}$, $w \models \chi$
iff for all $w \in W_K$, if $w \models \varphi$ then $w \models \chi$
iff for all $w \in W_K + \varphi$, $w \models \chi$
iff $W_K + \varphi \models \chi$.

$\chi \in K_W + \varphi$
iff $\chi \in Cn(K_W \cup \{\varphi\})$

iff $K_W \cup \{\varphi\} \vdash \chi$
iff for all w such that $w \models K_W \cup \{\varphi\}$, $w \models \chi$
iff for all w such that $w \models K_W$ and $w \models \varphi$, $w \models \chi$
iff for all $w \in W + \varphi$, $w \models \chi$
iff $W + \varphi \models \chi$. QED

This ends our account about expansion. Now we are going to propose a generalization of this operation to the multi-agent case.

Expansion and private announcement in a multi-agent setting

Assume we are now in a multi-agent setting and we follow the internal approach. Let us have a closer look at the case of private announcement. The event model of a private announcement of $\varphi \in \mathcal{L}$ to agent Y is depicted in Figure 4.2.

We will show that private announcements is the counterpart of AGM expansion.

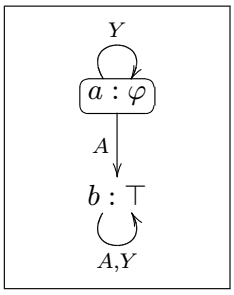


Figure 4.2: Private announcement of φ to agent Y

Proposition 4.2.6 *Let $(\mathcal{A}, \{a\}) = (\{a, b\}, R, Pre, \{a\})$ be the internal event model of a private announcement of $\varphi \in \mathcal{L}$ to Y. Then for all $n \in \mathbb{N}$, $\models_{Ext} \delta^n(b) \leftrightarrow \top$ and $\models_{Int} \delta^n(a) \leftrightarrow \varphi$.*

PROOF. We prove both results by induction on n. Clearly, the result holds for $n = 0$. Assume it is true for n.

Then $\delta^{n+1}(a) = \delta^0(a) \wedge \hat{B}_Y \delta^n(a) \wedge \hat{B}_A \delta^n(b) \wedge B_Y(Pre(a) \rightarrow \delta^n(a)) \wedge B_A(Pre(b) \rightarrow \delta^n(b))$ by Definition 3.2.12. So $\models_{Int} \delta^{n+1}(a) \leftrightarrow \varphi \wedge \hat{B}_Y \varphi \wedge \hat{B}_A \top \wedge B_Y(\varphi \rightarrow \varphi) \wedge B_A(\varphi \rightarrow \top)$ by induction hypothesis. Then $\models_{Int} \delta^{n+1}(a) \leftrightarrow \varphi \wedge \hat{B}_Y \varphi$. So $\models_{Int} \delta^{n+1}(a) \leftrightarrow \varphi$ by axiom T.

Besides, $\delta^{n+1}(b) = \delta^0(b) \wedge \hat{B}_Y \delta^n(b) \wedge \hat{B}_A \delta^n(b) \wedge B_Y(Pre(b) \rightarrow \delta^n(b)) \wedge B_A(Pre(b) \rightarrow \delta^n(b))$ by Definition 3.2.12. So $\models_{Ext} \delta^{n+1}(b) \leftrightarrow \top \wedge \hat{B}_Y \top \wedge \hat{B}_A \top \wedge B_Y(\top \rightarrow \top) \wedge B_A(\top \rightarrow \top)$, i.e. $\models_{Ext} \delta^{n+1}(b) \leftrightarrow \top$. QED

Corollary 4.2.7 *Let $(\mathcal{M}, W_a)$ be an internal model of type 1 and let $(\mathcal{A}, \{a\})$ be the internal event model of the private announcement of $\varphi \in \mathcal{L}$ to Y. Then $(\mathcal{M} \otimes_1 \mathcal{A}, W_a \otimes_1 \{a\})$ is defined and is an internal model of type 1 iff there is $w_a \in W_a$ such that $\mathcal{M}, w_a \models \varphi$.*

PROOF. Assume that $\mathcal{M} \otimes_1 \mathcal{A}$ is defined. We know that for all $n \in \mathbb{N}$, $\models_{Int} (Pre(a) \rightarrow \delta^n(a)) \leftrightarrow (Pre(a) \rightarrow Pre(a))$ by Proposition 4.2.6. So for all $w \in W_a$, $\mathcal{M}, w \models (Pre(a) \rightarrow \delta^n(a))$. This means that $\mathcal{M} \otimes_1 \mathcal{A}$ is an internal model by Proposition 3.3.11. So, if $\mathcal{M} \otimes_1 \mathcal{A}$ is defined then it is an internal model. But $\mathcal{M} \otimes_1 \mathcal{A}$ is defined iff there is $w_a \in W_a$ such that $\mathcal{M}, w_a \models \varphi$. So we get the result. QED

This corollary is obtained thanks to our study on seriality preservation for generated submodels made in Section 3.2.2. It tells us that as soon as the updated model of type 1 by a private announcement is defined then it must be an internal model, in other words it is serial. This might seem strange at first sight but the following crucial theorem provides a good explanation for that.

Theorem 4.2.8 *Let $(\mathcal{M}, W_a) = \{(M^1, w^1), \ldots, (M^n, w^n)\}$ be an internal model of type 1 and $(\mathcal{A}, \{a\})$ be the internal event model of a private announcement of $\varphi \in \mathcal{L}$ to Y. Then,*

$$\mathcal{M} \otimes_1 \mathcal{A} = \{(M^i, w^i) \mid \mathcal{M}, w^i \models \varphi\}$$

PROOF. It suffices to apply the definition of $\otimes_1$. QED

This theorem is very important. Indeed, it explains why an updated model of type 1 is an internal model of type 1 as soon as it is defined: it is because the updated model is a submodel of the original internal model of type 1. But more importantly, it bridges the gap between AGM belief revision theory and dynamic epistemic logic as viewed by BMS. Indeed, one can note that the definition of $\mathcal{M} \otimes_1 \mathcal{A}$ is very similar to the semantic definition of expansion in Definition 4.2.4. On the one hand, the expansion of a set of possible worlds by a propositional formula φ consists in the worlds that satisfy φ. On the other hand, the updated model of an internal model of type 1 by a private announcement of an epistemic formula φ consists in the *multi-agent* possible worlds that satisfy φ. This similarity is depicted in Figure 4.3 where the triangles represent multi-agent possible worlds.

So, informally, the BMS update by a private announcement can be viewed as a 'multi-agent' AGM expansion. In other words, AGM expansion can be viewed as a particular case of a BMS update by a private announcement in which Y is the only agent. This means that private announcement is the generalization of AGM expansion to the multi-agent case. This goes against van Ditmarsch, van der Hoek and Kooi's claim that *public* announcement can be viewed as a belief expansion [van Ditmarsch *et al.*, 2004]. However, their comparison is rather syntactical and in that respect they only consider special kinds of formulas to represent belief sets and (public) announcements, namely 'positive' formulas. Here, our results hold independently of any particular chosen language because we compare only the semantics of expansion and private announcement. We believe this semantic correspondence to be deeper than any syntactic one because the languages in the single agent case and the multi-agent case are anyway quite different and so do not allow for a straightforward comparison. Moreover, the fact that private announcement can be viewed as a generalization of expansion in a multi-agent setting is not accidental. Indeed, an important property of private announcement is that not only the actual world does not change but also the agents' beliefs do not change (except of course for agent Y's beliefs). For example, suppose $y$ou (Y) believe

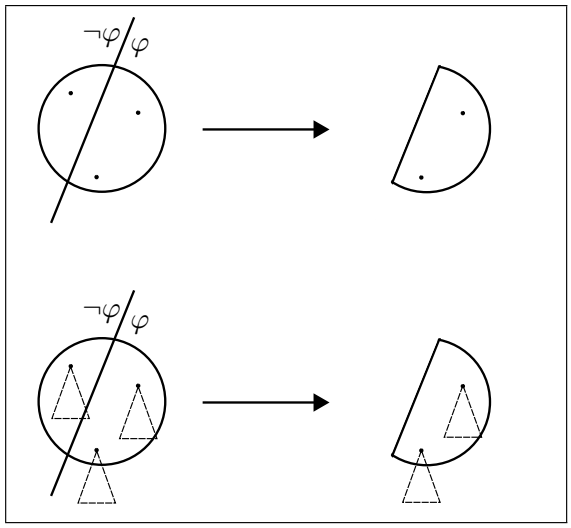


Figure 4.3: AGM expansion by φ (*above*) and BMS update by a private announcement of φ (*below*)

p, and agent j believes p (and perhaps even that p is common belief of Y and j). When a third external agent privately tells you that $\neg p$ then afterwards j still believes p and you still believe that j believes p (and that j believes that p is common belief). This static aspect of private announcements is similar to the static aspect of AGM belief revision in a single-agent case: in both cases the world does not change but only agent Y's beliefs about the world change.

Example 4.2.9 ('Coin' example)
Let us take up the 'coin example' and let us consider Bob's internal point of view. Bob's internal model of type 1 representing the initial situation is recalled in Figure 4.4. Then, according to Theorem 4.2.8, the resulting situation after the private announcement to Bob that the coin is heads up (p) is the internal model of type 1 composed only of the multi-agent possible world (M, w) on the left of Figure 4.4.

We can check that this result is correct. Indeed, we showed in Proposition 3.3.8 that the update products of type 1 and 2 are equivalent. So from the representation of type 2 of this scenario, we should get (M, w) as the resulting internal model of type 2 after the private announcement. This scenario is recalled in Figure 4.5. The first model represents Bob's internal model of type 2 of the original situation, which is equivalent to the internal model of type 1 depicted in Figure 4.4. The second model is Bob's internal event model of type 2 of the private announcement. The last model is Bob's updated model of type 2 after the private announcement. This last model is indeed the same model as (M, w) in Figure 4.4. □

So, now we know that the generalization of expansion to the multi-agent case is private announcement. We also know by Theorem 4.2.8 how to get easily the updated model of type 1 by a private announcement of φ. However, this updated model of type 1 is not necessarily

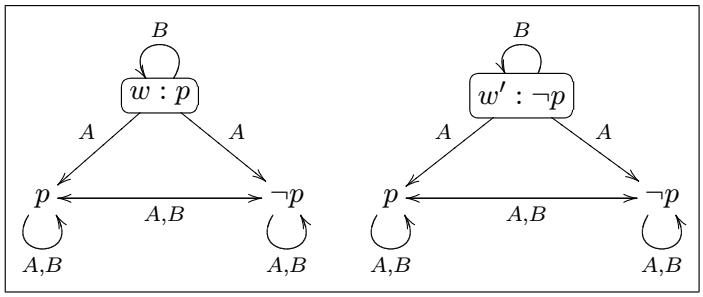


Figure 4.4: Bob's initial internal model of type 1 $\{(M, w), (M', w')\}$

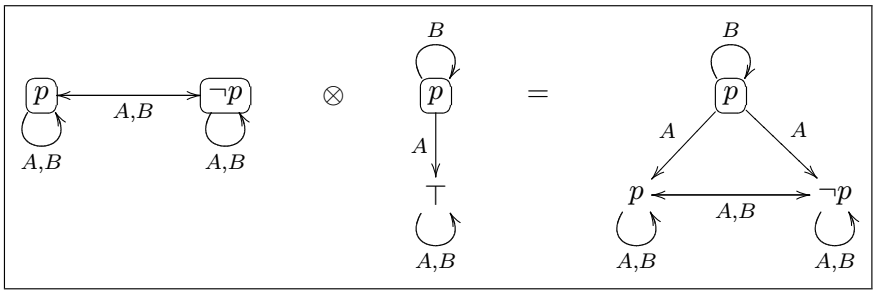


Figure 4.5: Bob's scenario

defined. For that by Corollary 4.2.7 there must be a multi-agent possible world of the initial internal model of type 1 that satisfies φ . But what happens when there is no multi-agent possible world that satisfies φ? In other words, what happens when the private announcement made to Y contradicts her beliefs? In that case, agent Y must *revise* her initial internal model. We call this kind of revision 'private multi-agent belief revision'.

This is what we will deal with in the next section. First we will recall AGM belief revision, focusing more particularly on the revision operation. Then we will generalize this framework to the multi-agent case and therefore study private multi-agent belief revision.

4.2.2 Revision

State of the art

In this paragraph, all formulas are propositional formulas. Just as for expansion, Gärdenfors and his colleagues proposed rationality postulates for revision operations. These postulates make precise what we mean by rational change, and more precisely rational revision. We will not provide intuitive motivations for these postulates (even if some of them have been criticized), see [Gärdenfors, 1988] for details. However, note that these postulates do not characterize a unique revision operation, unlike the postulates for expansion.

K*1 $K * \varphi$ is a belief set

K*2 $\varphi \in K * \varphi$

K*3 $K * \varphi \subseteq K + \varphi$

K*4 If $\neg\varphi \notin K$ then $K + \varphi \subseteq K * \varphi$

K*5 $K * \varphi = K_\perp$ iff φ is unsatisfiable

K*6 If $\varphi \leftrightarrow \varphi'$ then $K * \varphi = K * \varphi'$

K*7 $K * (\varphi \wedge \varphi') \subseteq (K * \varphi) + \varphi'$

K*8 If $\neg\varphi' \notin K * \varphi$ then $(K * \varphi) + \varphi' \subseteq K * (\varphi \wedge \varphi')$

Before going on, let us reconsider how we represent agent Y's epistemic state. So far we have proposed two equivalent formalisms: belief set and (finite) set of possible worlds. As we said, a belief set is an infinite set of formulas closed under logical consequence. However, this cannot be handled easily by computers because of its infinitude. We would like to have a more compact and finite representation of the agent's epistemic state. For that, we follow the approach of [Katsuno and Mendelzon, 1992].

As argued by Katsuno and Mendelzon, because Φ is finite, a belief set K can be equivalently represented by a mere propositional formula ψ. This formula is also called a belief base. Then $\chi \in K$ if and only if $\chi \in Cn(\psi)$. Besides, one can easily show that $\chi \in K + \varphi$ if and only if $\chi \in Cn(\psi \wedge \varphi)$. So in this approach, the expansion of the belief base ψ by φ is the belief base $\psi \wedge \varphi$, which is possibly an inconsistent formula. Now, given a belief base ψ and a formula φ, $\psi \circ \varphi$ denotes the revision of ψ by φ. But in this case, $\psi \circ \varphi$ is supposed to be consistent if φ is. Given a revision operation $*$ on belief sets, one can define a corresponding operation $\circ$ on belief bases as follows: $\psi \circ \varphi \rightarrow \chi$ if and only if $\chi \in Cn(\psi) * \varphi$. Thanks to this correspondence, Katsuno and Mendelzon set some rationality postulates for this revision operation $\circ$ on belief bases which are equivalent to the AGM rationality postulates for the revision operation $*$ on belief sets.

Lemma 4.2.10 [Katsuno and Mendelzon, 1992]
*Let * be a revision operation on belief sets and $\circ$ its corresponding operation on belief bases. Then * satisfies the 8 AGM postulates $K * 1 - K * 8$ iff $\circ$ satisfies the postulates $R1 - R6$ below:*

R1 $\vdash \psi \circ \varphi \rightarrow \varphi$.

R2 *if* $\psi \wedge \varphi$ *is satisfiable, then* $\vdash \psi \circ \varphi \leftrightarrow \psi \wedge \varphi$.

R3 *If* φ *is satisfiable, then* $\psi \circ \varphi$ *is also satisfiable.*

R4 *If* $\vdash \psi \leftrightarrow \psi'$ *and* $\vdash \varphi \leftrightarrow \varphi'$, *then* $\vdash \psi \circ \varphi \leftrightarrow \psi' \circ \varphi'$.

R5 $\vdash (\psi \circ \varphi) \wedge \varphi' \rightarrow \psi \circ (\varphi \wedge \varphi')$.

R6 *If* $(\psi \circ \varphi) \wedge \varphi'$ *is satisfiable, then* $\vdash \psi \circ (\varphi \wedge \varphi') \rightarrow (\psi \circ \varphi) \wedge \varphi'$.

So far our approach to revision was syntactically driven. Now we are going to give a semantical approach to revision and then set some links between the two approaches.

First some notations. $Mod(\psi)$ denotes the set of all logically possible worlds (also called models in that case) that make ψ true, i.e. $Mod(\psi) = \{w \in \mathcal{W} \mid w \models \psi\}$. If $\mathcal{M}$ is a set of possible worlds then $form(\mathcal{M})$ denotes a formula whose set of models is equal to $\mathcal{M}$.

Definition 4.2.11 (Faithful assignment)
A *pre-order* $\leq$ over $\mathcal{W}$ is a reflexive and transitive relation on $\mathcal{W}$. A pre-order is *total* if for every $w, w' \in \mathcal{W}$, either $w \leq w'$ or $w' \leq w$. Consider a function that assigns to each propositional formula ψ a pre-order $\leq_\psi$ over $\mathcal{W}$. We say this assignment is *faithful* if the following three conditions hold:

1. If $w, w' \in Mod(\psi)$, then $w <_\psi w'$ does not hold;
2. If $w \in Mod(\psi)$ and $w' \notin Mod(\psi)$, then $w <_\psi w'$ holds;
3. If $\vdash \psi \leftrightarrow \psi'$, then $\leq_\psi = \leq_{\psi'}$.

□

Intuitively, $w \leq_\psi w'$ means that the possible world w is closer to ψ than w'.

Definition 4.2.12 Let $\mathcal{M}$ be a subset of $\mathcal{W}$. A possible world w is *minimal* in $\mathcal{M}$ with respect to $\leq_\psi$ if $w \in \mathcal{M}$ and there is no $w' \in \mathcal{M}$ such that $w' <_\psi w$. Let

$$Min(\mathcal{M}, \leq_\psi) = \{w \mid w \text{ is minimal in } \mathcal{M} \text{ with respect to } \leq_\psi\}$$

□

The following representation theorem sets some connections between the semantic approach and the syntactic one.

Theorem 4.2.13 [Katsuno and Mendelzon, 1992]
Revision operation $\circ$ satisfies postulates $R1 - R6$ iff there exists a faithful assignment that maps each belief base ψ to a total pre-order $\leq_\psi$ such that

$$Mod(\psi \circ \varphi) = Min(Mod(\varphi), \leq_\psi).$$

This semantic revision process is described in Figure 4.6. In this figure, the dots represent possible worlds and the diagonal line separates the worlds satisfying φ from the worlds satisfying $\neg\varphi$. The worlds in the inner circle are the worlds that satisfy ψ and thus correspond to $Mod(\psi)$. The other circles represent the ordering $\leq_\psi$: if $w <_\psi w'$ then w is within a smaller circle than w' and if $w =_\psi w'$ then w and w' are in between the same successive circles. So the further a world is from the inner circle, the further it is from ψ. The worlds in the hatched part are then the worlds that satisfy φ and which are the closest to ψ. Therefore they represent $Mod(\psi \circ \varphi) = Min(Mod(\varphi), \leq_\psi)$.

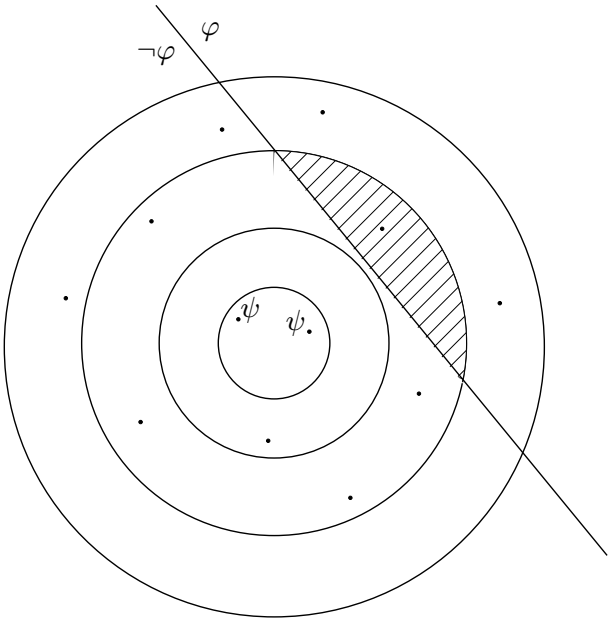


Figure 4.6: AGM belief revision

Grove proposed another semantic approach based on a system of spheres [Grove, 1988]. But one can show that his framework can be recast in the one just described.

This ends our account about revision. Now we are going to propose a generalization of revision to the multi-agent case.

Private multi-agent belief revision

In the multi-agent case like in the single-agent case, it does not make any sense to revise by formulas dealing with what the agent Y believes or considers possible. Indeed, due to the fact that positive and negative introspection are valid in KD45, Y already knows all she believes and all she disbelieves. So we restrict the epistemic language to a fragment that we call $\mathcal{L}^C_{\neq Y}$ defined as follows.

Definition 4.2.14 (Language $\mathcal{L}^C_{\neq Y}$)

$$\mathcal{L}^C_{\neq Y} : \varphi := \top \mid p \mid B_j\psi \mid \varphi \wedge \varphi \mid \neg\varphi,$$

where ψ ranges over $\mathcal{L}^C$ and j over $G - \{Y\}$. □

Note that by restricting ourselves to this kind of formulas, the two notions of validity of Definition 2.3.13 and Remark 2.3.14 coincide. That is to say, a formula $\varphi \in \mathcal{L}^C_{\neq Y}$ is positively valid if and only if it is negatively valid. It also entails that $\varphi \in \mathcal{L}^C_{\neq Y}$ is internally satisfiable

iff there is a multi-agent possible world (M, w) such that $M, w \models \varphi$. Besides, if we assume that agent Y is the only agent then $\mathcal{L}^C_{\neq Y}$ is just the propositional language, i.e. the language used in AGM belief revision theory.

We can now apply with some slight modifications the procedure spelled out for the single agent case in the previous paragraph.

First, we define $\mathcal{W}_G$ to be the set of all multi-agent possible worlds modulo bisimulation, and we pick the smallest multi-agent possible world among each class of bisimilarly indistinguishable multi-agent possible worlds. Then we define $Mod(\psi)$ by $Mod(\psi) = \{(M, w) \in \mathcal{W}_G \mid M, w \models \psi\}$. Thanks to Proposition 2.2.16 we can easily prove the following proposition.

Proposition 4.2.15 (Syntactic characterization of internal models)
Let $\mathcal{M}$ be an internal model. There is a formula $form(\mathcal{M}) \in \mathcal{L}^C_{\neq Y}$ such that $Mod(form(\mathcal{M})) = \mathcal{M}$.

PROOF. Let (M, w) be a multi-agent possible world. Then we set

$$\delta^*_M(w) = \left(\bigwedge_{\{p|w\in V(p)\}} p\right) \wedge \left(\bigwedge_{\{p|w\notin V(p)\}} \neg p\right) \wedge \bigwedge_{j\in G-\{Y\}} \left(\bigwedge_{v\in R_j(w)} \hat{B}_j \delta_M(v) \wedge B_j \left(\bigvee_{v\in R_j(w)} \delta_M(v)\right)\right).$$

Clearly $\delta^*_M(w) \in \mathcal{L}^C_{\neq Y}$, $M, w \models \delta^*_M(w)$ and for all multi-agent possible worlds (M', w'), if $M', w' \models \delta^*_M(w)$ then $M, w \leftrightarrows M', w'$ by applying Proposition 2.2.16. Let $\mathcal{M} = \{(M_1, w_1), \ldots, (M_n, w_n)\}$. We set $form(\mathcal{M}) = \delta^*_{M_1}(w_1) \vee \ldots \vee \delta^*_{M_n}(w_n)$. Then $form(\mathcal{M}) \in \mathcal{L}^C_{\neq Y}$ and $Mod(form(\mathcal{M})) = \mathcal{M}$. QED

The proof of this proposition is made possible because of the modularity of multi-agent possible worlds enforced by condition 2 in our definition of multi-agent possible world. Therefore, this is another motivation for this condition.

We then get the multi-agent generalization of Theorem 4.2.13 by replacing possible worlds w by multi-agent possible worlds (M, w) and replacing the propositional language $\mathcal{L}_0$ by $\mathcal{L}^C_{\neq Y}$, the rationality postulates being the same.

Theorem 4.2.16 *Revision operation $\circ$ on $\mathcal{L}^C_{\neq Y}$ satisfies postulates $R1 - R6$ in the internal logic* Int *iff there exists a faithful assignment that maps each belief base ψ to a total pre-order $\leq_\psi$ such that $Mod(\psi \circ \varphi) = Min(Mod(\varphi), \leq_\psi)$.*

PROOF. The proof is identical to the proof of Theorem 4.2.13 except that interpretations are replaced by multi-agent possible worlds and propositional formulas are replaced by formulas of $\mathcal{L}^C_{\neq Y}$.

('Only if') Assume that there is a revision operation satisfying conditions $R1 - R6$. We will define a total pre-order $\leq_\psi$ for each ψ by using the revision operation $\circ$. For any multi-agent possible worlds $(M, w), (M', w')$ ($(M, w) = (M', w')$ is permitted), we define a relation

$\leq_\psi$ as $(M,w) \leq_\psi (M',w')$ if and only if either $(M,w) \in Mod(\psi)$ or $(M,w) \in Mod(\psi \circ form((M,w),(M',w')))$.

We first show that $\leq_\psi$ is a total pre-order. It follows from $R1$ and $R3$ that $Mod(\psi \circ form((M,w),(M',w')))$ is a non-empty subset of $\{(M,w),(M',w')\}$. Hence, $\leq_\psi$ is total. In particular, if we consider the case where (M,w) is equal to (M',w'), $Mod(\psi \circ form((M,w))) = \{(M,w)\}$ holds. Hence, for any (M,w), $(M,w) \leq_\psi (M,w)$ (i.e. the reflexivity) holds.

We will show the transitivity. Assume that both $(M_1,w_1) \leq_\psi (M_2,w_2)$ and $(M_2,w_2) \leq_\psi (M_3,w_3)$ hold. We show $(M_1,w_1) \leq_\psi (M_3,w_3)$. There are three cases to consider.

1. $(M_1,w_1) \in Mod(\psi)$, $(M_1,w_1) \leq_\psi (M_3,w_3)$ follows from the definition of $\leq_\psi$

2. $(M_1,w_1) \notin Mod(\psi)$ and $(M_2,w_2) \in Mod(\psi)$. Since

$$Mod(\psi \wedge form((M_1,w_1),(M_2,w_2))) = \{(M_2,w_2)\}$$

holds, $Mod(\psi \circ form((M_1,w_1),(M_2,w_2)))$ follows from $R2$. Thus $(M_1,w_1) \not\leq (M_2,w_2)$ follows from $(M_1,w_1) \notin Mod(\psi)$. This contradicts $(M_1,w_1) \leq_\psi (M_2,w_2)$, so case 2 is impossible.

3. $(M_1,w_1) \notin Mod(\psi)$ and $(M_2,w_2) \notin Mod(\psi)$. By $R1$ and $R3$, $Mod(\psi \circ form((M_1,w_1),(M_2,w_2),(M_3,w_3)))$ is a nonempty subset of $\{(M_1,w_1),(M_2,w_2),(M_3,w_3)\}$.

$$(3.1) Mod(\psi \circ form((M_1,w_1),(M_2,w_2),(M_3,w_3))) \cap \{(M_1,w_1),(M_2,w_2)\} = \emptyset$$

$Mod(\psi \circ form((M_1,w_1),(M_2,w_2),(M_3,w_3))) = \{(M_3,w_3)\}$ holds in this case. If we regard φ and φ' as $form((M_1,w_1),(M_2,w_2),(M_3,w_3))$ and $form((M_2,w_2),(M_3,w_3))$ respectively in postulates $R5$ and $R6$, we obtain

$$Mod(\psi \circ form((M_1,w_1),(M_2,w_2),(M_3,w_3))) \cap \{(M_2,w_2),(M_3,w_3)\} =$$
$$Mod(\psi \circ form((M_2,w_2),(M_3,w_3))).$$

Hence, $Mod(\psi \circ form((M_2,w_2),(M_3,w_3))) = \{(M_3,w_3)\}$. This contradicts $(M_2,w_2) \leq_\psi (M_3,w_3)$ and $(M_2,w_2) \notin Mod(\psi)$. Thus (3.1) is not possible.

$$(3.2) Mod(\psi \circ form((M_1,w_1),(M_2,w_2),(M_3,w_3))) \cap \{(M_1,w_1),(M_2,w_2)\} \neq \emptyset.$$

Since $(M_1,w_1) \leq_\psi (M_2,w_2)$ and $(M_1,w_1) \notin Mod(\psi)$, $(M_1,w_1) \in Mod(\psi \circ form((M_1,w_1),(M_2,w_2)))$ holds. Hence, by regarding φ and φ' as $form((M_1,w_1),(M_2,w_2),(M_3,w_3))$ and

$form((M_1,w_1),(M_2,w_2))$ in postulates $R5$ and $R6$, we obtain

$$Mod(\psi \circ form((M_1,w_1),(M_2,w_2),(M_3,w_3))) \cap \{(M_1,w_1),(M_2,w_2)\} =$$
$$Mod(\psi \circ form((M_1,w_1),(M_2,w_2))).$$

Thus, $(M_1,w_1) \in Mod(\psi \circ form((M_1,w_1),(M_2,w_2),(M_3,w_3))) \cap \{(M_1,w_1),(M_2,w_2)\}$ holds. By using conditions $R5$ and $R6$ again in a similar way, we can obtain $(M_1,w_1) \in Mod(\psi \circ form((M_1,w_1),(M_3,w_3)))$. Therefore, $(M_1,w_1) \leq_\psi (M_3,w_3)$ holds.

Next we show that the assignment mapping ψ to $\leq_\psi$ is faithful. The first condition of the faithfulness easily follows from the definition of $\leq_\psi$. The third condition follows from $R4$. We show the second condition. Assume that $(M, w) \in Mod(\psi)$ and $(M', w') \notin Mod(\psi)$. Then $Mod(\psi \circ form((M, w), (M', w'))) = \{(M, w)\}$ follows from $R2$. Therefore, $(M, w) <_\psi (M', w')$ holds.

Finally, we show $Mod(\psi \circ \varphi) = Min(Mod(\varphi), \leq_\psi)$. If φ is unsatisfiable then $Mod(\psi \circ \varphi) = \emptyset = Min(Mod(\varphi), \leq_\psi)$ holds. Hence we assume that φ is satisfiable in the following. We will show $Mod(\psi \circ \varphi) \subseteq Min(Mod(\varphi), \leq_\psi)$.[2] Assume that $(M, w) \in Mod(\psi \circ \varphi)$ and $(M, w) \notin Min(Mod(\varphi), \leq_\psi)$. By postulate $R1$, (M, w) is a model of φ. Hence, there is a model (M', w') of φ such that $(M', w') <_\psi (M, w)$.

1. $(M', w') \in Mod(\psi)$. Since $(M', w') \in Mod(\varphi)$, $\psi \wedge \varphi$ is satisfiable. Hence, by postulate $R2$, $\psi \circ \varphi \leftrightarrow \psi \wedge \varphi$ holds. Thus, $(M, w) \in Mod(\psi)$ follows from $(M, w) \in Mod(\psi \circ \varphi)$. Therefore, $(M, w) \leq_\psi (M', w')$ holds. This contradicts $(M', w') <_\psi (M, w)$.

2. $Mod(\psi \circ form((M, w), (M', w'))) = \{(M', w')\}$. Since both (M, w) and (M', w') are models of φ, $(\varphi \wedge form((M, w), (M', w'))) \leftrightarrow form((M, w), (M', w'))$ holds. Thus,

$$Mod(\psi \circ \varphi) \cap \{(M, w), (M', w')\} \subseteq Mod(\psi \circ form((M, w), (M', w')))$$

   follows from postulate $R5$. Since we assume $Mod(\psi \circ form((M, w), (M', w'))) = \{(M', w')\}$, we obtain $(M, w) \notin Mod(\psi \circ \varphi)$. This is a contradiction.

Now, to prove the other containment $Min(Mod(\varphi), \leq_\psi) \subseteq Mod(\psi \circ \varphi)$, we assume the opposite, i.e. we assume $(M, w) \in Min(Mod(\varphi), \leq_\psi)$ and $(M, w) \notin Mod(\psi \circ \varphi)$. Since we also assume that φ is internally satisfiable, it follows from postulate $R3$ that there is a multi-agent possible world (M', w') such that $(M', w') \in Mod(\psi \circ \varphi)$. Since both (M, w) and (M', w') are models of φ, $\vdash_{\mathsf{Int}} (form((M, w), (M', w')) \wedge \varphi) \leftrightarrow form((M, w), (M', w'))$ holds. By using postulates $R5$ and $R6$, we obtain

$$Mod(\psi \circ \varphi) \cap \{(M, w), (M', w')\} = Mod(\psi \circ form((M, w), (M', w')).$$

Since $(M, w) \notin Mod(\psi \circ \varphi), Mod(\psi \circ form((M, w), (M', w')) = \{(M', w')\}$ holds. Hence, $(M', w') \leq_\psi (M, w)$ holds. On the other hand, since (M, w) is minimal in $Mod(\varphi)$ with respect to $\leq_\psi$, $(M, w) \leq_\psi (M', w')$ holds. Since
$Mod(\psi \circ form((M, w), (M', w'))) = \{(M', w')\}, (M, w) \in Mod(\psi)$ holds. Therefore, $(M, w) \in Mod(\psi \circ \varphi)$ follows from postulate $R2$. This is a contradiction.

('If') Assume that there is a faithful assignment that maps ψ to a total pre-order $\leq_\psi$. We define a revision operation $\circ$ by

$$Mod(\psi \circ \varphi) = Min(Mod(\varphi), \leq_\psi).$$

[2] It should be noted that $R6$ is not used in the proof of this containment.

We show that $\circ$ satisfies conditions $R1 - R6$. It is obvious that postulate $R1$ follows from the definition of the revision operator $\circ$. It is also obvious that postulates $R3$ and $R4$ follow from the definition of the faithful assignment.

We show postulate $R2$. It suffices to show if $Mod(\psi \wedge \varphi)$ is not empty then $Mod(\psi \wedge \varphi) = Min(Mod(\varphi), \leq_\psi)$. $Mod(\psi \wedge \varphi) \subseteq Min(Mod(\varphi), \leq_\psi)$ follows from the conditions of the faithful assignment. To prove the other containment, we assume that $(M, w) \in Min(Mod(\varphi), \leq_\psi)$ and $(M, w) \notin Mod(\psi \wedge \varphi)$. Since $Mod(\psi \wedge \varphi)$ is not empty, there is an interpretation (M', w') such that $(M', w') \in Mod(\psi \wedge \varphi)$. Then $(M, w) \not\leq_\psi (M', w')$ follows from the conditions of the faithful assignment. Moreover, $(M', w') \leq_\psi (M, w)$ follows from the conditions of the faithful assignment. Hence, (M, w) is not minimal in $Mod(\varphi)$ with respect to $\leq_\psi$. This is a contradiction.

We show postulates $R5$ and $R6$. It is obvious that if $(\psi \circ \varphi) \wedge \varphi'$ is not internally satisfiable then $R6$ holds. Hence, it suffices to show that if $Min(Mod(\varphi), \leq_\psi) \cap Mod(\varphi')$ is not empty then

$$Min(Mod(\varphi), \leq_\psi) \cap Mod(\varphi') = Min(Mod(\varphi \wedge \varphi'), \leq_\psi)$$

holds.

Assume that

$$(M, w) \in Min(Mod(\varphi), \leq_\psi) \cap Mod(\varphi')$$

and

$$(M, w) \notin Min(Mod(\varphi \wedge \varphi'), \leq_\psi).$$

Then, since $(M, w) \in Mod(\varphi \wedge \varphi')$, there is a multi-agent possible world (M', w') such that $(M', w') \in Mod(\varphi \wedge \varphi')$ and $(M', w') <_\psi (M, w)$. This contradicts $(M, w) \in Min(Mod(\varphi), \leq_\psi)$. Therefore, we obtain

$$Min(Mod(\varphi), \leq_\psi) \cap Mod(\varphi') \subseteq Min(Mod(\varphi \wedge \varphi'), \leq_\psi).$$

To prove the other containment, we assume that

$$(M, w) \in Min(Mod(\varphi \wedge \varphi'), \leq_\psi)$$

and

$$(M, w) \notin Min(Mod(\varphi), \leq_\psi) \cap Mod(\varphi').$$

Since $(M, w) \in Mod(\varphi')$, $(M, w) \notin Min(Mod(\varphi), \leq_\psi)$ holds. Since we assume that $Min(Mod(\varphi), \leq_\psi) \cap Mod(\varphi')$ is not empty, suppose that (M', w') is a multi-agent possible world of $Min(Mod(\varphi), \leq_\psi) \cap Mod(\varphi')$. Then $(M', w') \in Mod(\varphi \wedge \varphi')$ holds. Since we assume that $(M, w) \in Min(Mod(\varphi \wedge \varphi'), \leq_\psi)$ and $\leq_\psi$ is total[3], $(M, w) \leq_\psi (M', w')$ holds. Thus $(M, w) \in Min(Mod(\varphi), \leq_\psi)$ follows from $(M', w') \in Min(Mod(\varphi), \leq_\psi)$. This is a contradiction. QED

This similarity between Theorem 4.2.13 and Theorem 4.2.16 is depicted in Figure 4.7. We see in this figure that possible worlds of AGM belief revision are just replaced by multi-agent possible worlds which are represented by triangles.

[3]The totality of $\leq_\psi$ is used only here in the proof of (If).

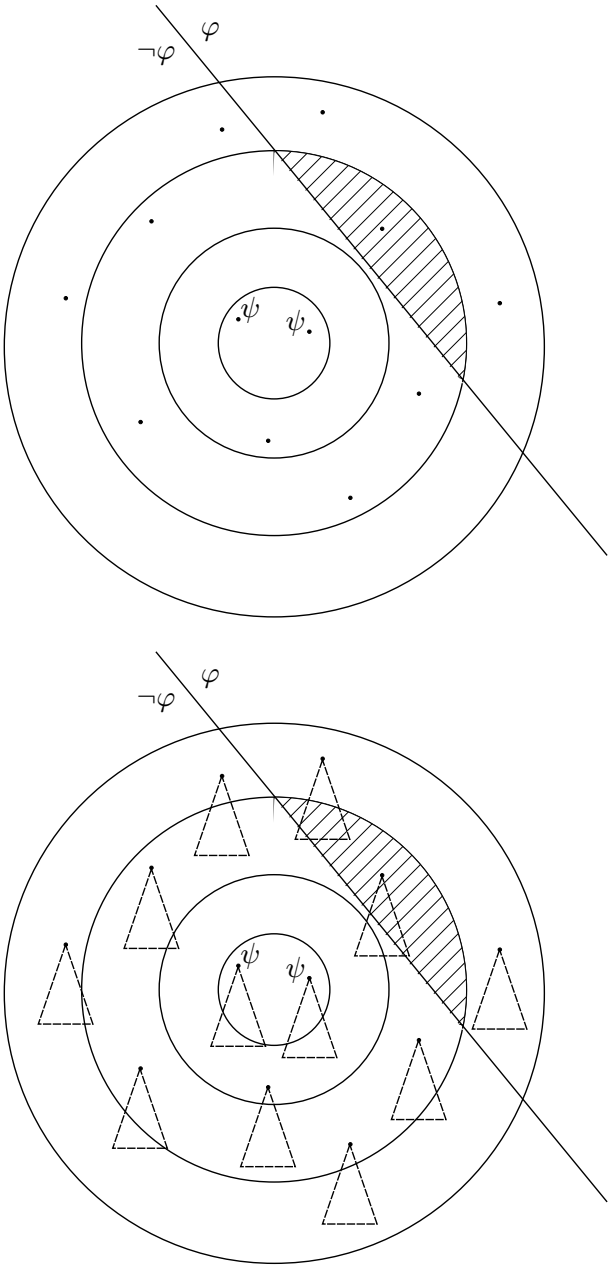


Figure 4.7: AGM belief revision (*above*) and private multi-agent belief revision (*below*)

Remark 4.2.17 (Important)
We have picked only one of the theorems of [Katsuno and Mendelzon, 1992] but in fact all the theorems present in [Katsuno and Mendelzon, 1992] transfer to the multi-agent case. It includes in particular the following theorem, where $\leq_\psi$ is a partial order instead of a total

order:

Revision operation $\circ$ satisfies postulates $R1 - R5$, $R7$ and $R8$ if and only if there exists a faithful assignment that maps each belief base ψ to a *partial* pre-order $\leq_\psi$ such that $Mod(\psi \circ \varphi) = Min(Mod(\varphi), \leq_\psi)$; where

R7 If $\vdash (\psi \circ \varphi_1) \rightarrow \varphi_2$ and $\vdash (\psi \circ \varphi_2) \rightarrow \varphi_1$ then $\vdash (\psi \circ \varphi_1) \leftrightarrow (\psi \circ \varphi_2)$.

R8 $\vdash (\psi \circ \varphi_1) \wedge (\psi \circ \varphi_2) \rightarrow \psi \circ (\varphi_1 \vee \varphi_2)$.

□

This theorem entails that we can use all the techniques and methods of AGM belief revision to deal with private multi-agent belief revision. This similarity will be exploited in the last section of this chapter. We stress that this transfer is made possible thanks to the notion of multi-agent possible world which generalizes the notion of possible world.

4.3 Multi-agent rationality postulates

In this section we are going to investigate some multi-agent rationality postulates. Indeed, because we are now in a multi-agent setting, it is natural to study how (agent Y's beliefs about) the other agents' beliefs evolve during a private announcement.

As we said, in a private announcement the beliefs of the other agents different from Y actually do not change, and agent Y knows this. Consequently, agent Y's beliefs about the agents who are not concerned by the formula announced should not change as well. To formalize this idea, we first need to define who are the agents who are concerned by a formula.

4.3.1 On the kind of information a formula is about

First note that an input may not only concern agents but also the objective state of nature, i.e. propositional facts, that we write pf. For example, the formula $p \wedge B_j B_i \neg p$ concerns agent j's beliefs but also propositional facts (namely p). Besides, a formula cannot be about Y's beliefs because $\varphi \in \mathcal{L}^C_{\neq Y}$ by assumption. So what an input is about includes propositional facts but excludes agent Y's beliefs. This leads us to the following definition.

Definition 4.3.1 (Operator A)
Let $A_0 = (G \cup \{\mathtt{pf}\}) - \{Y\}$.
We define by induction the agents that a formula is about as follows:

- $A(p) = \mathtt{pf}$; $A(B_j\varphi) = \{j\}$;
- $A(\neg\varphi) = A(\varphi)$; $A(\varphi \wedge \varphi') = A(\varphi) \cup A(\varphi')$.

□

For example, $A(p \vee (q \wedge B_j B_i r) \wedge B_k r)) = \{\mathtt{pf}, j, k\}$, and $A(B_i p \vee B_j B_k \neg p) = \{i, j\}$.

We then define a language $\mathcal{L}_{A_1}$ whose formulas concern only agents in A_1, and possibly propositional facts if $\mathtt{pf} \in A_1$.

Definition 4.3.2 (Language $\mathcal{L}^C_{A_1}$)
Let $A_1 \subseteq A_0$. We define the language $\mathcal{L}^C_{A_1}$ as follows.

$$\mathcal{L}^C_{A_1} : \varphi ::= \top \mid P \mid B_j \psi \mid \varphi \wedge \varphi \mid \neg \varphi,$$

where j ranges over $A_1 - \{\mathtt{pf}\}$ and ψ over formulas of $\mathcal{L}^C$ as defined in Definition 2.2.10. Besides, $P = \Phi$ if $\mathtt{pf} \in A_1$ and $P = \emptyset$ otherwise. □

Now we define a notion supposed to tell us whether two pointed and finite epistemic models contain the same information about some agents' beliefs and possibly about propositional facts.

Definition 4.3.3 (A_1-bisimilarity)
Let $A_1 \subseteq A_0$. We say that (M, w) and (M', w') are A_1-*bisimilar*, written $M, w \leftrightarrows_{A_1} M', w'$, iff

- if $\mathtt{pf} \in A_1$ then for all $p \in \Phi$, $w \in V(p)$ iff $w' \in V'(p)$; and
- for all $j_1 \in A_1$,
  if $v \in R_{j_1}(w)$ then there is $v' \in R_{j_1}(w')$ such that $M, v \leftrightarrows M', v'$,
  if $v' \in R_{j_1}(w')$ then there is $v \in R_{j_1}(w)$ such that $M, v \leftrightarrows M', v'$.

□

Proposition 4.3.4 *Let $A_1 \subseteq A_0$. Then,*

$$M, w \leftrightarrows_{A_1} M', w' \text{ iff for all } \varphi \in \mathcal{L}^C_{A_1},\ M, w \models \varphi \text{ iff } M', w' \models \varphi.$$

PROOF. We assume that $\mathtt{pf} \in A_1$, the proof without this assumption is essentially the same.

- Assume $M, w \leftrightarrows_{A_1} M', w'$. We are going to prove by induction on $\varphi \in \mathcal{L}^C_{A_1}$ that $M, w \models \varphi$ iff $M', w' \models \varphi$.
  - $\varphi = p$. As $\mathtt{pf} \in A_1$, $M, w \models p$ iff $M', w' \models p$.
  - $\varphi = \varphi_1 \wedge \varphi_2, \varphi = \neg \varphi'$ work by induction hypothesis.
  - $\varphi = B_{j_1} \varphi', j_1 \in A_1$. Assume $M, w \models B_{j_1} \varphi'$ then for all $v \in R_{j_1}$, $M, v \models \varphi'$ (*). But for all $v' \in R_{j_1}(w')$ there is $v \in R_{j_1}(w)$ such that $M, v \leftrightarrows M', v'$.
    So for all $v' \in R_{j_1}(w')$, $M', v' \models \varphi'$ by property of the bisimulation and (*). Finally $M', w' \models B_{j_1} \varphi'$, i.e. $M', w' \models \varphi$.
    The other way around we could show that if $M', w' \models B_{j_1} \varphi$ then $M, w \models B_{j_1} \varphi$.
- Assume that for all $\varphi \in \mathcal{L}^C_{A_1}$, $M, w \models \varphi$ iff $M', w' \models \varphi$ (*).

– Clearly for all $p \in \Phi$, $w \in V(p)$ iff $w' \in V'(p)$.
– Let $j_1 \in A_1$ and $v \in R_{j_1}(w)$.
Assume for all $v' \in R_{j_1}(w')$ it is not the case that $M, v \leftrightarrows M', v'$ (**).
Then for all $v' \in R_{j_1}(w')$ there is $\varphi(v') \in \mathcal{L}$ such that $M, v \models \neg\varphi(v')$ and $M', v' \models \varphi(v')$.
As by hypothesis W' is finite, let $\varphi(w') = B_{j_1}\left(\bigvee_{v' \in R_{j_1}(w')} \varphi(v')\right)$; then $\varphi(w') \in \mathcal{L}^C_{A_1}$.
Besides $M', w' \models \varphi(w')$ but $M, w \models \neg\varphi(w')$. This is impossible by (*), so (**) is false.
The other part of the definition of $\leftrightarrows_{A_1}$ is proved similarly.

QED

Proposition 4.3.4 ensures us that the notion we just defined captures what we wanted. Its proof uses that the models are finite (otherwise the if direction would not hold).

Definition 4.3.5 Let $\mathcal{M}$ and $\mathcal{M}'$ be two sets of multi-agent possible worlds, we set $\mathcal{M} \leftrightarrows_{A_1} \mathcal{M}'$ iff for all $(M, w) \in \mathcal{M}$ there is $(M', w') \in \mathcal{M}'$ such that $M, w \leftrightarrows_{A_1} M', w'$, and for all $(M', w') \in \mathcal{M}'$ there is $(M, w) \in \mathcal{M}$ such that $M, w \leftrightarrows_{A_1} M', w'$. □

4.3.2 Some rationality postulates specific to our multi-agent approach

As we said before, in a private announcement, agent Y's beliefs about the beliefs of the agents who are not concerned by the formula should not change. This can be captured by the following postulate:

RG1 Let $\psi, \varphi, \varphi' \in \mathcal{L}^C_{\not\in Y}$ such that $A(\varphi) \cap A(\varphi') = \emptyset$.
If $\vdash_{\mathsf{Int}} \psi \to \varphi'$ then $\vdash_{\mathsf{Int}} (\psi \circ \varphi) \to \varphi'$

This postulate is the multi-agent version of Parikh and Chopra's postulate [Chopra and Parikh, 1999]. The example in the paragraph just before Example 4.2.9 illustrates this postulate: there, $\varphi = \neg p$ and $\varphi' = B_j p \wedge B_j C_G p$. Now the semantic counterpart of $RG1$:

Proposition 4.3.6 *Revision operation $\circ$ satisfies $RG1$ iff for all $\varphi \in \mathcal{L}^C_{\not\in Y}$, for all $(M', w') \in Mod(\psi \circ \varphi)$, there is $(M, w) \in Mod(\psi)$ such that $M, w \leftrightarrows_{A'} M', w'$, with $A' = A_0 - A(\varphi)$.*

PROOF. We first prove a lemma which is the counterpart of Proposition 2.2.16 for this notion of A_1-bisimilarity.

Lemma 4.3.7 *Let $A_1 \subseteq A_0$, let M be a finite epistemic model and $w \in M$. Then there is a formula $\delta^{A_1}_M(w)$ such that*

1. $M, w \models \delta^{A_1}_M(w)$;

2. *for every finite epistemic model M' and world $w' \in M'$, if $M', w' \models \delta_M^{A_1}(w)$ then $M, w \leftrightarrows_{A_1} M', w'$.*

PROOF. We only sketch the proof. If $\mathtt{pf} \in A_1$, take

$$\delta_M^{A_1}(w) = \bigwedge_{\{p|w\in V(p)\}} p \wedge \bigwedge_{\{p|w\notin V(p)\}} \neg p \wedge \bigwedge_{j\in A_1} \left(\bigwedge_{v\in R_j(w)} \hat{B}_j \delta_M(v) \wedge B_j \left(\bigvee_{v\in R_j(w)} \delta_M(v) \right) \right)$$

otherwise if $\mathtt{pf} \notin A_1$, take

$$\delta_M^{A_1}(w) = \bigwedge_{j\in A_1} \left(\bigwedge_{v\in R_j(w)} \hat{B}_j \delta_M(v) \wedge B_j \left(\bigvee_{v\in R_j(w)} \delta_M(v) \right) \right)$$ QED

We now prove the proposition. The "if" part is straightforward. Let us prove the "only if" part. Let $\varphi \in \mathcal{L}^C_{\neq Y}$ and let $(M', w') \in Mod(\psi \circ \varphi)$. Assume that for all $(M, w) \in Mod(\psi)$, it is not the case that $M', w' \leftrightarrows_{A'} M, w$. Then for all $(M, w) \in Mod(\psi)$, $M, w \models \neg\delta_{M'}^{A'}(w')$ by lemma 4.3.7. So $\vdash_{\mathsf{Int}} \psi \to \neg\delta_{M'}^{A'}(w')$. Then $\vdash_{\mathsf{Int}} \psi \circ \varphi \to \neg\delta_{M'}^{A'}(w')$. Hence $M', w' \models \neg\delta_{M'}^{A'}(w')$, which is contradictory. QED

Let us consider the converse of $RG1$.

RG2 Let $\psi, \varphi, \varphi' \in \mathcal{L}^C_{\neq Y}$ such that $A(\varphi) \cap A(\varphi') = \emptyset$.
If $\psi \wedge \varphi'$ is internally satisfiable then $(\psi \circ \varphi) \wedge \varphi'$ is internally satisfiable.

This postulate means that if agent Y considers φ' originally possible ($\psi \wedge \varphi'$ is internally satisfiable) then after revising by φ, which is not about the same kind of information ($A(\varphi) \cap A(\varphi') = \emptyset$), agent Y still considers φ' possible ($(\psi \circ \varphi) \wedge \varphi'$ is internally satisfiable). (This postulate is formally equivalent to: If $\vdash_{\mathsf{Int}} (\psi \circ \varphi) \to \varphi'$ then $\vdash_{\mathsf{Int}} \psi \to \varphi'$.) Its semantic counterpart is:

Proposition 4.3.8 *Revision operation $\circ$ satisfies $RG2$ iff for all $\varphi \in \mathcal{L}^C_{\neq Y}$, for all $(M, w) \in Mod(\psi)$ there is $(M', w') \in Mod(\psi \circ \varphi)$ such that $M, w \leftrightarrows_{A'} M', w'$, with $A' = A_0 - A(\varphi)$.*

PROOF. Similar to Proposition 4.3.6. QED

Unlike $RG1$, $RG2$ is not really suitable for revision because all the worlds representing Y's epistemic state 'survive' the revision process if $RG2$ is fulfilled. This is not the case in general because new information can discard some previous possibilities. This is however the case for update where we apply the update process to each world independently (see [Katsuno and Mendelzon, 1991] for an in depth analysis). So $RG2$ is more suitable for an update operation. In that respect, one can show that all the results about propositional update in [Katsuno and Mendelzon, 1992] transfer to the multi-agent case as well, and so still thanks to our notion of multi-agent possible world.

For example, consider $\psi = B_i p \vee B_j p$ and $\varphi = \neg B_i p$. Then the revised formula is $\psi \circ \varphi = B_j p \wedge \neg B_i p$ according to postulate $R2$. But according to postulate $RG2$, after the revision $\neg B_j p$ should be satisfiable because $\psi \wedge \neg B_j p$ was satisfiable. This exemplifies the similarity of postulate RG2 with the propositional update postulate $U8$.

Postulates $RG1$ and $RG2$ together are equivalent to:

RG1+RG2 Let $\psi, \varphi, \varphi' \in \mathcal{L}^C_{\neq Y}$ such that $A(\varphi) \cap A(\varphi') = \emptyset$.
$\vdash_{\mathsf{Int}} \psi \to \varphi'$ iff $\vdash_{\mathsf{Int}} (\psi \circ \varphi) \to \varphi'$.

And the semantic counterpart:

Proposition 4.3.9 *Revision operation $\circ$ satisfies $RG1$ and $RG2$ iff for all $\varphi \in \mathcal{L}^C_{\neq Y}$, $Mod(\psi) \leftrightarrows_{A'} Mod(\psi \circ \varphi)$, with $A' = A_0 - A(\varphi)$.*

4.4 A revision operation

In this section we propose a revision operation based on a degree of similarity between multi-agent possible worlds defined very much in the same way as in [Lehmann *et al.*, 2001; Ben-Naim, 2006]. Besides, for sake of simplicity, we assume that formulas representing belief bases and private announcements belong to the language associated to Y *without* common belief, written $\mathcal{L}_{\neq Y}$:

$$\mathcal{L}_{\neq Y} : \varphi ::= \top \mid p \mid B_j \psi \mid \varphi \wedge \varphi \mid \neg \varphi,$$

where ψ ranges over $\mathcal{L}$ and j over $G - \{Y\}$. One should note that in this setting, the 'if' direction of Theorem 4.2.16 and Proposition 4.3.6 still hold, but not the 'only if' direction.

4.4.1 Mathematical preliminaries

Anti-lexicographic ordering

We first recall the definition of an anti-lexicographic ordering.

Definition 4.4.1 (Anti-lexicographic ordering)
Let $k \in \mathbb{N}$ and $(l_0, \ldots, l_k), (l'_0, \ldots, l'_k) \in [0; 1]^{k+1}$. We set

$$(l_0, \ldots, l_k) <^k (l'_0, \ldots, l'_k) \quad \text{iff} \quad \begin{cases} l_k < l'_k \text{ or} \\ l_k = l'_k, \ldots, l_{k-j+1} = l'_{k-j+1} \text{ and } l_{k-j} < l'_{k-j} \\ \text{for some } 1 \leq j \leq k. \end{cases}$$

□

Now we define the *Supremum* of a set of tuples with respect to the anti-lexicographic ordering by using the supremum Sup of real numbers.

Definition 4.4.2 (Anti-lexicographic supremum)
Let $k \in \mathbb{N}$ and $\{(l_0^i, \ldots, l_k^i) \mid i \in S\} \subseteq [0;1]^{k+1}$ (where S is an index set which is possibly infinite). The *anti-lexicographic supremum* $Sup^k\{(l_0^i, \ldots, l_k^i) \mid i \in S\} = (A_0, \ldots, A_k)$ is defined as follows.

$$A_k = Sup\{l_k^i \mid i \in S\}; \text{and for all } m < k,$$

$$A_m = \begin{cases} Sup\{l_m^i \mid l_j^i = A_j \text{ for all } k \geq j > m\} & \text{if there is } i \text{ such that } l_j^i = A_j \\ & \text{for all } k \geq j > m \\ Sup\{l_m^i \mid i \in S\} & \text{otherwise.} \end{cases}$$

where Sup is the usual supremum on real numbers. □

This definition is well-defined because the supremum of a non-empty set of real numbers with an upper bound always exists. Finally, we check that this anti-lexicographic supremum does correspond to the maximum of tuples when this one exists.

Proposition 4.4.3 *Let $L = \{(l_0^i, \ldots, l_k^i) \mid i \in S\} \subseteq [0;1]^{k+1}$ and $(l_0^{i_0}, \ldots, l_k^{i_0}) \in L$ (where S is an index set which is possibly infinite).*

$$\text{If } (l_0^{i_0}, \ldots, l_k^{i_0}) \geq^k (l_0^i, \ldots, l_k^i) \text{ for all } i \in S, \text{ then } (l_0^{i_0}, \ldots, l_k^{i_0}) = Sup^k(L).$$

PROOF. Let $(A_0, \ldots, A_k) = Sup^k(L)$. We prove by induction on m that $A_m = l_m^{i_0}$.

- $A_k = Sup\{l_k^i \mid i \in S\} = l_k^{i_0}$ by definition of $\leq^k$.
- Assume for all $k \geq j > m$ that $l_j^{i_0} = A_j$. Then
  $A_m = Sup\{l_m^i \mid l_j^i = A_j \text{ for all } j > m\}$
  $= Sup\{l_m^i \mid l_j^i = l_j^{i_0} \text{ for all } j > m\}$ by induction hypothesis
  $= l_m^{i_0}$.

QED

n-bisimulation

Our definition of n-bisimulation is a slight modification of the definition of n-bisimulation in [Balbiani and Herzig, 2007; Blackburn *et al.*, 2001].

Definition 4.4.4 (n-bisimulation)
Let $M = (W, R, V)$ and $M' = (W', R', V')$ be two epistemic models, and let $w \in M$, $w' \in M'$. Let $Z \subseteq W \times W'$. We recursively define the property of Z being *n-bisimulation* in w and w', written $Z : M, w \leftrightarrows_n M', w'$:

1. $Z : M, w \leftrightarrows_0 M', w'$ iff wZw' and there is $p \in \Phi$ such that $w \in V(p)$ and $w' \notin V'(p)$;

2. $Z : M, w \leftrightarrows_1 M', w'$ iff wZw' and for all $p \in \Phi$, $w \in V(p)$ iff $w' \in V'(p)$;

3. For all $n \geq 1$, $Z : M, w \leftrightarrows_{n+1} M', w'$ iff wZw' and $val(w) = val(w')$ and for all $j \in G$,

   - for all $v \in R_j(w)$ there is $v' \in R_j(w')$ such that $Z : M, v \leftrightarrows_n M', v'$.
   - for all $v' \in R_j(w')$ there is $v \in R_j(w)$ such that $Z : M, v \leftrightarrows_n M', v'$.

Now we can define *n-bisimilarity* between w and w', written $M, w \leftrightarrows_n M', w'$ by $M, w \leftrightarrows_n M', w'$ iff there exists a relation Z such that $Z : M, w \leftrightarrows_n M', w'$. □

Two worlds being n-bisimilar (with $n \geq 1$) intuitively means that they have the same modal structure up to modal depth $n-1$, and thus they satisfy the same formulas of degree at most $n-1$. For example, in the epistemic models of Figure 4.8, we have $M, w \leftrightarrows_1 M', w'$, but $M, w \leftrightarrows_2 M', w'$ is not the case.

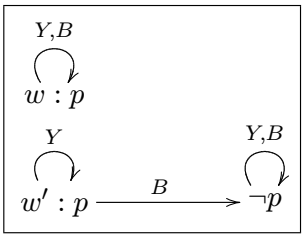


Figure 4.8: Epistemic model (M, w) (*above*) and (M', w') (*below*)

The usual definition of Z being a bisimulation corresponds to $Z : M, w \leftrightarrows_n M', w'$ for all $n \in \mathbb{N}^*$. In fact, it suffices that two finite epistemic models be n-bisimilar up to a certain modal depth to be bisimilar, as the following proposition shows.

Proposition 4.4.5 [Balbiani, 2007]
Let M and M' be two finite epistemic models and $w \in M$, $w' \in M'$. Let $n = |M| \cdot |M'| + 1$. Then,

$$M, w \leftrightarrows_n M', w' \text{ iff } M, w \leftrightarrows M', w'.$$

4.4.2 Definition of the revision operation

First we are going to define a degree of similarity between two multi-agent possible worlds that will allow for an anti-lexicographic order.

Definition 4.4.6 (Degree of similarity between multi-agent possible worlds)
Let (M, w) and (M', w') be two multi-agent possible worlds, let $v \in M$ and $v' \in M'$, let S and S' be two finite sets of possible worlds, and let $\mathcal{M}$ and $\mathcal{M}'$ be two sets of multi-agent possible worlds (possibly infinite). Let $n = |M| \cdot |M'| + 1$ and $k \in \mathbb{N}$.

If E is a finite set of real numbers, we write $m(E)$ the average of E, i.e. $m(E) = \frac{1}{|E|} \sum_{e \in E} e$.

- $\sigma(v, v') = max\{\frac{i}{n} \mid M, v \leftrightarrows_i M', v' \text{ and } i \in \{0, \ldots, n\}\}$;

- $\sigma(S, S') = \frac{1}{2}\left(m\{\sigma(s, S') \mid s \in S\} + m\{\sigma(S, s') \mid s' \in S'\}\right)$
  where $\sigma(s, S') = max\{\sigma(s, s') \mid s' \in S'\}$ and $\sigma(S, s') = max\{\sigma(s, s') \mid s \in S\}$;
- $s^k((M, w), (M', w')) = (\sigma(w, w'), m\{\sigma\left(R_j(w), R_j(w')\right) \mid j \in G, j \neq Y\}, \ldots,$
  $m\{\sigma\left(R_{j_1} \circ \ldots \circ R_{j_k}(w), R_{j_1} \circ \ldots \circ R_{j_k}(w')\right) \mid j_1, \ldots, j_k \in G, j_i \neq j_{i+1}, j_1 \neq Y\})$;
- $s^k\left(\mathcal{M}, \mathcal{M}'\right) = Sup^k\{s^k\left((M, w), (M', w')\right) \mid (M, w) \in \mathcal{M}, (M', w') \in \mathcal{M}'\}$.

□

$\sigma(v, v')$ measures a degree of similarity between the worlds v and v'. For example in Figure 4.8, we have $\sigma(w, w') = \frac{1}{3}$. Note that $0 \leq \sigma(v, v') \leq 1$ for all v and v'. If $\sigma(v, v') = 1$ then the worlds v and v' are bisimilar by Proposition 4.4.5. So their degree of similarity is the highest possible. If $\sigma(v, v') = 0$, that is $M, v \leftrightarrows_0 M', v'$ then their degree of similarity is the lowest possible because they differ even on propositional facts. Likewise, $\sigma(S, S')$ measures a degree of similarity between the sets of worlds S and S'. Note also that $0 \leq \sigma(S, S') \leq 1$ for all S and S'. If $\sigma(S, S') = 1$ then for all worlds $v \in S$ there is $v' \in S'$ such that v is bisimilar with v', and vice versa, for all $v' \in S'$ there is $v \in S$ such that v' is bisimilar with v. So the degree of similarity between S and S' is the highest possible. If $\sigma(S, S') = 0$ then for all $v \in S$ there is no $v' \in S'$ such that v and v' agree on all propositional letters, and vice versa, for all $v' \in S'$ there is no $v \in S$ such that v and v' agree on all propositional letters. So the degree of similarity is the lowest possible. To be more precise, $\sigma(v, S')$ is the degree of similarity of a world v with S'. So $m\{\sigma(v, S') \mid v \in S\}$ is the average degree of similarity of a world $v \in S$ with S'. Likewise, $m\{\sigma(S, v') \mid v' \in S'\}$ is the average degree of similarity of a world $v' \in S'$ with S. So the degree of similarity between S and S' is just the average of these two degrees. $s^k((M, w), (M', w'))$ is a tuple which represents by how much two multi-agent possible worlds are similar relatively to their respective modal depth. For example in Figure 4.8 we have $s^2((M, w), (M', w')) = (\frac{1}{3}, 0, 0)$. Note that for a given modal depth we only compare the degree of similarity of worlds which have the same history (i.e. they are all accessed from w and w' by the same sequence of accessibility relations $R_{j_1}, \ldots, R_{j_k}$). Doing so, in our comparison we stick very much to the modal structure of both multi-agent possible worlds. Besides we take the average of their degree of similarity for every possible history in order to give the same importance to these different possible histories.

Definition 4.4.7 (Revision operation $\circ$)
Let $\psi \in \mathcal{L}_{\neq Y}$ and $k = deg(\psi)+1$. We assign to ψ a total pre-order $\leq_\psi$ on multi-agent possible worlds defined as follows:

$$(M, w) \leq_\psi (M', w') \text{ iff } s^k(Mod(\psi), (M, w)) \geq^k s^k(Mod(\psi), (M', w')).$$

The revision operation $\circ$ associated to this pre-order $\leq_\psi$ is defined semantically in the usual way (see Theorem 4.2.13) by:

$$Mod(\psi \circ \varphi) = Min(Mod(\varphi), \leq_\psi).$$

□

So (M, w) is closer to ψ than (M', w') when its degree of similarity with the models of ψ is higher than the degree of similarity of (M', w') with the models of ψ. In the next section, we are going to motivate our use of anti-lexicographic ordering and explain why we compare the modal structures of the multi-agent possible worlds only until modal depth $k = deg(\psi)+1$.

4.4.3 Properties of the revision operation

Proposition 4.4.8 *Let (M, w) be a multi-agent possible world and $\psi \in \mathcal{L}_{\neq Y}$ a satisfiable formula such that $deg(\psi) = d$. Then there is $(M_\psi, w_\psi) \in Mod(\psi)$ such that*

$$m\{\sigma(R_{j_1} \circ \ldots \circ R_{j_{d+1}}(w), R_{j_1} \circ \ldots \circ R_{j_{d+1}}(w_\psi)) \mid j_1, \ldots, j_{d+1} \in G, j_i \neq j_{i+1}, j_1 \neq Y\} = 1.$$

PROOF. We first need to introduce a technical device that will be used in the proof of the proposition.

Definition 4.4.9 (Tree-like multi-agent possible world)
Let $d \in \mathbb{N}$. A *tree-like multi-agent possible world of height d* is a finite pointed epistemic model $(M^t, w^t) = (W^t, R^t, V^t, w^t)$ of height d (see Definition 2.2.15) generated by w^t such that:

1. $R_Y(w^t) = \{w^t\}$;
2. for all $j \in G$, R_j is transitive and euclidean;
3. for all $v^t \neq w^t$ there are two unique sequences $v^t_0 = w^t, \ldots, v^t_n = v^t$ and $j_1, \ldots, j_n$ such that $j_i \neq j_{i+1}$, $j_1 \neq Y$ and $w^t = v^t_0 R_{j_1} v^t_1 R_{j_2} \ldots R_{j_n} v^t_n = v^t$;
4. for all v^t and j such that $v^t \in R_j(v^t)$,
   - if $h(v^t) < d$ then for all i, $R_i(v^t) \neq \emptyset$;
   - if $h(v^t) = d$ then for all $i \neq j$, $R_i(v^t) = \emptyset$.

□

Now we can prove the proposition.

- One can easily show that there is a tree-like multi-agent possible world of height d, $(M^t, w^t) = (W^t, R^t, V^t, w^t)$, such that:
  - $M^t, w^t \models \psi$
  - for all $j_1, \ldots, j_d$ with $j_i \neq j_{i+1}$, $j_1 \neq Y$ $|R_{j_1} \circ \ldots \circ R_{j_d}(w^t)| \geq |R_{j_1} \circ \ldots \circ R_{j_d}(w)|$.

  For all $j_1, \ldots, j_d$, let $f_{j_1,\ldots,j_d}$ be a surjection from $R_{j_1} \circ \ldots \circ R_{j_d}(w^t)$ to $R_{j_1} \circ \ldots \circ R_{j_d}(w)$.
  For all $j_1, \ldots, j_d$ and $v^t \in R_{j_1} \circ \ldots \circ R_{j_d}(w^t)$, we write $M^{v^t} = (W^{v^t}, R^{v^t}, V^{v^t})$ the submodel of M generated by $\bigcup_{i \neq j_d} R_i(f_{j_1,\ldots,j_d}(v^t))$. Then we define $Plug((M^t, w^t), (M, w)) = (W', R', V', w')$ as follows.

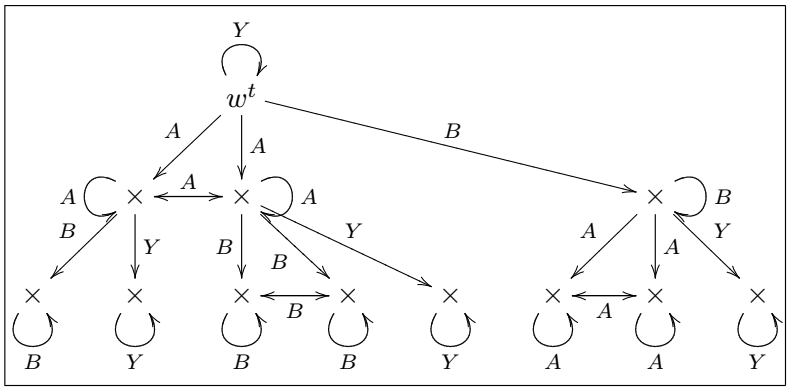


Figure 4.9: A tree-like multi-agent possible world of height 2

- $W' = W^t \cup \{W^{v^t} \mid v^t \in R_{j_1} \circ \ldots \circ R_{j_d}(w^t), j_i \neq j_{i+1}, j_1 \neq Y\}$;
- $R'_j = R^t_j \cup \{R^{v^t}_j \mid v^t \in R_{j_1} \circ \ldots \circ R_{j_d}(w^t), j_i \neq j_{i+1}, j_1 \neq Y\} \cup \{(v^t, u^t) \mid v^t \in R_{j_1} \circ \ldots \circ R_{j_d}(w^t), j_i \neq j_{i+1}, j_1 \neq Y, j_d \neq j, u^t \in R_j(f_{j_1,\ldots,j_d}(v^t))$(in M)$\}$;
- $V'(p) = V^t(p) \cup \{V^{f_{j_1,\ldots,j_d}(v^t)}(p) \mid v^t \in R_{j_1} \circ \ldots \circ R_{j_d}(w^t), j_i \neq j_{i+1}, j_1 \neq Y\}$;
- $w' = w^t$.

- Now we prove that $Plug((M^t, w^t), (M, w))$ is a multi-agent possible world. We first prove that $Plug((M^t, w^t), (M, w))$ is serial.
  - For all v such that $h(v) < d$, $R_j(v) \neq \emptyset$ for all j by definition of a tree-like multi-agent possible world;
  - for all v such that $h(v) > d$, $R_j(v) \neq \emptyset$ for all j by definition of a generated submodel;
  - for all v such that $h(v) = d$, $R_j(v) \neq \emptyset$ for all j by definition of R'_j.

  We prove that condition 2 of the definition of a multi-agent possible world is fulfilled.
  - If $d = 0$ then condition 2 is fulfilled by definition of R'_j;
  - if $d > 0$ then condition 2 is fulfilled by condition 3 of the definition of a tree-like multi-agent possible world.

  The other conditions are obvious.

- Because $deg(\psi) = d$ and the restriction of $Plug((M^t, w^t), (M, w))$ to the worlds of height at most d is bisimilar to (M^t, w^t), we get that $Plug((M^t, w^t), (M, w)), w' \models \psi$.

- Let $v' \in R_{j_1} \circ \ldots \circ R_{j_{d+1}}(w')$ with $j_i \neq j_{i+1}$ and $j_1 \neq Y$. Then there is $v^t \in R_{j_1} \circ \ldots \circ R_{j_d}(w')$ such that $v' \in R_{j_{d+1}}(v^t)$. Then $f_{j_1,\ldots,j_d}(v^t) \in R_{j_1} \circ \ldots \circ R_{j_d}(w)$ and there is $v \in R_{j_{d+1}}(f_{j_1,\ldots,j_d}(v^t))$ such that $Plug((M^t, w^t), (M, w)), v' \leftrightarrows M, v$ by definition of $Plug((M^t, w^t), (M, w))$. So $v \in R_{j_1} \circ \ldots \circ R_{j_{d+1}}(w)$ and $M, v \leftrightarrows Plug((M^t, w^t), (M, w)), v'$.

Likewise, let $v \in R_{j_1} \circ \ldots \circ R_{j_{d+1}}(w)$ with $j_i \neq j_{i+1}$ and $j_1 \neq Y$. Then there is $u \in R_{j_1} \circ \ldots \circ R_{j_d}(w)$ such that $v \in R_{j_{d+1}}(u)$. Then there is $v^t \in R_{j_1} \circ \ldots \circ R_{j_d}(w')$ such that $f_{j_1,\ldots,j_d}(v^t) = u$ because $f_{j_1,\ldots,j_d}$ is surjective. Besides, there is $v' \in R_{j_{d+1}}(v^t)$ such that $Plug((M^t, w^t), (M, w)), v' \leftrightarrows M, v$ by definition of $Plug((M^t, w^t), (M, w))$. So $v' \in R_{j_1} \circ \ldots \circ R_{j_{d+1}}(w')$ and $Plug((M^t, w^t), (M, w)), v' \leftrightarrows M, v$.

So $\sigma(R_{j_1} \circ \ldots \circ R_{j_{d+1}}(w), R_{j_1} \circ \ldots \circ R_{j_{d+1}}(w')) = 1$ for all $j_1, \ldots, j_{d+1}$ such that $j_i \neq j_{i+1}$ and $j_1 \neq Y$.

Therefore $m\{\sigma\left(R_{j_1} \circ \ldots \circ R_{j_{d+1}}(w), R_{j_1} \circ \ldots \circ R_{j_{d+1}}(w')\right) \mid j_1, \ldots, j_{d+1} \in G, j_i \neq j_{i+1}, j_1 \neq Y\} = 1$.

- Finally, we define (M_ψ, w_ψ) as the bisimulation contraction of $Plug((M^t, w^t), (M, w))$. Then all the results for $Plug((M^t, w^t), (M, w))$ still hold for (M_ψ, w_ψ) and besides $(M_\psi, w_\psi) \in Mod(\psi)$.

QED

This proposition tells us that, given a formula ψ of degree d and a multi-agent possible world (M, w), there is a multi-agent possible world that satisfies ψ and whose structure is the same as (M, w) beyond modal depth d. That is why, in $s^k(Mod(\psi), (M, w))$, we stop at modal depth $k = d + 1$ when we compare models of ψ with (M, w): we know that there is anyway a model of ψ whose modal structure is the same as (M, w) beyond this modal depth, so there is no need to check it further. Moreover, we would like to give priority to this similarity when we compare models of ψ with (M, w). That is to say, we would like to ensure that the models of ψ closest to (M, w) are such that their modal structure beyond this modal depth is the same as the one of (M, w). We do so by using the anti-lexicographic order defined in Definition 4.4.1.

The following proposition shows that we need to consider only finitely many models of ψ in $s^k(Mod(\psi), (M, w)) = Sup^k\{s^k((M', w'), (M, w)) \mid (M', w') \in Mod(\varphi)\}$.

Proposition 4.4.10 *Let (M, w) be a multi-agent possible world. For all $k \in \mathbb{N}^*$, there are finitely many multi-agent possible worlds (M', w') such that*

$$m\{\sigma(R_{j_1} \circ \ldots \circ R_{j_k}(w'), R_{j_1} \circ \ldots \circ R_{j_k}(w)) \mid j_1, \ldots, j_k \in G, j_i \neq j_{i+1}, j_1 \neq Y\} = 1.$$

PROOF. We first prove a lemma.

Lemma 4.4.11 *Let $\mathcal{M} = \{(M^1, w^1), \ldots, (M^n, w^n)\}$ be an internal model of type 1 for agent j.*[4]

For all $k \in \mathbb{N}$, there are finitely many multi-agent possible worlds (M', w') for agent j such that

$(*)$ for all $j_1, \ldots, j_k$ with $j_1 \neq j$ and $j_i \neq j_{i+1}$, for all $v \in R_{j_1} \circ \ldots \circ R_{j_k}(w')$, there is $(M^i, w^i) \in \mathcal{M}$ and $v^i \in R_{j_1} \circ \ldots \circ R_{j_k}(w^i)$ such that $M, v \leftrightarrows M^i, v^i$.

[4] An internal model or a multi-agent possible world *for agent j* is an internal model or a multi-agent possible world where the designated agent is j instead of Y.

PROOF. First, note that every multi-agent possible world (M', w') for agent j can be seen as the 'connection' of an interpretation (the root w) with a finite number of multi-agent possible worlds for each agent $l \neq j$.

Now we prove the lemma by induction on k.

k=1 Because Φ is finite, there are finitely many interpretations. So there are finitely many (valuations for the) roots of multi-agent possible worlds.

By $(*)$, there are also finitely many worlds accessible from each root modulo bisimulation. So, by the remark at the beginning of this proof, there are finitely many multi-agent possible worlds satisfying $(*)$.

k+1 For all $l \neq j$, for all $(M^i, w^i) \in \mathcal{M}$, let M^i_l be the submodel of M^i generated by $R_l(w^i)$. M^i_l is an internal model of type 2 for agent l. Let $\mathcal{M}^i_l = \{(M^1_l, w^1_l), \ldots, (M^{n_i}_l, w^{n_i}_l)\}$ be its associated internal model of type 1. Let $\mathcal{M}_l = \bigcup_{i \in \{1,\ldots,n\}} \mathcal{M}^i_l = \bigcup_{i \in \{1,\ldots,n\}} \{(M^1_l, w^1_l), \ldots, (M^{n_i}_l, w^{n_i}_l)\}$.

Now, using the remark at the beginning of this proof,

there are finitely many multi-agent possible worlds for agent j satisfying (*)

iff for all $l \neq j$ there are finitely many multi-agent possible worlds (M', w') for agent l such that for all $j_1, \ldots, j_k$ with $j_1 \neq l$ and $j_i \neq j_{i+1}$,
for all $v' \in R_{j_1} \circ \ldots \circ R_{j_k}(w')$, there is $(M^i, w^i) \in \mathcal{M}$ and $v^i \in R_l \circ R_{j_1} \circ \ldots \circ R_{j_k}(w^i)$ such that $M', v' \leftrightarrows M^i, v^i$.

iff for all $l \neq j$ there are finitely many multi-agent possible worlds (M', w') for agent l such that
for all $j_1, \ldots, j_k$ with $j_1 \neq l$ and $j_i \neq j_{i+1}$,
for all $v' \in R_{j_1} \circ \ldots \circ R_{j_k}(w')$, there is $(M^i_l, w^i_l) \in \mathcal{M}_l$ and $v^i_l \in R_{j_1} \circ \ldots \circ R_{j_k}(w^i_l)$ such that $M', v' \leftrightarrows M^i_l, v^i_l$,
which is true by induction hypothesis.

QED

The proof follows easily from the lemma. Indeed, we just take $\mathcal{M} = \{(M, w)\}$ and we can then apply the lemma because for all $k \in \mathbb{N}^*$ and all multi-agent possible world (M', w'), if

$$m\{\sigma(R_{j_1} \circ \ldots \circ R_{j_k}(w'), R_{j_1} \circ \ldots \circ R_{j_k}(w)) \mid j_1, \ldots, j_k \in G, j_i \neq j_{i+1}, j_1 \neq Y\} = 1$$

then $(*)$ is fulfilled. QED

Corollary 4.4.12 *Let (M, w) be a multi-agent possible world, $\psi \in \mathcal{L}_{\neq Y}$ and $k = deg(\psi) + 1$. Then there is $(M', w') \in Mod(\psi)$ such that $s^k((M', w'), (M, w)) = s^k(Mod(\psi), (M, w))$.*

PROOF. It follows for Propositions 4.4.8 and 4.4.10. QED

In other words, this corollary tells us that $s^k(Mod(\psi),(M,w)) = Sup^k\{s^k((M',w'),(M,w)) \mid (M',w') \in Mod(\psi)\}$ is actually a maximum.

Finally, we have the following nice property.

Proposition 4.4.13 *The assignment defined in Definition 4.4.7 is a faithful assignment. Therefore the operator $\circ$ defined in Definition 4.4.7 satisfies the postulates $R1-R6$. Besides, $\circ$ satisfies also postulate $RG1$.*

PROOF.

- Clearly $\leq_\psi$ is a total pre-order because $\leq^k$ is a total pre-order. We are going to show that it is faithful.
  - If $(M,w),(M',w') \in Mod(\psi)$ then $s^k(Mod(\psi),(M,w)) = s^k(Mod(\psi),(M',w')) = (1,\ldots,1)$ by definition of s^k. So we cannot have $(M,w) <_\psi (M',w')$.
  - If $(M,w) \in Mod(\psi)$ and $(M',w') \notin Mod(\psi)$ then $s^k(Mod(\psi),(M,w)) = (1,\ldots,1)$ and $s^k(Mod(\psi),(M',w')) = (l_1,\ldots,l_k)$ with $l_1 < 1$.
    So $s^k(Mod(\psi),(M,w)) >^k s^k(Mod(\psi),(M',w'))$, i.e. $(M,w) <_\psi (M',w')$.
  - Finally, if $\vdash_{\mathsf{Int}} \psi \leftrightarrow \psi'$ then clearly $\leq_\psi = \leq_{\psi'}$.
- We are going to show that $\circ$ satisfies postulate $(RG1)$. Let $\varphi \in \mathcal{L}_{\neq Y}$ and $(M',w') \in Mod(\psi \circ \varphi)$. Assume that for all $(M,w) \in Mod(\psi)$, it is not the case that $M,w \leftrightarrows_{A'} M',w'$ with $A' = A_0 - A(\varphi)$.

  Let $(M,w) \in Mod(\psi)$ such that $s^k((M,w),(M',w')) = s^k(Mod(\psi),(M',w'))$. Such a (M,w) exists by Corollary 4.4.12.

  Assume that $pf \notin A'$, the case $pf \in A'$ is dealt with similarly. Then by definition of $\leftrightarrows_{A'}$,

  there is $j_0 \in A'$, $v \in R_{j_0}(w)$ such that for all $v' \in R_{j_0}(w')$ it is not the case that $M,w \leftrightarrows M',v'$ (1)

  or there is $j_0 \in A'$, $v' \in R_{j_0}(w')$ such that for all $v \in R_{j_0}(w)$ it is not the case that $M,v \leftrightarrows M',v'$ (2).

  Assume w.l.o.g. that (1) is the case. Then $\sigma(R_{j_0}(w),R_{j_0}(w')) < 1$.

  Using generated submodels, we can easily build a multi-agent possible world (M'',w'') such that $M'',w'' \leftrightarrows_{A_0-\{j_0\}} M',w'$ and $M'',w'' \leftrightarrows_{j_0} M,w$.
  - Then for all $j \neq j_0$, $\sigma(R_j(w''),R_j(w)) = \sigma(R_j(w'),R_j(w))$ and $\sigma(R_{j_0}(w''),R_{j_0}(w)) = 1 > \sigma(R_{j_0}(w',R_{j_0}(w))$.

– So for all $n \in \mathbb{N}^*$, all $j_1, \ldots, j_n$ such that $j_1 \neq j_0, j_i \neq j_{i+1}$, $\sigma(R_{j_0} \circ R_{j_1} \circ \ldots \circ R_{j_n}(w''), R_{j_0} \circ R_{j_1} \circ \ldots \circ R_{j_n}(w)) = 1 \geq \sigma(R_{j_0} \circ R_{j_1} \circ \ldots \circ R_{j_n}(w'), R_{j_0} \circ R_{j_1} \circ \ldots \circ R_{j_n}(w))$.
Besides, for all $n \in \mathbb{N}^*$, all $j_1, \ldots, j_n$ such that $j_1 \neq j_0, j_1 \neq Y, j_i \neq j_{i+1}$, $\sigma(R_{j_1} \circ \ldots \circ R_{j_n}(w''), R_{j_1} \circ \ldots \circ R_{j_n}(w)) = \sigma(R_{j_1} \circ \ldots \circ R_{j_n}(w'), R_{j_1} \circ \ldots \circ R_{j_n}(w))$ because $M'', w'' \leftrightarrows_{A_0 - \{j_0\}} M', w'$.

So for all $n \geq 2$,

$$m\{\sigma(R_{j_1} \circ \ldots \circ R_{j_n}(w''), R_{j_1} \circ \ldots \circ R_{j_n}(w)) \mid j_i \neq j_{i+1}, j_1 \neq Y\} \geq$$

$$m\{\sigma(R_{j_1} \circ \ldots \circ R_{j_n}(w'), R_{j_1} \circ \ldots \circ R_{j_n}(w)) \mid j_i \neq j_{i+1}, j_1 \neq Y\}$$

and

$$m\{\sigma(R_j(w''), R_j(w)) \mid j \in G, j \neq Y\} > m\{\sigma(R_j(w'), R_j(w)) \mid j \in G, j \neq Y\}.$$

So $s^k((M'', w''), (M, w)) >^k s^k((M', w'), (M, w))$.

Finally, because $\varphi \in \mathcal{L}_{A_0 - \{j_0\}}$, $M'', w'' \leftrightarrows_{A_0 - \{j_0\}} M', w'$ and $M', w' \models \varphi$, we have $M'', w'' \models \varphi$. So $(M'', w'') \in Mod(\varphi)$. Then $(M', w') \notin Mod(\psi \circ \varphi)$ which is impossible by assumption.

QED

4.4.4 Concrete example

The revision operations $\circ$ we introduced so far were syntactic. But in fact we could also define revision operations directly on internal models. Indeed, as we said internal models are formal representations that agent Y has 'in her mind'. So we need revision mechanisms that she could use to revise her formal representation when she receives an input under the form of an epistemic formula. Such revision operations would then take an internal model and an input formula as arguments and would yield another internal model. The following definition gives an example of such a revision operation.

Definition 4.4.14 (Revision operation $*$)
Let $\mathcal{M}$ be an internal model (of type 1) and $\varphi \in \mathcal{L}_{\neq Y}$ a satisfiable formula. We define the *revision of $\mathcal{M}$ by φ*, written $\mathcal{M} * \varphi$, as follows.

$$\mathcal{M} * \varphi = Min\left(Mod(\varphi), \leq_{\mathcal{M}}\right).$$

where for all multi-agent possible worlds (M, w) and (M', w'),

$$(M, w) \leq_{\mathcal{M}} (M', w') \text{ iff } s^k(\mathcal{M}, (M, w)) \geq^k s^k(\mathcal{M}, (M', w'))$$

where $k = deg(\varphi) + 1$. □

The reason why we stop at modal depth $k = deg(\varphi) + 1$ is the same reason why we stopped at modal depth $k = deg(\psi) + 1$ for $s^k(Mod(\psi), (M, w))$ in Definition 4.4.7. It is because we know thanks to Proposition 4.4.8 that there is a model of φ and a multi-agent possible world of $\mathcal{M}$ which agree on their modal structure beyond modal depth $deg(\varphi)$.

However, note that if $\mathcal{M}$ is an internal model then $\mathcal{M} * \varphi$ might be infinite and therefore not an internal model. The following proposition ensures us that it is not the case.

Proposition 4.4.15 *Let $\mathcal{M}$ be an internal model (of type 1) and $\varphi \in \mathcal{L}_{\neq Y}$ a satisfiable formula. Then $\mathcal{M} * \varphi$ is an internal model (of type 1).*

PROOF. Let $k = deg(\varphi) + 1$. By Proposition 4.4.8, we know that there is $(M', w') \in Mod(\varphi)$ and $(M, w) \in \mathcal{M}$ such that

$$m\{\sigma(R_{j_1} \circ \ldots \circ R_{j_k}(w'), R_{j_1} \circ \ldots \circ R_{j_k}(w)) \mid j_1, \ldots, j_k \in G, j_i \neq j_{i+1}, j_1 \neq Y\} = 1.(**)$$

So $\mathcal{M} * \varphi = \{(M', w') \in Mod(\varphi) \mid s^k((M', w'), \mathcal{M}) = s^k(Mod(\varphi), \mathcal{M})\} = \{(M', w') \in Mod(\varphi) \mid$ there is $(M, w) \in \mathcal{M}$ such that (M', w') satisfies (**) and $s^k((M', w'), (M, w)) = s^k(Mod(\varphi), \mathcal{M})\}$. By proposition 4.4.10, this last set is finite. So $\mathcal{M} * \varphi$ is finite, i.e. $\mathcal{M} * \varphi$ is an internal model (of type 1). QED

Example 4.4.16 ('Coin' example)
Let us take up the 'coin' example after the private announcement to Bob that the coin is heads up (p). Ann's internal model of type 1 $\{(M, w), (M', w')\}$ after this private announcement is depicted in Figure 4.10. The internal model of type 2 associated to $\{(M, w), (M', w')\}$ is depicted in Figure 4.11. These internal models are the same as before the private announcement to Bob because for her it is as if nothing happened. Now, suppose that the

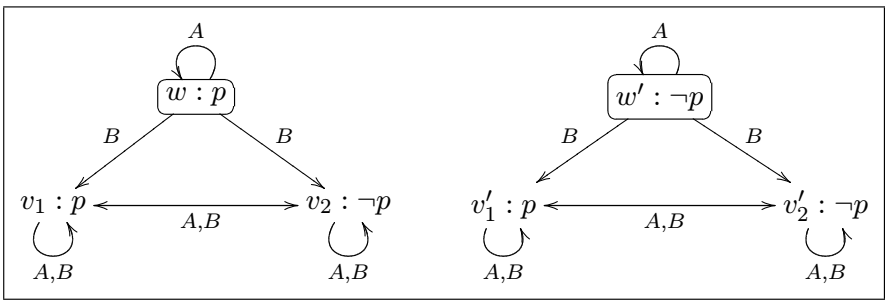


Figure 4.10: Ann's internal model of type 1 $\{(M, w), (M', w')\}$ after the private announcement to Bob that p is true

quizmaster announces her *privately* that Bob believes that the coin is heads up (formally $B_B p$). This announcement contradicts of course her beliefs and she has to revise her internal model. The following proposition tells us that the revised model is $\{(M^r, w^r), (M^{r'}, w^{r'})\}$, which is depicted in Figure 4.12.

Proposition 4.4.17 $\{(M, w), (M', w')\} * B_B p = \{(M^r, w^r), (M^{r'}, w^{r'})\}$

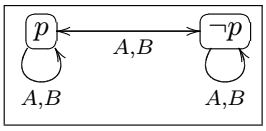


Figure 4.11: Internal model of type 2 associated to $\{(M, w), (M', w')\}$

PROOF. We first prove a series of lemmas.

Lemma 4.4.18 *Let (M'', w'') such that $M'', w'' \models B_B p$ and $s^2(\{(M, w), (M', w')\}, (M'', w'')) = s^2(\{(M, w), (M', w')\}, Mod(B_B p))$. Then $|M''| \geq 4$.*

PROOF. W.l.o.g. we assume that $M'', w'' \models p$. Let $v'' \in R_B \circ R_A(w'')$. We know by Proposition 4.4.8 that $s^2(\{(M, w), (M', w')\}, (M'', w'')) = (\alpha, \beta, 1)$. So there is $v \in R_B \circ R_A(w)$ such that $M'', v'' \leftrightarrows M, v$.

Then there are $v''_1, v''_2 \in R_B \circ R_A(w'')$ such that $M'', v''_1 \models p \wedge \neg B_B p \wedge \neg B_A p$ and $M'', v''_2 \models \neg p \wedge \neg B_B p \wedge \neg B_A p$. There is also $v''_3 \in R_B(w'')$ such that $M'', v''_3 \models B_B p \wedge \neg B_A p$. Finally, $M'', w'' \models B_A p \wedge B_B p$. So we have 4 worlds w'', v''_1, v''_2 and v''_3 satisfying different formulas. Therefore, there are at least 4 worlds in M''. QED

Lemma 4.4.19 *Let (M'', w'') such that $M'', w'' \models B_B p$. Then $s^2(\{(M, w), (M', w')\}, (M'', w'')) = \left(\frac{1}{3|M''|+1}, \frac{3}{4(3|M''|+1)}, \alpha\right)$ for some $\alpha \in [0; 1]$.*

Therefore $s^2(\{(M, w), (M', w')\}, Mod(B_B p)) \leq (\frac{1}{13}, \frac{3}{52}, 1)$.

PROOF. Assume w.l.o.g. that $M'', w'' \models p$. Then $max\{i \mid M'', w'' \leftrightarrows_i M, w\} = 1$.

Let $v'' \in R_B(w'')$. Then $max\{i \mid M'', v'' \leftrightarrows_i M, v \text{ and } v \in R_B(w)\} = 1$ because $M'', v'' \models B_B p$ and $M, v_i \not\models B_B p$ for $i = 1, 2$.

$max\{i \mid M'', v'' \leftrightarrows_i M, v_1 \text{ and } v'' \in R_B(w'')\} = 1$ because for all $v'' \in R_B(w'')$, $M'', v'' \models B_B p$ and $M, v_i \not\models B_B p$ for $i = 1, 2$.

$max\{i \mid M'', v'' \leftrightarrows_i M, v_2 \text{ and } v'' \in R_B(w'')\} = 0$ because for all $v'' \in R_B(w'')$, $M'', v'' \models p$ and $M, v_2 \models \neg p$.

So $\sigma(w, w') = \frac{1}{|M||M''|+1}$, and

$\sigma(R_B(w), R_B(w'')) = \frac{1}{2}\left(\frac{1}{2}\left(\frac{1}{|M||M''|+1} + \frac{0}{|M||M''|+1}\right) + \frac{1}{|R_B(w'')|}\sum_{v'' \in R_B(w'')} \frac{1}{|M||M'|+1}\right) =$

$\frac{1}{2}\left(\frac{1}{2}\frac{1}{|M||M''|+1} + \frac{1}{|M||M''|+1}\right) = \frac{3}{4(|M||M''|+1)}$.

Therefore $s^2(\{(M, w), (M', w')\}, (M'', w'')) = \left(\frac{1}{3|M''|+1}, \frac{3}{4(3|M''|+1)}, \alpha\right)$ for some $\alpha \in [0; 1]$.

We get that $s^2(\{(M, w), (M', w')\}, Mod(B_B p)) \leq \left(\frac{1}{13}, \frac{3}{52}, 1\right)$ thanks to Lemma 4.4.18. QED

Lemma 4.4.20

$s^2(\{(M, w), (M', w')\}, (M^r, w^r)) = \left(\frac{1}{13}, \frac{3}{52}, 1\right)$.

$s^2(\{(M, w), (M', w')\}, (M^{r'}, w^{r'})) = \left(\frac{1}{13}, \frac{3}{52}, 1\right)$.

Lemma 4.4.21 *Let (M'', w'') such that $M'', w'' \models B_B p$. Then,*
if $s^2((M, w), (M'', w'')) = \left(\frac{1}{13}, \frac{3}{52}, 1\right)$ then $M'', w'' \leftrightarrows M^r, w^r$;
if $s^2((M', w'), (M'', w'')) = \left(\frac{1}{13}, \frac{3}{52}, 1\right)$ then $M'', w'' \leftrightarrows M^{r'}, w^{r'}$.

PROOF. Assume $s^2\left((M, w), (M'', w'')\right) = \left(\frac{1}{13}, \frac{3}{52}, 1\right)$. Then $|M''| = 4$ by Lemma 4.4.19. Then one can easily show that $|R_B(w'')| = 1$ and $|R_B \circ R_A(w'')| = 2$. We set $R_B(w'') = \{v_1\}$ and $R_B \circ R_A(w'') = \{v_3'', v_4''\}$ with $M'', v_3'' \models p$ and $M'', v_4'' \models \neg p$.

Let $Z = \{(w^r, w''), (v_2^r, v_2''), (v_3^r, v_3''), (v_4^r, v_4'')\}$. One can easily show that Z is a bisimulation between (M^r, w^r) and (M'', w'').

The proof is similar if $s^2\left((M', w'), (M'', w'')\right) = \left(\frac{1}{13}, \frac{3}{52}, 1\right)$. QED

The proof of Proposition 4.4.17 then follows easily from Lemma 4.4.19, Lemma 4.4.20 and Lemma 4.4.21. QED

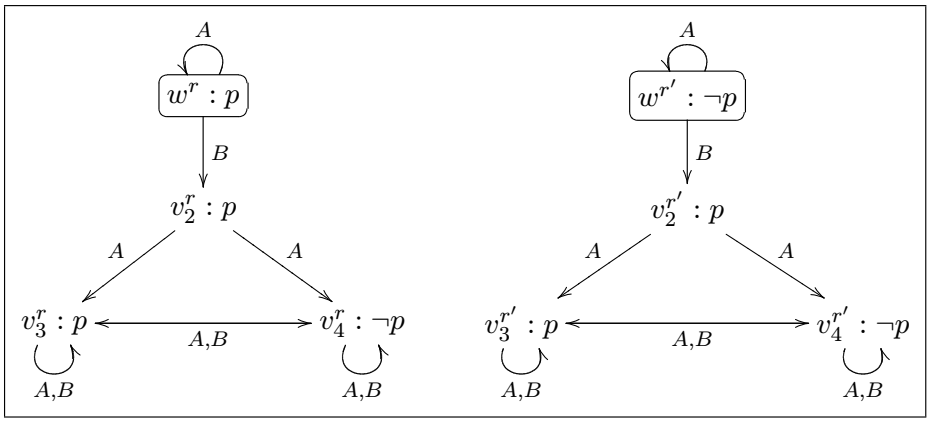


Figure 4.12: Revised internal model $\{(M^r, w^r), (M^{r'}, w^{r'})\}$ after the private announcement made to her that Bob believes the coin is heads up ($B_B p$)

The internal model of type 2 associated to $\{(M^r, w^r), (M^{r'}, w^{r'})\}$ is depicted in Figure 4.13. If we compare this internal model with the original internal model of Figure 4.11, we observe that Ann still does not know whether the coin is heads or tails up. This is what we should expect since the announcement was only about Bob's beliefs and did not give any information about the actual state of the coin (as it would have been the case if the private announcement was that Bob *knows* that the coin is heads up). Of course, Ann's beliefs about Bob's beliefs have changed because she now believes that Bob believes that the coin is heads up, unlike before. But (Ann's beliefs about) Bob's beliefs about Ann's beliefs have not changed. This is also what we should expect. Indeed, Bob is not aware of this private announcement to Ann, so his beliefs about Ann's beliefs do not change, and Ann knows that they do not change. And because these beliefs are independent from his beliefs about propositional facts like p, Ann's beliefs about Bob's beliefs about Ann's beliefs should not change during the revision

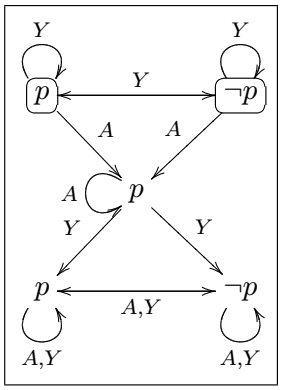


Figure 4.13: Internal model of type 2 associated to $\{(M^r, w^r), (M^{r'}, w^{r'})\}$

process. More generally, Ann's beliefs about beliefs of degree greater than 1, i.e. larger than the degree of $B_B p$, should not change. Formally, this is exactly what our anti-lexicographic ordering and Proposition 4.4.8 ensure. □

Remark 4.4.22 This example suggests that we could strengthen our postulate $RG1$ and require more demanding conditions. For example, if $\varphi = p \wedge (B_j B_i q \vee B_i p) \wedge B_i B_j B_i B_j p$, then the information this formula is about is not really $A(\varphi) = \{\mathtt{pf}, j, i\}$ as we defined it, but is more precisely made of the set of sequences of agents $S(\varphi) = \{\mathtt{pf}, (j,i), (i), (i,j,i,j)\}$. So, what should not change during a revision by φ are all beliefs φ' whose corresponding set of sequences $S(\varphi')$ does not intersect with $S(\varphi)$, which includes here all formulas of degree higher than 4 (because $deg(\varphi) = 4$). Formally, this corresponds to replacing $RG1$ by the following postulate.

RG1' Let $\psi, \varphi, \varphi' \in \mathcal{L}^C_{\neq Y}$ such that $S(\varphi) \cap S(\varphi') = \emptyset$.

If $\vdash_{\mathsf{Int}} \psi \to \varphi'$ then $\vdash_{\mathsf{Int}} \psi \circ \varphi \to \varphi'$.

□

4.5 Conclusion

We have shown in this chapter that generalizing belief revision theory to the multi-agent case amounts to studying private announcement. Indeed, when the incoming information is consistent with agent Y's beliefs, we have shown that expanding by φ in the AGM style corresponds to updating by the private announcement of φ in the BMS style. As we said, this result bridges the gap between AGM theory and BMS theory: AGM expansion can be seen as a particular kind of BMS update. On the other hand, when the incoming information is not consistent with agent Y's beliefs then agent Y has to revise her beliefs. To do so, we have generalized the techniques of AGM theory to the multi-agent case by replacing the notion of possible world by the notion of multi-agent possible world. Afterwards, we have

proposed rationality postulates for private multi-agent belief revision which are specific to our multi-agent setting. Finally, we have given a concrete example of private multi-agent belief revision based on the 'coin' example.

However, we have studied revision in the internal approach only for the case of private announcement. It still remains to study revision in the internal approach for any kind of events, and not only for private announcement. In general, revision has to take place for an internal model $(\mathcal{M}, W_a)$, given an internal event model $(\mathcal{A}, A_a)$, if the updated model is not serial (and thus not an internal model). Indeed, if the updated model is not serial then this means that for agent Y it is not common belief that the agents' beliefs are consistent (as we said in Section 3.5). The ultimate goal is thus to preserve seriality. Formally speaking, by Proposition 3.3.10, revision has to take place when for all $a \in A_a$ and all $w \in W_a$, $\mathcal{M}, w \not\models \delta^n(a)$ (where $n = |\mathcal{M}| \cdot |\mathcal{A}|$). One obvious solution would be to revise $(\mathcal{M}, W_a)$ by the formula $\bigvee_{a \in A_a} \delta^n(a)$ and then update by $(\mathcal{A}, A_a)$. The problem is that this formula does not necessarily belong to $\mathcal{L}^C_{\neq Y}$ and our revision mechanisms work only for formulas of $\mathcal{L}^C_{\neq Y}$. Besides, even if we managed to revise by $\bigvee_{a \in A_a} \delta^n(a)$, it is not necessarily the case that the resulting updated model would be serial. Indeed, the model revised by $\bigvee_{a \in A_a} \delta^n(a)$ could be of size larger than $|\mathcal{M}|$, and so Proposition 3.3.10 could not be applied anymore. So we do not know how to proceed for such general cases. Nevertheless, revision in the internal approach for any kind of internal event model is essential, not only for the internal approach but also for the external approach. Indeed, from the external model corresponding to an initial situation, one can get easily the internal model of each agent by Definition 2.3.16; and from the external event model corresponding to an event performed in this initial situation, one can easily get the internal event model for each agent by Definition 3.4.1. Now, in reality, when an event takes place, each agent updates and possibly revises her internal model on the basis of her internal event model. So given these internal models and internal event models, if we knew how to revise internal models, we should also be able to obtain in any case the updated and possibly revised internal model of each agent. Then the updated and possibly revised external model after the event would be obtained by applying Definition 2.3.19 to this set of updated and possibly revised internal models. So we see that knowing how to revise internal models by any kind of internal event model is essential if we want to revise external models as well. Steiner and Studer in [Steiner, 2006; Steiner and Studer, 2007] have proposed a system that preserves seriality in updated external models, but when the formula that is announced contradicts the beliefs of (some of) the agents, this formula is just ignored. By doing so, they simply avoid the issue of revising external models.

However, there are other ways than external models to represent epistemic situations from an external point of view. These formalisms often enrich modal logic with probability or plausibility. This enables to provide a more fine grained account not only of the epistemic state of agents but also of the process of belief change. In particular, they allow for belief revision. In the following chapter, we are going to propose such a general formalism for the external approach.

Chapter 5

External approach: a general formalism

5.1 Introduction

As we said in Section 2.3.1, the external approach has rather applications in cognitive psychology or game theory. In these fields, one needs to have a formalism which models as accurately as possible epistemic states of (human) agents and the dynamics of belief and knowledge. In that respect, the logical dynamics underlying the interpretation of an event by a human agent are complex and are rather neglected in the literature. To get a glimpse of them, let us have a look at two examples. Assume that you see somebody drawing a ball from an urn containing n balls which are either black or white. If you *believe* that it is equally probable that there are 0,1,..., or n black balls in the urn then you expect with equal probability that he draws a white ball or a black ball; but if you *believe* there are more black than white balls in the urn then you expect with a higher probability that he draws a black ball rather than a white ball. We see in this example that your beliefs about the situation contribute actively to interpret the event: they determine the probability with which you expect the white-ball drawing to happen. But this expectation, determined by your beliefs about the situation, can often be balanced consciously or unconsciously by what you actually obtain by the pure perception of the event happening. For example, assume that you listen to a message of one of your colleagues on your answering machine which says that he will come to your office on Tuesday afternoon, but you cannot distinguish precisely due to some noise whether he said Tuesday or Thursday. From your beliefs about his schedule on Tuesday and Thursday, you would have expected him to say that he would come on Tuesday because, say you believe he is busy on Thursday. But this expectation has to be balanced by what you actually perceive and distinguish from the message on the answering machine: you might consider more probable of having heard Tuesday than Thursday, which is independent of this expectation. Thus we see in these examples that in the process of interpreting an event, there is an interplay between two main informational components: your expectation of the event to happen and your pure perception (observation) of the event happening.

Up to now, this kind of phenomenon, although very common in everyday life is not dealt

with neither in dynamic epistemic logic [Baltag and Moss, 2004], [van Benthem, 2003] nor in the situation calculus [Bacchus *et al.*, 1999]. To model such phenomena, we have to resort to probability because the formalism for the external approach based on modal logic that we used in the preceding chapters is not expressive enough. Besides, in order to represent the agent's epistemic state accurately we also use hyperreal numbers. This enables us to model both degrees of belief and degrees of potential surprise, thanks to infinitesimals. The expressiveness of this formalism will then allow us to tackle the dynamics of belief appropriately, notably revision, and also to express in a suitable language what would surprise the agent (and how much) by contradicting her beliefs. Finally, for sake of generality, the events we consider in this chapter can also change the (propositional) facts of the situation.

This chapter is organized as follows. In Section 5.2 we will introduce some mathematical objects needed to represent the epistemic state of agents accurately. In Section 5.3 we will propose a general formalism for the external approach but we will deal only with the notion of belief. This formalism will be built in the BMS style. Then in Section 5.4 we will add knowledge to this formalism and in Section 5.5 we will generalize our formalism to the multi-agent setting. In Section 5.6, we will compare our formalism to existing ones and finally in Section 5.7 we will conclude.

5.2 Mathematical preliminaries

In our formalism, the probabilities of worlds and formulas will take values in a particular mathematical structure $(\mathbb{V}, \lessapprox)$ (abusively written $(\mathbb{V}, \leq)$) different from the real numbers, based on hyperreal numbers $({}^*\mathbb{R}, \leq)$. (The approach in [Adams, 1975] uses them as well to give a probabilistic semantics to conditional logic.) In this section, we will briefly recall the main features of hyperreal numbers that will be useful in the sequel (for details see [Keisler, 1986]). Afterwards we will motivate and introduce our particular structure $(\mathbb{V}, \lessapprox)$.

Roughly speaking, hyperreal numbers are an extension of the real numbers to include certain classes of infinite and infinitesimal numbers. A hyperreal number x is said to be infinitesimal iff $|x| < 1/n$ for all integers n, finite iff $|x| < n$ for some integer n, and infinite iff $|x| > n$ for all integers n. Infinitesimal numbers are typically written ε, finite numbers are written x and infinite numbers are written ∞. Note that an infinitesimal number is a finite number as well, that $\frac{1}{\varepsilon}$ is an infinite number and that $\frac{1}{\infty}$ is an infinitesimal number. Two hyperreal numbers x and y are said to be infinitely close to each other if their difference $x - y$ is infinitesimal. If x is finite, the *standard part* of x, denoted by $St(x)$, is the unique real number which is infinitely close to x. So for example $St(1 + \varepsilon) = 1$, $St(\varepsilon) = 0$.

Hyperreal numbers will be used to assign probabilities to facts (formulas). A fact is considered consciously probable by the agent when its probability is real. A fact would surprise the agent if she learnt that it was true when its current probability is infinitesimal. We want to refine the ordering given by the infinitesimals and introduce a global ranking among these potentially surprising facts. Indeed, a fact of probability ε^2 is infinitely more surprising than a fact of probability ε and hence, for the agent, the importance of the former should be negligible compared to the importance of the latter. This can be done algebraically by approximating our expressions. More precisely, in case a hyperreal number x is infinitely

smaller than y, i.e. there is an infinitesimal ε such that $x = \varepsilon \cdot y$, then we want $y + x = y$. For example we want $1 + \varepsilon = 1$ (here $x = \varepsilon$ and $y = 1$), $\varepsilon + \varepsilon^2 = \varepsilon$ (here $x = \varepsilon^2$ and $y = \varepsilon$),... In other words, in case x is *negligible* compared to y, then $y + x = y$. The hyperreal numbers do not allow us to do that, so we are obliged to devise a new structure $(\mathbb{V}, \lessapprox)$.

First we introduce some definitions. By *semi-field* (resp. *ordered semi-field*), we mean a field (resp. ordered field) which lacks the property of 'existence of additive inverse'. For example, $(\mathbb{R}^+, +, \cdot, \leq)$ and $({}^*\mathbb{R}^+, +, \cdot, \leq)$ are ordered semi-fields, where ${}^*\mathbb{R}^+$ denotes the positive hyperreal numbers. Now we define $\mathbb{V}$, which will be the quotient structure of the set of positive hyperreal numbers ${}^*\mathbb{R}^+$ by a particular equivalence relation.

Definition 5.2.1 (Equivalence relation $\approx$)
Let $x, y \in {}^*\mathbb{R}^+$, we set

$$x \approx y \text{ iff } \begin{cases} St(\frac{x}{y}) = 1 & \text{if } y \neq 0 \\ x = 0 & \text{if } y = 0. \end{cases}$$

We can easily check that $\approx$ is an equivalence relation on ${}^*\mathbb{R}^+$. □

For instance, we have $1 + \varepsilon \approx 1$ and $\varepsilon + \varepsilon^2 \approx \varepsilon$.

Theorem 5.2.2 *The quotient structure $\mathbb{V} = ({}^*\mathbb{R}^+/_{\approx}, \overline{+}, \overline{\cdot})$ is a semi-field. (Elements of $\mathbb{V}$, being equivalence classes of ${}^*\mathbb{R}^+$, are classically denoted $\overline{x}$. And $\overline{\cdot}, \overline{+}$ denote the quotient relations of $\cdot$ and $+$.)*

PROOF. We only need to prove that $\overline{+}$ and $\overline{\cdot}$ are well defined, since the rest is standard and straightforward.

Assume $\overline{x} = \overline{x'}$ and $\overline{y} = \overline{y'}$. We have to show $\overline{x}\overline{+}\overline{y} = \overline{x'}\overline{+}\overline{y'}$ and $\overline{x}\overline{\cdot}\overline{y} = \overline{x'}\overline{\cdot}\overline{y'}$.

First, let us show that $\overline{x}\overline{+}\overline{y} = \overline{x'}\overline{+}\overline{y'}$.

Assume $\overline{x} = 0$ (similar proof for $\overline{y} = 0$). Then $x = x' = 0$. In that case $\overline{x}\overline{+}\overline{y} = \overline{x + y} = \overline{y} = \overline{y'} = \overline{0 + y'} = \overline{x' + y'} = \overline{x'}\overline{+}\overline{y'}$.

Assume $\overline{x} \neq 0$ and $\overline{y} \neq 0$. Then x, x', y, y' are all different from 0. So $x + y \neq 0$ because $x, y \geq 0$.

$\overline{x}\overline{+}\overline{y} = \overline{x'}\overline{+}\overline{y'}$
iff $\overline{x + y} = \overline{x' + y'}$
iff $St(\frac{x'+y'}{x+y}) = 1$ because $x + y \neq 0$ (see above)
iff $St(\frac{x}{x+y}) + St(\frac{y'}{x+y}) = 1$
iff $St(\frac{x'}{x} \cdot \frac{1}{1+\frac{y}{x}}) + St(\frac{y'}{y} \cdot \frac{1}{1+\frac{x}{y}}) = 1$
iff $St(\frac{x'}{x}) \cdot St(\frac{1}{1+\frac{y}{x}}) + St(\frac{y'}{y}) \cdot St(\frac{1}{1+\frac{x}{y}}) = 1$
iff $St(\frac{1}{1+\frac{y}{x}}) + St(\frac{1}{1+\frac{x}{y}}) = 1$ because $St(\frac{x'}{x}) = St(\frac{y'}{y}) = 1$
iff $\frac{1}{1+St(\frac{y}{x})} + \frac{1}{1+St(\frac{x}{y})} = 1$

iff $\frac{1}{1+St(\frac{y}{x})} + \frac{1}{1+\frac{1}{St(\frac{y}{x})}} = 1$ which is true.

Now, let us show that $\overline{x \cdot y} = \overline{x'} \cdot \overline{y'}$.

Assume $\overline{x} = 0$ (similar proof for $\overline{y} = 0$). Then $x = x' = 0$ and the equality is fulfilled.

Assume $\overline{x} \neq 0$ and $\overline{y} \neq 0$. Then x, x', y, y' are different from 0. So $x \cdot y \neq 0$.

$\overline{x \cdot y} = \overline{x'} \cdot \overline{y'}$
iff $\overline{x \cdot y} = \overline{x' \cdot y'}$
iff $St(\frac{x' \cdot y'}{x \cdot y}) = 1$ because $x \cdot y \neq 0$
iff $St(\frac{x'}{x}) \cdot St(\frac{y'}{y}) = 1$
iff $1 \cdot 1 = 1$ which is true. QED

Now we need to define the ordering relation $\lessapprox$ on $\mathbb{V}$.

Definition 5.2.3 We define a relation $\lessapprox$ on $\mathbb{V}$ by

$$\overline{x} \lessapprox \overline{y} \text{ iff there are } x \in \overline{x},\ y \in \overline{y} \text{ such that } x \leq y$$

□

$\mathbb{V}$ equipped with $\lessapprox$ turns out to be an ordered semi-field thanks to the following lemma:

Lemma 5.2.4 *If $\overline{x} \lessapprox \overline{y}$ then for all $x' \in \overline{x}$ and all $y' \in \overline{y}$, $x' \leq y'$*

PROOF.

1. Assume $\overline{x} \neq 0$ and $\overline{y} \neq 0$ (then x, y are different from 0).
   (a) If $\overline{x} = \overline{y}$ then we have the result.
   (b) If $\overline{x} \neq \overline{y}$ then by definition there are $x_0 \in \overline{x}$ and $y_0 \in \overline{y}$ such that $x_0 < y_0$.
   Assume there are $x' \in \overline{x}, y' \in \overline{y}$ such that $x' > y'$.

   - Assume that either $x_0 \leq y' \leq x' \leq y_0$ or $x_0 \leq y' \leq y_0 \leq x'$ or $x_0 \leq y_0 \leq y' \leq x'$. Then $x_0 \leq y' \leq x'$, so $\frac{x_0}{x_0} \leq \frac{y'}{x_0} \leq \frac{x'}{x_0}$ because $x_0 \neq 0$. Then $1 \leq St(\frac{y'}{x_0}) \leq St(\frac{x'}{x_0}) = 1$ because $x', x_0 \in \overline{x}$. Then $St(\frac{y'}{x_0}) = 1$, $y' \in \overline{x_0} = \overline{x}$, $\overline{x} = \overline{y'} = \overline{y}$ which is impossible by assumption.

   - Assume that either $y' \leq x_0 \leq x' \leq y_0$ or $y' \leq x' \leq x_0 \leq y_0$ or $y' \leq x_0 \leq y_0 \leq x'$. Then $y' \leq x_0 \leq y_0$, so $\frac{y'}{y_0} \leq \frac{x_0}{y_0} \leq 1$ because $y_0 \neq 0$. Then $1 = St(\frac{y'}{y_0}) \leq St(\frac{x_0}{y_0}) \leq 1$. Then $St(\frac{x_0}{y_0}) = 1$. So $\overline{x} = \overline{x_0} = \overline{y_0} = \overline{y}$ which is impossible by assumption.

   So in all possible cases we reach a contradiction. This means that for all $x' \in \overline{x}$, all $y' \in \overline{y}$ $x' \leq y'$

2. (a) If $\overline{x} = 0$ and $\overline{y} \neq 0$, then $0 \lessapprox \overline{y}$. So there is $y_0 \in \overline{y}$ such that $0 < y_0$.

   Assume there is $y' \in \overline{y}$ such that $0 > y'$. Then
   $y' < 0 \leq y_0$
   $\frac{y'}{y_0} < 0 \leq \frac{y_0}{y_0} = 1$
   $1 = St(\frac{y'}{y_0}) \leq 0 \leq 1$ because $y', y_0 \in \overline{y}$
   i.e. $0 = 1$, which is counterintuitive.

   So for all $y' \in \overline{y}$, $0 \leq y'$.

   (b) if $\overline{y} = 0$ and $\overline{x} \neq \overline{0}$ then
   $\overline{x} \leq 0$, so there is $x_0 \in \overline{x}$ such that $x_0 < 0$.

   Assume there is $x' \in \overline{x}$ such that $x' > 0$.
   $x_0 \leq 0 < x'$
   $1 = \frac{x_0}{x_0} \leq 0 < \frac{x'}{x_0}$
   $1 \leq 0 \leq St(\frac{x'}{x_0}) = 1$ because $x', x_0 \in \overline{x}$
   i.e. $0 = 1$ which is again counterintuitive.

   So for all $x' \in \overline{x}$, $x' \leq 0$.

QED

Theorem 5.2.5 *The structure $(\mathbb{V}, \lessapprox)$ is an ordered semi-field.*

PROOF. First we prove a lemma.

Lemma 5.2.6 *$\lessapprox$ is a total order on ${}^*\mathbb{R}^+$ such that*

1. *if $\overline{x} \lessapprox \overline{y}$ then $\overline{x}\overline{+}\overline{z} \lessapprox \overline{y}\overline{+}\overline{z}$,*
2. *if $0 \lessapprox \overline{x}$ and $0 \lessapprox \overline{y}$ then $0 \lessapprox \overline{x}\overline{\cdot}\overline{y}$.*

PROOF. Follows easily from Lemma 5.2.4 and the fact that $\leq$ is a total order on ${}^*\mathbb{R}^+$ satisfying also conditions 1 and 2 above. QED

The proof then follows easily from the lemma above and Theorem 5.2.2. QED

In the sequel, we will denote abusively $\overline{x}$, $\overline{+}$, $\overline{\cdot}$, $\lessapprox$ by $x, +, \cdot, \leq$. The elements $\overline{x}$ containing an infinitesimal will be called abusively *infinitesimals* and denoted $\varepsilon, \delta, \ldots$. Those containing a real number will be called abusively *reals* and denoted $a, b, \ldots$ Those containing an infinite number will be called *infinites* and denoted $\infty, \infty', \ldots$ Moreover, when we refer to intervals, these intervals will be in $\mathbb{V}$; so for example $]0; 1]$ refers to $\{x \in \mathbb{V} \mid 0 \lessapprox x \lessapprox 1 \text{ and not } x \approx 0\}$.

We can then easily check that now we do have at our disposal the following identities: $1 + \varepsilon = 1; 0.6 + \varepsilon = 0.6; \varepsilon + \varepsilon^2 = \varepsilon; \ldots$

5.3 Starting with beliefs as probabilities

As in the BMS system, we divide our task into three parts. Firstly, we propose a formalism called *p*roba-*d*oxastic (pd) model to represent how the actual world is perceived by the agent from a static point of view (Section 5.3.1). Secondly, we propose a formalism called generic event model to represent how an event occurring in this world is perceived by the agent (Section 5.3.2). Thirdly, we propose an update mechanism which takes as arguments a pd-model and a generic event model, and yields a new pd-model; the latter is the agent's representation of the world after the event represented by the above generic event model took place in the world represented by the above pd-model (Section 5.3.3). In this section, our account will be presented for a single agent to highlight the main new ideas and we will focus only on the notion of belief.

5.3.1 The static part

The Notion of pd-model

Definition 5.3.1 (Proba-doxastic model)
A *proba-doxastic model (pd-model)* is a tuple $M = (W, P, V, w_a)$ where:

1. W is a finite set of possible worlds;
2. w_a is the possible world corresponding to the actual world;
3. $P : W \to]0; 1]$ is a probability measure such that
$$\sum\{P(w) \mid w \in W\} = 1;$$
4. $V : \Phi \to 2^W$ assigns a set of possible worlds to each propositional letter.

□

Intuitive interpretations.

The possible worlds W are determined by the modeler (who is somebody different from the agent and who has perfect knowledge of the situation). One of them, w_a, corresponds to the actual world. Among these worlds W there are some worlds that the agent conceives as potential candidates for the world in which she dwells, and some that she would be surprised to learn that they actually correspond to the world in which she dwells (disregarding whether this is true or false). The first ones are called *conceived worlds* and the second *surprising worlds*. The conceived worlds are assigned by P a real value and the surprising worlds are assigned an infinitesimal value, both different from 0. For example, some people would be surprised if they learnt that some swans are black, although it is true. To model this situation, we introduce two worlds: one where all swans are white (world w) and one where some swans are not white (world v). So for these people the actual world v is a surprising world, whereas the world w is a conceived world.

Of course for the agent some (conceived) worlds are better candidates than others, and this is expressed by the probability value of the world: the larger the real probability value of the (conceived) world is, the more likely it is for the agent. But that is the same for the surprising worlds: the agent might be more surprised to learn about some worlds than others. For example, if you play poker with somebody you trust, you will never suspect that he cheats. However he does so, and so carefully that you do not suspect anything. When at the end of the game he announces to you that he has cheated, you will be surprised: it is true in the actual world, but this world was a surprising world for you. But you will be even more surprised if he tells you that he has cheated five times before. So the world where he has cheated five times will be more surprising than the world where he has cheated once, and these are both surprising worlds for you. Infinitesimals enable us to express this: the larger the infinitesimal probability value of the (surprising) world is, the less the agent would be surprised by this world. Anyway, that is why we need to introduce hyperreal numbers: to express these degrees of potential surprise that cannot be expressed by a single number like 0, which then becomes useless for us.

The agent does not think consciously that the surprising worlds are possible (unlike conceived worlds), she is just not aware of them. In other words, from an internal point of view, the agent's representation of the surrounding world is composed only of the conceived worlds. So these surprising worlds are useless to represent her beliefs which we assume are essentially conscious. But still, they are relevant for the modeling from an external point of view. Indeed they provide some information about the epistemic state of the agent: namely what would surprise her and how firmly she holds to her belief. Intuitively, something that you do not consider consciously as possible and that contradicts your beliefs is often surprising for you if you learn that it is true. These worlds will moreover turn out to be very useful technically in case the agent has to revise her beliefs (see Section 5.3.3).

In Figure 5.1 an example of a pd-model (without any valuation) is depicted. The dots correspond to worlds and the numbers next to them represent their respective probabilities. The conceived worlds are the ones in the inner circle, the other worlds are surprising worlds. Note that both the sum of the probabilities of the conceived worlds and the global sum of all these worlds (conceived and surprising) are equal to 1: $\sum\{P(v) \mid v \in W\} = \sum\{P(v) \mid v \text{ is real}\} = 1$ (see Section 5.2).

Examples

In this chapter we will follow step by step two examples: the 'Urn' example and the 'Answering Machine' example of the introduction.

Example 5.3.2 ('Urn' example)
Suppose the agent is in a fair, and there is an urn containing $n = 2.k > 0$ balls which are either white or black. The agent does not know how many black balls there are in the urn but believes that it is equally probable that there are 0,1,..., or n black balls in the urn. Now say there is actually no black ball in the urn. This situation is depicted in Figure 5.2. The worlds are within squares and the double bordered world is the actual world. The numbers within squares stand for the probabilities of the worlds and the propositional letter p_i stands

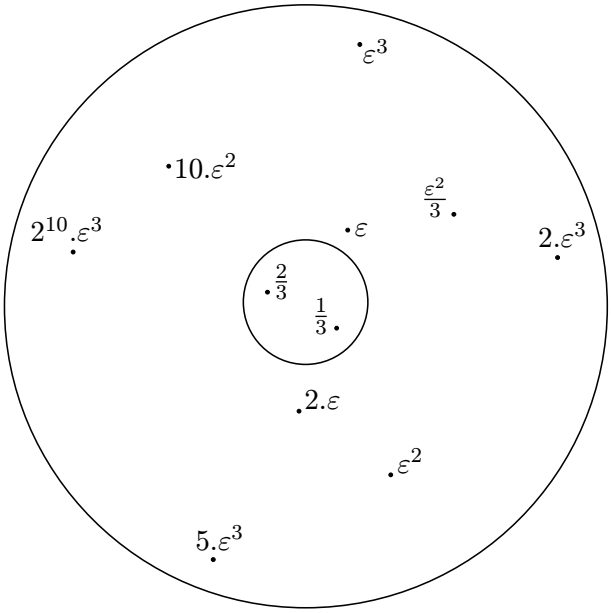


Figure 5.1: Example of pd-model

for: "there are i black balls in the urn". □

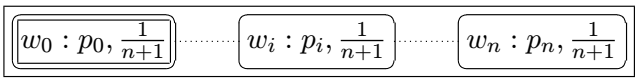


Figure 5.2: 'Urn' example

Example 5.3.3 ('Answering machine' example)
Assume a professor (the agent) comes back after lunch to her office. She ate with a colleague who just told her at lunch about his new timetable for this year. However she does not remember quite precisely what he said, in particular she is a bit uncertain whether his 1.5 hour lecture at 2.00 pm is on Tuesday or on Thursday, and she believes with probability $\frac{4}{5}$ (resp. $\frac{1}{5}$) that his lecture is on Tuesday (resp. Thursday). In fact it is on Tuesday at 2.00 pm. We represent this situation in the model of Figure 5.3. As in Figure 5.2, the double bordered world is the actual world. The numbers within squares stand for the probabilities of the worlds and p (resp. $\neg p$) stands for "her colleague has his 1.5 hour lecture on Tuesday (resp. Thursday)"[1]. □

[1]Note that strictly speaking we would need two propositional variables p and p', where p (resp. p') would stand for "her colleague has his lecture on Tuesday (resp. Thursday)" because the fact that he does not have his lecture on Tuesday ($\neg p$) should not imply logically that he has his lecture on Thursday (p').

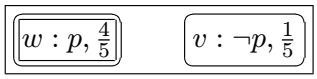


Figure 5.3: 'Answering Machine' Example

Static language

We can define naturally a language $\mathcal{L}_{St}$ for pd-models (St standing for Static).

Definition 5.3.4 (Language $\mathcal{L}_{St}$)
The syntax of the language $\mathcal{L}_{St}$ is defined by

$$\mathcal{L}_{St} : \varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid P(\varphi) \geq x \mid B\varphi$$

where $x \in [0; 1[$, and p ranges over Φ.

Its semantics is inductively defined by

$$\begin{array}{lll} M, w \models \top & & \\ M, w \models p & \text{iff} & w \in V(p) \\ M, w \models \neg\varphi & \text{iff} & \text{not } M, w \models \varphi \\ M, w \models \varphi \wedge \psi & \text{iff} & M, w \models \varphi \text{ and } M, w \models \psi \\ M, w \models P(\varphi) \geq x & \text{iff} & \sum\{P(v) \mid M, v \models \varphi\} \geq x \\ M, w \models B\varphi & \text{iff} & \sum\{P(v) \mid M, v \models \varphi\} = 1 \end{array}$$

□

$M, w \models P(\varphi) \geq x$ should be read "φ has probability greater than x for the agent". (Note that the world w does not really play a role here because we could pick any world of the pd-model M.) $M, w \models B\varphi$ should be read "the agent believes φ". However we have to be careful about what notion of belief we refer to in this definition. Indeed the term "belief" refers in natural language to different concepts (that we distinguish here through $\mathcal{L}_{St}$). Assume that you conjecture an arithmetical theorem φ from a series of examples and particular cases. The more examples you have checked, the more you will "believe" in this theorem. This notion of belief corresponds to the type of formula $P(\varphi) \geq a$ for a real and smaller than 1; the bigger a is the more you "believe" in φ. But if you come up with a proof of this theorem that you have checked several times, you will still "believe" in this theorem but this time with a different strength. Your belief will be a conviction and corresponds here to the formula $B\varphi$. However, note that this conviction (belief) might still be false if there is a mistake in the proof that you did not notice (like what happened to Wilson for his first proof of Fermat's theorem). This notion of belief actually corresponds to the notion of belief we used in the preceding chapters. Moreover, if we define the operator B_{w} by $M, w \models B_{\mathsf{w}}\varphi$ iff $M, w \models P(\varphi) > 0.5$, then B_{w} corresponds to Lenzen's notion of "weak belief" and B and B_{w} satisfy Lenzen's axioms defined in [Lenzen, 1978]. This operator B satisfies the axioms K, D, 4, 5 but not the axiom T. So it does not correspond to the notion of knowledge. For a more in

depth account on the different significations of the term "belief", see for example [Lenzen, 1978].

Moreover, we can also express in this language what would surprise the agent, and how much so. Indeed, in case x is an infinitesimal ε, $P(\varphi) = \varepsilon$ should be read "the agent would be surprised with degree ε if she learnt that φ ".[2] ($P(\varphi) = \varepsilon$ is defined by $P(\varphi) \geq \varepsilon \wedge P(\neg\varphi) \geq 1 - \varepsilon$.) Note that the smaller x is, the higher the intensity of surprise is. But this use of infinitesimals could also express how firmly we believe something, in Spohn's spirit. Indeed, $P(\neg\varphi) = \varepsilon > \varepsilon' = P(\neg\varphi')$ would then mean that φ' is believed more firmly than φ.

In the truth condition for $P(\varphi) \geq x$, note that if x is real then only the conceived worlds have to be considered in the sum $\sum\{P(v) \mid M, v \models \varphi\}$ because the sum of any real (different from 0) with an infinitesimal is equal to this real (see Section 5.2). Likewise, the semantics of B amounts to say that φ is true at least in all the *conceived* worlds of W. So it is quite possible to have a surprising actual world where $\neg\varphi$ is true and still the agent believing φ (i.e. $B\varphi$): just take φ="All swans are white" in the above example.

5.3.2 The dynamic part

Definition 5.3.5 (Generic event model)
A *generic event model* is a structure $A = (E, S, P, \{P^\Gamma \mid \Gamma \text{ is a maximal consistent subset of S }\}, \{Pre_a \mid a \in E\}, a_a)$ where

1. E is a finite set of possible events;
2. a_a is the actual event;
3. S is a set of formulas of $\mathcal{L}_{St}$ closed under negation;
4. $P^\Gamma : E \to [0; 1]$ is a probabilistic measure indexed by a maximal consistent subset Γ of S, and assigning to each possible event a a *real* number in $[0; 1]$ such that
$$\sum\{P^\Gamma(a) \mid a \in E\} = 1;$$
5. $P : E \to]0; 1]$ is a probabilistic measure such that
$$\sum\{P(a) \mid a \in E\} = 1;$$
6. $Pre_a : \Phi \to \mathcal{L}_{St}$ is a function indexed by each possible event a.

□

[2] Note that this does not necessarily model *all* the things that would surprise the agent since there might still be things that the agent might conceive as possible with a low real probability and still be surprised to hear them claimed by somebody else (Gerbrandy, private communication). For a more in depth formal account on the notion of surprise, see [Lorini and Castelfranchi, 2007]

Intuitive interpretation.

Items 1 and 2 are similar to Definition 5.3.1. It remains to give an interpretation to items 3-6 of the definition.

Items 3 and 4. S corresponds to the set of facts about the world that are relevant to determine the probabilities P^Γ of Item 4. The maximal consistent sets Γ then cover all the relevant eventualities needed for the modeling. The choice of the formulas of S is left to the modeler but they should be as elementary and essential as possible in order to give rise to all the relevant eventualities. In that respect, one should avoid infinite sets S.

$P^\Gamma(a)$ is the probability that the agent *would expect (or would have expected)* a to happen (among the possible events of E), if the agent assumed that she was in a world where the formulas of Γ are true. In other words, $P^\Gamma(a)$ can be viewed as the probability for the agent that the event a would occur in a world where the formulas of Γ are true. Note that this is a conditional probability of the form $P(a|\Gamma)$. Moreover, because we assume the agent to be rational, the determination of the value of this probability can often be done objectively and coincides with the agent's subjective determination (see examples).

This probability value is real and cannot be infinitesimal (unlike P), and (still unlike P) we can have $P^\Gamma(a) = 0$. This last case intuitively means that the event a cannot *physically* be performed in a world where Γ is true. These operators P^Γ generalize the binary notion of precondition in [Baltag and Moss, 2004], [Aucher, 2004] and [van Benthem, 2003].

Remark 5.3.6 In our definition of generic event model, instead of referring directly to worlds w of a particular pd-model, we refer to maximal consistent subsets Γ of a set S. This is for several reasons. Firstly, the determination of the probability $P^w(a)$ does not depend on all the information provided by the world w but often on just a part of it expressed by Γ. Secondly, for philosophical reasons, the *same* event might be performed in different situations, modelled by different pd-models, and be perceived differently by the agent in each of these situations, depending on what her epistemic state is (see Example 18). The use of maximal consistent sets enables us to capture this *generic* aspect of events: in our very definition we do not refer to a particular pd-model. Finally, for computational reasons, we do not want that in practice the definition of generic event models depends on a particular pd-model because we want to be able to iterate the event without having to specify at each step of the iteration the new event model related to the new pd-model. The use of maximal consistent sets allows us to do so. However, from now on and for better readability we write $\boldsymbol{P^w(a) = P^\Gamma(a)}$ **for the unique** $\boldsymbol{\Gamma}$ **such that** $\boldsymbol{M, w \models \Gamma}$ (such a Γ exists because S contains the negation of each formula it contains). □

Item 5. $P(a)$ is the probability for the agent that a actually occurs among E, determined solely by the agent's *perception* and *observation* of the event happening. This probability is independent of the agent's beliefs of the static situation which could alter and modify this determination consciously or unconsciously (see the 'answering machine' example). This probability is thus independent of the probability that the agent would have expected a to happen, because to determine this last probability the agent takes into account what she believes about the static situation (this fact will be relevant in Section 5.3.3).

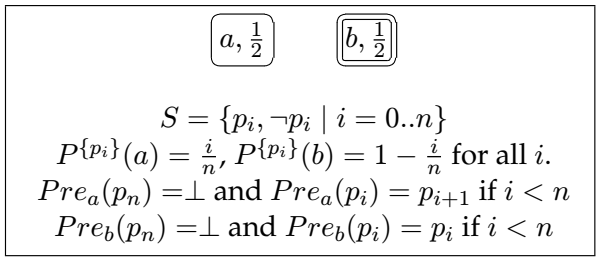


Figure 5.4: Someone else draws a (white) ball and puts it in his pocket and the agent is uncertain whether he draws a white (b) or a black ball (a).

Just as in the static case, we have *conceived* events (which are assigned a real number) and *surprising* events (which are assigned an infinitesimal). The former are events that the agent conceives as possible candidates while one of the events of E actually takes place. The latter are events that the agent would be surprised to learn that they actually took place while one of the other events took place. To take up the poker example of Section 5.3.1, if you play poker with somebody you trust and at a certain point he cheats while you do not suspect anything, then the actual event of cheating will be a surprising event for you (of value ε) whereas the event where nothing particular happens is a conceived event (of value 1). Just as in the static case, the relative strength of the events (conceived and surprising) is expressed by the value of the operator P. The probabilities values are different from 0 for the same reasons that the probability values in a pd-model are different from 0 (see Section 5.3.1).

Item 6. The function Pre_a deals with the problem of determining what facts will be true in a world after the event a takes place. Intuitively, $Pre_a(p)$ represents the necessary and sufficient Precondition in any world w for p to be true after the performance of a in this world w.

Example 5.3.7 ('Urn' example)
Consider the event whereby someone else draws a ball from the urn (which is actually a white ball) and puts it in his pocket, the agent sees him doing that but she cannot see the ball. This event is depicted in Figure 5.4. Action a (resp. b) stands for "someone else draws a black (resp. white) ball and puts it in his pocket". The numbers within the squares stand for the probabilities $P(a)$ of the possible events. The maximal consistent sets are represented by their 'positive' components, so $\{p_i\}$ refers to the set $\{p_i, \neg p_k \mid k \neq i\}$.

The observation and perception of the event in itself does not provide the agent any reason to have a preference between him drawing a black ball or a white ball; so we set $P(a) = P(b) = \frac{1}{2}$.[3] However if the agent assumed she was in a world where there are i black balls then the probability that she would (have) expect(ed) him drawing a black (resp. white)

[3]We could nevertheless imagine some exotic situation where the agent's observation of him drawing the ball would give her some information on the color of the ball he is drawing. For example, if black balls were much heavier than white balls and the agent sees him having difficulty drawing a ball, she could consider more probable that he is drawing a black ball.

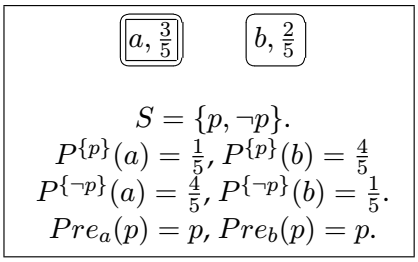


Figure 5.5: The agent is uncertain whether her colleague says that he will come on Tuesday (a) or on Thursday (b) between 2.00 pm and 4.00 pm, but she considers more probable having heard Tuesday than Thursday.

ball would be $\frac{i}{n}$ (resp. $1 - \frac{i}{n}$); so we set $P^{\{p_i\}}(a) = \frac{i}{n}$ and $P^{\{p_i\}}(b) = 1 - \frac{i}{n}$. Moreover there cannot be n black balls in the urn after he put one ball in his pocket; so we set $Pre_a(p_n) = \bot$ and $Pre_b(p_n) = \bot$. But if he draws a black ball then there is one black ball less; so we set $Pre_a(p_i) = p_{i+1}$ for all $i < n$. Otherwise if he draws a white ball the number of black balls remains the same; so we set $Pre_b(p_i) = p_i$ for all $i < n$. □

Example 5.3.8 ('Answering machine' example)
Assume now that when the professor enters her office, she finds a message on her answering machine from her colleague. He tells her that he will bring her back a book he had borrowed next Tuesday in the beginning of the afternoon between 2.00 pm and 4.00 pm. However, there is some noise on the message and she cannot distinguish precisely whether he said Tuesday or Thursday. Nevertheless she considers more probable of having heard Tuesday rather than Thursday. We model this event in Figure 5.5. a stands for "her colleague says that he will come on Tuesday between 2.00 pm and 4.00 pm" and b stands for "her colleague says that he will come on Thursday between 2.00 pm and 4.00 pm". $P(a)$ and $P(b)$ represent the probabilities the agent assigns to a and b on the sole basis of what she has heard and distinguished from the answering machine (and are depicted within the squares). These probabilities are determined on the basis of her sole perception of the message. On the other hand, $P^{\{p\}}(a)$ is the probability that she would have expected her colleague to say that he will come on Tuesday (rather than Thursday) if she assumed that his lecture was on Tuesday. This probability can be determined objectively. Indeed, because we assume that her colleague has a lecture on Tuesday from 2.00 pm to 3.30 pm, the only time he could come on Tuesday would be between 3.30 pm and 4.00 pm (only 0.5 hour). So we set $P^{\{p\}}(a) = \frac{0.5hr}{2.5hrs} = \frac{1}{5}$. Similarly, $P^{\{p\}}(b) = \frac{2hrs}{2.5hrs} = \frac{4}{5}$. Finally, the message does not change the fact of her colleague having a lecture or not on Tuesday; so we set $Pre_a(p) = p$ and $Pre_b(p) = p$. □

5.3.3 The update mechanism

Definition 5.3.9 (Update product)

Given a pd-model $M = (W, P, V, w_a)$ and a generic event model $A = (E, S, P,$

$\{P^\Gamma \mid \Gamma \text{ is a m.c. subset of } S\}, \{Pre_a \mid a \in E\}, a_a)$, we define their *update product* to be the pd-model $M \otimes A = (W', P', V', w'_a)$, where:

1. $W' = \{(w, a) \in W \times E \mid P^w(a) > 0\}$;
2. We set

$$P'(a) = \frac{P(a) \cdot P^W(a)}{\sum\{P(b) \cdot P^W(b) \mid b \in E\}} \text{ where } P^W(a) = \sum\{P(v) \cdot P^v(a) \mid v \in W\};$$

then

$$P'(w, a) = \frac{P(w)}{\sum\{P(v) \mid v \in W \text{ and } P^v(a) > 0\}} \cdot P'(a);$$

3. $V'(p) = \{(w, a) \in W' \mid M, w \models Pre_a(p)\}$;
4. $w'_a = (w_a, a_a)$.

□

Intuitive interpretation and motivations.

Items 1 and 4 As in BMS, in the new model we consider all the possible worlds (w, a) resulting from the occurrence of the possible event a in the possible world w, granted that this event a can physically take place in w (i.e. $P^w(a) > 0$). The new actual world is the result of the performance of the actual event a_a in the actual world w_a.[4]

Item 2. We want to determine $P'(w, a) = P(W \cap A)$, where W stands for the event 'we were in world w before a occurred' and A for 'event a just occurred'. More formally, in the probability space W', W stands for $\{(w, b) \mid b \in E\}$ and A for $\{(v, a) \mid v \in W\}$ and we can check that $W \cap A = \{(w, a)\}$. But of course to determine these probabilities we have to rely only on M and A.
Probability theory tells us that

$$P(W \cap A) = P(W|A) \cdot P(A).$$

We first determine $P(W|A)$, i.e. the probability that the agent was in world w *given* the extra assumption that event a occurred in this world. We claim it is reasonable to assume

$$P(W|A) = \frac{P(w)}{\sum\{P(v) \mid v \in W \text{ and } P^v(a) > 0\}}.$$

That is to say, we *conditionalize* the probability of w for the agent (i.e. $P(w)$) to the worlds where the event a took place and that may correspond for the agent to the actual world w

[4]In this chapter we follow the external approach. So our pd-models and generic event models are correct and the actual world and actual event of our models do correspond to the actual world and actual event in reality. It is then natural to assume that the actual event a_a can physically be performed in the actual world w_a: $P^{w_a}(a_a) > 0$. Hence, the existence of the (actual) world (w_a, a_a) is justified.

(i.e. $\{v \mid v \in W \text{ and } P^v(a) > 0\}$). That is the way it would be done in classical probability theory. The intuition behind it is that we now possess the extra piece of information that a occurred in w. So the worlds where the event a did *not* occur do not play a role anymore for the determination of the probability of w. We can then get rid of them and conditionalize on the remaining relevant worlds.

It remains to determine $P(A)$, which we also denote $P'(a)$; that is to say the probability for the agent that a occurred. We claim that

$$P(A) = P_1 \cdot P_2$$

where P_1 is the probability for the agent that a actually occurred, determined on the sole basis of her *perception* and *observation* of the event happening; and P_2 is the probability that the agent would have expected a to happen determined on the sole basis of her epistemic state (i.e. her beliefs). Because P_1 and P_2 are independent, we simply multiply them to get $P'(a)$.

By the very definition of $P(a)$ (see Definition 5.3.5), $P_1 = P(a)$.

As for P_2, the agent's epistemic state is represented by the worlds W. So she could expect a to happen in any of these worlds, each time with probability $P^v(a)$. We might be tempted to take the average of them: $P_2 = \frac{\sum\{P^v(a) | v \in W\}}{n}$, where n is the number of worlds in W. But we have more information than that on the agent's epistemic state. The agent does not know in which world of W she is, but she has a preference among them, which is expressed by P. So we can refine our expression above and take the *center of mass* (or barycenter) of the $P^v(a)$s balanced respectively by the weights $P(v)$s (whose sum equals 1), instead of taking roughly the average (which is actually also a center of mass but with weights $\frac{1}{n}$). We get $P_2 = P^W(a) = \sum\{P(v) \cdot P^v(a) \mid v \in W\}$. (Note that this expression could also be viewed as an application of a theorem of conditional probability if we rewrote $P^v(a)$ to $P(a|v)$.)

Finally, we normalize $P'(a)$ on the set of events E to get a probabilistic space. We get

$$P(A) = P'(a) = \frac{P(a) \cdot P^W(a)}{\sum\{P(b) \cdot P^W(b) \mid b \in E\}} \text{ where } P^W(a) = \sum\{P(v) \cdot P^v(a) \mid v \in W\}.$$

We can easily check that $\sum\{P'(w,a) \mid (w,a) \in W'\} = 1$, which ensures that P' is a probability measure on $2^{W'}$.

Item 3. Intuitively, this formula says that a fact p is true after the performance of a in w iff the necessary and sufficient precondition for p to be true after a was satisfied in w before the event occurred.[5]

Example 5.3.10 ('Urn' example)
Assume now that someone else just drew a (white) ball from the urn and put it in his pocket, event depicted in Figure 5.4. Then because the agent considered equally probable that there was 0,1,..., or n black balls in the urn, she would expect that the other person drew a black ball or a white ball with equal probability. That is indeed the case:

[5]This solution of determining which propositional facts are true after an update was first proposed by Renardel De Lavalette [Renardel De Lavalette, 2004] and Kooi [Kooi, 2007], [van Ditmarsch *et al.*, 2005].

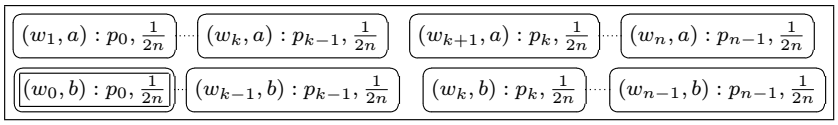


Figure 5.6: Situation of the urn example after somebody else drew a white ball and put it in his pocket (a).

$$P^W(a) = \sum\{P(v) \cdot P^v(a) \mid v \in W\} = \sum\{\tfrac{1}{n+1} \cdot \tfrac{i}{n} \mid i = 0..n\} = \tfrac{1}{2} = P^W(b).$$

Independently from that, her perception of the event did not provide her any reason to prefer a over b (i.e. $P(a) = P(b)$). So, in the end she should believe equally that the other agent drew a black ball or a white ball, and this is indeed the case: $P'(a) = P'(b)$.

If we perform the full update mechanism, then we get the pd-model of Figure 5.6. In this model all the worlds are equally probable for the agent. Note that there cannot be n black balls in the urn (p_n) any longer since one of them has been withdrawn.

Now consider another scenario where this time the agent initially believes that there are more black balls than white balls (for example she believed somebody else that told her so in the beginning). This can be modelled by assigning the initial probabilities $P(w_i) = \varepsilon$ for $i = 0, .., k$ and $P(w_i) = \frac{1}{k}$ for $i = k+1, .., n$ to the worlds of the model depicted in Figure 5.2 (recall that there are $n = 2.k$ balls). Then, if somebody else draws a ball from the urn and we compute again the probabilities of the events a and b, we get what we expect, namely that the agent considers more probable that a black ball has been withdrawn rather than a white ball:

$$P'(a) = P^W(a) = \sum\{\varepsilon \cdot \tfrac{i}{n} \mid i = 0..n\} + \sum\{\tfrac{1}{k} \cdot \tfrac{i}{n} \mid i = k+1..n\} = \tfrac{3}{4} + \tfrac{1}{4k} > \tfrac{1}{4} - \tfrac{1}{4k} = P^W(b) = P'(b).$$

□

Example 5.3.11 ('Answering machine' example)
Now that the professor has heard the message, she updates her representation of the world with this new information. We are not going to perform the full update and display the new pd-model, but rather just concentrate on how she computes her event probabilities $P'(a)$ and $P'(b)$.

After this computation, the probability $P'(a)$ that her colleague said that he will come on Tuesday is a combination of: (1) how much she would have expected him to say so, based on what she knew and believed of the situation, and (2) what she actually distinguished and heard from the answering machine.

The first value (1) is $P^W(a) = \sum\{P(v) \cdot P^v(a) \mid v \in W\}$, and the second (2) is $P(a)$. We get $P'(a) = \frac{12}{29} < \frac{17}{29} = P'(b)$. The important thing to note here is that a has become less probable than b for her: $P'(a) < P'(b)$ while before the update $P(a) > P(b)$. On the one hand, because the probability that she heard her colleague saying that he would come on Tuesday

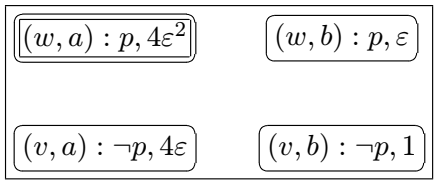


Figure 5.7: Situation after she believed that her colleague's lecture is on Thursday ($\neg p$) and she heard him saying that he would come on Thursday (b).

has decreased due to her lower expectation of him to say so: $P^W(a) = \frac{8}{25} < \frac{3}{5} = P(a)$; expectation which is based on her belief that he is busy on Tuesday because he has got a lecture ($P(p) = \frac{4}{5}$). On the other hand, this is due to the fact that the probability that she heard her colleague saying that he would come on Thursday has increased due to her higher expectation of him to say so: $P(b) = \frac{2}{5} < \frac{17}{25} = P^W(b)$).

Now consider a second scenario where this time she initially believes with a higher probability that he has got a lecture on Thursday than on Tuesday ($P(w) = \frac{1}{5}$ and $P(v) = \frac{4}{5}$ in Figure 5.3). This time her belief that she heard him saying that he will come on Tuesday (resp. Thursday) is strengthened (resp. weakened) by her independent expectation of him to say so: $P'(a) > P(a) > P(b) > P'(b)$. □

Example 5.3.12 ('Answering machine' example 2)
In this variant of the answering machine example, we are going to show the usefulness of infinitesimals and show an example of belief revision.

Basically, we consider the same initial situation, except that the agent's beliefs are different. This time she is convinced that her colleague's lecture is on Thursday and she is also convinced that she heard him saying that he would come on Thursday. Formally, everything remains the same except that now in Figure 5.3 $P(w) = \varepsilon$ and $P(v) = 1$, and in Figure 5.5 $P(a) = \varepsilon$ and $P(b) = 1$. If we apply the full update mechanism with these new parameters we get the model depicted in Figure 5.7. In this model she is convinced that he has got a lecture on Thursday and that he said he would come on Thursday (world (v, b)). Moreover, she would be surprised (with degree 5ε) if she learnt that her colleague has got his lecture on Tuesday *or* he said he will come on Tuesday (worlds (w, b), (v, a) and (w, a); remember that $5\varepsilon + 4\varepsilon^2 = 5\varepsilon$). But she would be much more surprised (with degree ε^2) if she learnt that he has got his lecture on Tuesday *and* he said he will come on Tuesday (world (w, a)), because that contradicts twice her original convictions.

Now later her colleague tells her that he has got his lecture on Tuesday, then she will have to revise her beliefs. The event model of this announcement is depicted in Figure 5.8. The resulting model is depicted in Figure 5.9. In this model, she now believes that he has got his lecture on Tuesday, but she is still convinced that her colleague said that he will come on Thursday because no new information has contradicted this. This is made possible thanks to the (global) ranking of surprising worlds by infinitesimals. What happened during this revision process is that the least surprising world where p is true became the only conceived

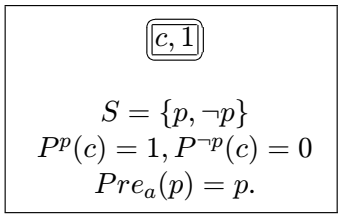


Figure 5.8: Her colleague announces to her that his lecture is on Tuesday.

Figure 5.9: Situation after her colleague announced his lecture is on Tuesday (c) and she then revised her beliefs.

world. □

5.4 Adding knowledge

5.4.1 State of the art

The definition of knowledge and its relationship with the notion of belief is an old issue of epistemology dating back at least to Plato. In the Theaetetus, Plato came to the conclusion that knowledge is true belief plus something else, namely *logos*. Ayer is more specific than Plato and claims that "the necessary and sufficient conditions for knowing that something is the case are first that one is said to know be true, secondly that one be sure of it, and thirdly that one should have the right to be sure" [Ayer, 1956]. So, we could say more succinctly that until then knowledge was conceived as 'justified true belief'. Seven years later, Gettier [Gettier, 1963] provided two counterexamples to such an analysis of knowledge. One of these two examples is the following. Suppose that Smith has strong evidence that 'Jones owns a Ford' (1) (for instance, Jones has always owned a Ford since Smith knows him). Then, because of (1) and by propositional logic, Smith is also justified in believing that 'Jones owns a Ford *or* his friend Brown is in Barcelona' (2), even if Smith has no clue about where Brown is at the moment. However it turns out that Jones does not own a Ford and that by pure coincidence Brown is actually in Barcelona. Then, (a) (2) is true, (b) Smith believes (2), and (c) Smith is justified in believing (2). So Smith has a true and justified belief in (2). However, intuitively, one could not say that Smith knows (2). This counterexample sparked a lot of discussion and many other definitions of knowledge were proposed, analyzed and refined to avoid the now-called 'Gettier problem' [Lycan, 2006]. However no consensus came out of this epistemological industry.

At about the same time Hintikka inaugurated contemporary epistemic logic with his

seminal book *Knowledge and Belief. An Introduction to the Logic of the Two Notions* [Hintikka, 1962]. His aim was not to propose another definition of knowledge like in epistemology but rather to propose a formal approach to the topic in order to study the logic of the notions of knowledge and belief. So he was more concerned with the investigation of reasonable principles that these notions could fulfill. In that respect, he claims in his book that the logic of knowledge is S4, which is obtained by adding to the normal modal logic K the axioms T and 4: if the agent knows something then this thing is true (formally $K\varphi \rightarrow \varphi$, axiom T) and if the agent knows something then she knows that she knows it (formally $K\varphi \rightarrow KK\varphi$, axiom 4). We believe these are reasonable assumptions to adopt for the notion of knowledge and refer to [Lenzen, 1978] or [Hintikka, 1962] for justifications.

Then comes the problem of adding belief to the picture. An attempt was proposed by Kraus and Lehman in [Kraus and Lehmann, 1988]. For this purpose, they introduce in the semantics two accessibility relations D and R, D modeling the notion of belief and R modeling the notion of knowledge. They then propose to add to the logic S4 of knowledge the following axioms to capture the interactions between these two notions.

D	$B\varphi \rightarrow \neg B\neg\varphi$	(Consistency)
5	$\neg K\varphi \rightarrow K\neg K\varphi$	(Negative introspection)
KB	$K\varphi \rightarrow B\varphi$	(Bridge axiom)
KB'	$B\varphi \rightarrow KB\varphi$	(Bridge axiom')

However, Voorbraak observed that these axioms entail the following theorem: $BK\varphi \rightarrow K\varphi$.[6] This theorem says that one cannot believe to know a false proposition, which is of course counterintuitive. Besides, we can also prove the following one: $B\varphi \rightarrow \varphi$, which is even more counterintuitive.

Van der Hoek in [van der Hoek, 1993] has proposed a systematic approach to this problem. He showed thanks to correspondence theory that any bimodal axiomatic system that includes the axioms D, KB, 5 and the axiom SB below entails the theorem $B\varphi \rightarrow K\varphi$. This theorem is of course counterintuitive because it produces a collapse of the distinction between knowledge and belief. However, he also showed that for each proper subset of $\{\mathsf{D}, \mathsf{KB}, 5, \mathsf{SB}\}$, counter-models can be built which show that none of those sets of axioms entail the collapse of the distinction between knowledge and belief.

SB	$B\varphi \rightarrow BK\varphi$	(Strong belief)

[6] Here is the proof:

1	$K\varphi \rightarrow B\varphi$	Axiom KB
2	$K\neg K\varphi \rightarrow B\neg K\varphi$	$KB : \neg K\varphi/\varphi$
3	$B\varphi \rightarrow \neg B\neg\varphi$	Axiom D
4	$B\neg\varphi \rightarrow \neg B\varphi$	3, contraposition
5	$B\neg K\varphi \rightarrow \neg BK\varphi$	$4 : K\varphi/\varphi$
6	$\neg K\varphi \rightarrow K\neg K\varphi$	Axiom 5
7	$\neg K\varphi \rightarrow B\neg K\varphi$	6,2, Modus Ponens
8	$\neg K\varphi \rightarrow \neg BK\varphi$	7,5, Modus Ponens
9	$BK\varphi \rightarrow K\varphi$	8, contraposition.

So we have to drop at least one principle in $\{\mathsf{D}, \mathsf{KB}, \mathsf{5}, \mathsf{SB}\}$. Voorbraak proposed to drop the axiom KB. His notion of knowledge is therefore unusual in so far as it does not require the agent to be aware of its belief state. We do not believe this is a reasonable principle, at least for modeling human agents. Axioms SB and D are also intuitively correct: SB expresses that our notion of belief corresponds to a notion of conviction or of 'being sure of', while axiom D expresses that our beliefs are consistent. On the other hand, most philosophers (including Hintikka) have quite rightly attacked the axiom 5. Let us take up the 'answering machine' example and assume as in Example 5.3.12 that the agent believes (is sure that) her colleague's lecture is on Thursday (formally $B\neg p$). She is actually wrong because it is actually on Tuesday. Therefore, she does not know that her colleague's lecture is on Tuesday ($\neg Kp$). If we assume that axiom 5 is valid then we should conclude that she knows that she does not know that her colleague's lecture is on Tuesday ($K\neg Kp$) (and therefore she also believes that she does not know it ($B\neg Kp$)). This is of course counterintuitive. More generally, axiom 5 is invalidated when the agent has wrong beliefs. Despite such blatant counterexamples, axiom 5 is still considered in computer science and economics as a valid principle for the notion of knowledge. It is however true that most examples and particular models studied in these fields do validate this axiom (such as Halpern et al.'s interpreted systems [Fagin *et al.*, 1995]), but that should not be a reason to accept this principle *in general*. We should instead propose a general system for knowledge and belief which boils down to accept axiom 5 when we restrict our attention to the examples and cases studied in these fields. Such systems have been proposed for example by Lenzen [Lenzen, 1978; Lenzen, 1979] and Stalnaker [Stalnaker, 2006].

Lenzen and Stalnaker proposed to add to the logic S4 of knowledge the following axioms to capture the interactions between the notions of belief and knowledge.

PI	$B\varphi \to KB\varphi$	(Positive introspection)
NI	$\neg B\varphi \to K\neg B\varphi$	(Negative introspection)
KB	$K\varphi \to B\varphi$	(Bridge axiom)
D	$B\varphi \to \neg B\neg\varphi$	(Consistency)
SB	$B\varphi \to BK\varphi$	(Strong belief)

These axioms are clearly intuitively correct. One can even show that the belief operator B satisfies the KD45 logic. In fact, the belief operator B can be defined in terms of the knowledge operator K because one can prove the following theorem: $B\varphi \leftrightarrow \neg K\neg K\varphi$. Thanks to this result, one can also show that the knowledge operator satisfies the logic S4.2 which is obtained by adding to the logic S4 the axiom .2 (see Section 2.2.1). This logic is also the logic of the notion of justified knowledge studied by Voorbraak in [Voorbraak, 1993].

Thanks to this connection between the belief and the knowledge operator, one can define an accessibility relation D for the notion of belief on the basis of the accessibility relation R for the notion of knowledge: for any worlds w and v, we set wDv iff for all u, wRu implies uRv. Then one can show that this accessibility relation is indeed serial, euclidean and transitive. Finally, note that if we are in a world w such that wDw then the accessibility relation for knowledge R is euclidean at w. So in this system, when the agent does not have wrong beliefs, axiom 5 is indeed valid (and the notions of knowledge and belief collapse).

But we could go the other way around and try to define from the accessibility relation D for belief an accessibility relation R for knowledge which satisfies the above principles PI, NI, KB, D and SB. We know by the bridge axiom that if wDv then wRv. We also know by the strong belief axiom that if wDw then (wRv iff wDv). So we still have to specify the worlds accessible by R for worlds w such that it is not the case that wDw. In [Stalnaker, 2006], Stalnaker introduces four such possible extensions of the belief accessibility relations D to the knowledge accessibility relations R. The first extension consists in the reflexive closure of the accessibility relation D. This is the minimal extension possible and it yields the objectionable definition of knowledge as true belief. This yields the logic S4.4.[7] The second extension consists in defining wRv as ((wDw and wDv) or (not wDw)). This is the maximal extension possible and it yields the logic S4.3.2 (which has been used in non-monotonic logic).[8] The third extension consists in defining knowledge as true belief which cannot be defeated by any true fact. In other words, a fact is known if and only if it is true and it will still be believed after any possible truthful announcement. [9] This last extension yields the logic S4.3.[10] This last condition was also proposed to be added to the classical notion of knowledge as justified true belief by Lehrer and Paxson [Lehrer and Paxson, 1969] in order to cope with the 'Gettier Problem'. In fact, this definition of knowledge as undefeated true belief can be formalized thanks to the arbitrary announcement modality $\Box$ introduced in [Balbiani *et al.*, 2007]. Intuitively, $\Box\varphi$ means that φ is true after *any* truthful announcement of an epistemic formula. So in any logic whose revision mechanism satisfies AGM postulate K*2 of Section 4.2.2 one can show that this definition of knowledge can be formalized as $K\varphi = \Box B\varphi$ (whether the announced formulas are propositional or epistemic does not make any difference here). However, Stalnaker argues further that this definition of knowledge as undefeated true belief should not be a sufficient and necessary condition for knowledge but only a sufficient one. This contention gives the last possible extension to the accessibility relation for knowledge.

5.4.2 Our proposal

We are not going to propose a new approach to the logics of knowledge and belief since we agree with Lenzen's and Stalnaker's approach. Instead, in order to exemplify the richness and expressivity of our formalism, we are going to show how the different notions of knowledge spelled out in the last paragraph can be captured in our framework. This includes in particular the notion of knowledge conceived as undefeated justified true belief. So for that,

[7] That is S4 plus .4: $(\varphi \wedge \hat{K}K\psi) \rightarrow K(\varphi \vee \psi)$ (True belief); see Section 2.2.

[8] That is S4 plus .3.2: $(\hat{K}\varphi \wedge \hat{K}K\psi) \rightarrow K(\hat{K}\varphi \vee \psi)$; see Section 2.2.

[9] For this definition to be consistent, we have to add another constraints that Stalnaker does not mention: in this definition, knowledge should only deal with propositional facts belonging to the propositional language $\mathcal{L}_0$. Indeed, assume that the agent believes non-p (formally $B\neg p$). Then clearly the agent knows that she believes non-p by PI (formally $KB\neg p$). However, assume that p is actually true. If we apply this definition of knowledge then if she learnt that p (which is true), she should still believe that she believes non-p (formally $BB\neg p$), so she should still believe non-p (formally $B\neg p$), which is of course counterintuitive. This restriction to propositional knowledge does not produce a loss of generality because we assumed that the agent knows everything about her own beliefs and disbeliefs.

[10] That is S4 plus .3: $\hat{K}\varphi \wedge \hat{K}\psi \rightarrow \hat{K}(\varphi \wedge \hat{K}\psi) \vee \hat{K}(\varphi \wedge \psi) \vee \hat{K}(\psi \wedge \hat{K}\varphi)$ (Weakly connected); see Section 2.2.

we first need to study a bit more how belief revision is performed in our framework. That is what we are going to do in the next section.

Structure of pd-models and belief revision

We first study a bit more the structure of pd-model because it will be relevant to better understand how belief revision is performed in our formalism.

In Section 5.2, we said that a fact of probability ε^2 should be negligible compared to a fact of probability ε because it is infinitely more surprising. We even refined the structure of hyperreal numbers in order to account for such a phenomenon. We can specify it a bit further by defining an equivalence relation $\approx^0$ which states in which case one number is not negligible compared to another.

Definition 5.4.1 (Equivalence relation $\approx^0$)
Let $\mathbb{V}' = \{x \in \mathbb{V} \mid x \text{ is real or } x \text{ is infinitesimal}\}$. Let $x, y \in \mathbb{V}'$, we set

$$x \approx^0 y \text{ iff } \begin{cases} \frac{x}{y} \text{ is real different from } 0 & \text{if } y \neq 0 \\ x = 0 & \text{if } y = 0. \end{cases}$$

We can easily check that $\approx^0$ is an equivalence relation on $\mathbb{V}'$. □

So for example, ε is negligible compared to $\frac{1}{3}$ but not compared to $3.\varepsilon$, i.e. $\varepsilon \approx^0 3.\varepsilon$.

Definition 5.4.2 (Ranking and global degree of a world)
We define the *ranking* $\mathbb{V}^0 = (V^0, +^0, .^0, 0^0, 1^0, \leq^0)$ as the quotient structure of $\mathbb{V}'$ by the equivalence relation $\approx^0$.

Given a pd-model, the *global degree of a world* w in this pd-model is defined as $\alpha((P(w))^0) \in P(w)^0$, where α is a given choice function. □

Example 5.4.3 In Figure 5.10 is depicted the pd-model of Figure 5.1 to which we have added other circles which specify the global degrees of its worlds. So all the worlds in between two circles have the same global degree. Besides, if the world w is in a circle closer to the inner circle than v, then $w^0 \geq^0 v^0$, i.e. the world w has a global degree larger than v. But among the worlds of the same global degree exists also a *local* ranking corresponding to the usual order relation. So for example, in the figure we have the local ranking $\frac{\varepsilon^2}{3} < \varepsilon^2$, even if the worlds with these probabilities have the same global degree. This refinement of the global ranking will play a role during the revision process. □

One can then show that our structure $\mathbb{V}'$ is isomorphic to a cumulative algebra, which is a notion introduced by Weydert [Weydert, 1994].

Theorem 5.4.4 *$\mathbb{V}'$ is isomorphic to a cumulative algebra.*

PROOF. We assume the validity of the axiom of choice and so the existence of a function $\alpha : \mathbb{V}^0 \to \mathbb{V}'$ which assigns to each element x^0 of $\mathbb{V}^0$ (which is a subset of $\mathbb{V}'$) an element of $\mathbb{V}'$ such that $\alpha(x^0) \in x^0$.

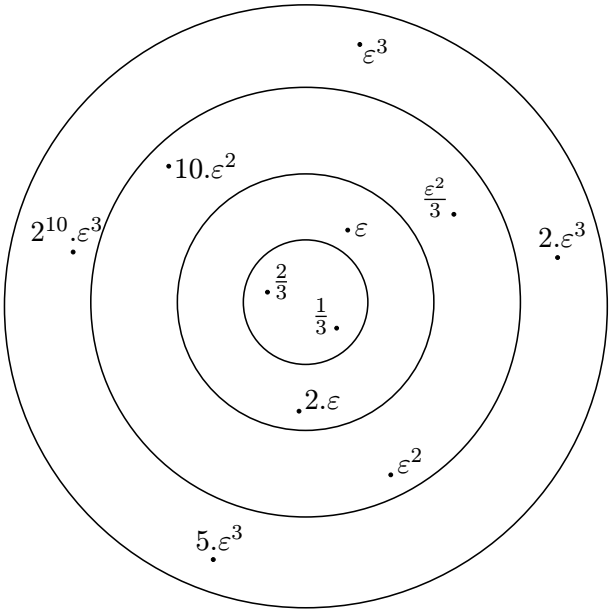


Figure 5.10: Example of pd-model with (global) ranking

Now we can define an isomorphism $f : \mathbb{V}' \to \mathbb{H}(\mathbb{V}^0, \mathbb{R}^+)$ between $\mathbb{V}'$ and the cumulative algebra with global structure $\mathbb{V}^0$ and local structure $\mathbb{R}^+$ as follows: $f(x) = (x^0, \frac{x}{\alpha(x^0)})$. Its inverse isomorphism $g : \mathbb{H}(\mathbb{V}^0, \mathbb{R}^+) \to \mathbb{V}'$ is defined by $g(x^0, y) = \alpha(x^0).y$. QED

So Weydert's comparisons with Spohn's theory [Spohn, 1988] and possibility theories [Dubois and Prade, 1991] transfer. In particular the ranking $\mathbb{V}^0$ determines a *global* ranking of worlds similar in spirit to Spohn's degrees of disbelief or possibility degrees in possibility theory or to the faithful ordering induced by a formula in Definition 4.2.11. More precisely, our conceived worlds correspond to Spohn's worlds of plausibility 0, and our surprising worlds correspond to Spohn's worlds of plausibility strictly larger than 0. Moreover our global degrees correspond to Spohn's plausibility degrees (although the order has to be reversed). The same correspondence applies for possibility theory and for the notion of faithful assignment.

There are actually several other proposals in the literature which cumulate features of ranking theories (Spohn-type or possibility theories) and probability: for example generalized qualitative probability [Lehmann, 1996], lexicographic probability [Blume *et al.*, 1991], big-stepped probabilities [Dubois and Fargier, 2004]. All these proposals are very similar in spirit and seem to be equivalent in one way or another.

Remark 5.4.5 In the literature (including myself in [Aucher, 2004]), one often considers degrees of possibility/plausibility (present in possibility theory and Spohn's theory) and probabilities as two different means to represent and tackle the same kind of information. However, as it is stressed in this chapter, for us they are meant to model two related but different

kinds of information. In our sense, the first rather corresponds to *degrees of potential surprise* about facts absent from the agent's mind. The second rather corresponds to *degrees of belief or acceptance* about facts present (or accessible) in the agent's mind (which can be a knowledge base for example). The same distinction is also present in [Gärdenfors, 1988]. □

Our structure $\mathbb{V}$ is richer than the cumulative algebra $\mathbb{V}'$ because it allows its elements to have multiplicative inverses. This feature turns out to be quite useful in a dynamic setting because it allows conditionalization and in particular belief revision, that we are now going to study.

The event model corresponding to the truthful announcement of φ is depicted in Figure 5.10. Given a pd-model (M, w_a), its revised model by φ is then defined by $M \otimes A(\varphi), (w_a, a)$.

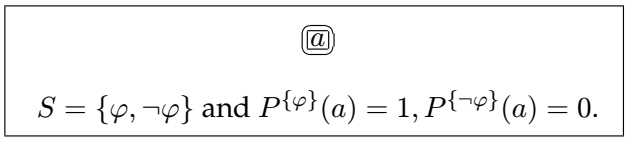


Figure 5.11: Event model $(A(\varphi), a)$ corresponding to the truthful announcement of φ to the agent.

Proposition 5.4.6 *Let (M, w_a) be a pd-model and $(A(\varphi), a)$ the event model depicted in Figure 5.11. Then the conceived worlds of $M \otimes A(\varphi)$ consist of the restriction of M to $Max(\{w \in M \mid M, w \models \varphi\}, \leq^0)$, where for a given set of possible worlds W', $Max(W', \leq^0) = \{w \in W' \mid$ for all $v \in W'$, $v^0 \leq^0 w^0\}$.*

PROOF. $(v, a) \in M \otimes A(\varphi)$ is a conceived world
iff $P(v, a)$ is real and $M, v \models \varphi$
iff $\frac{P(v)}{\sum\{P(w) \mid M, w \models \varphi\}}$ is real and $M, v \models \varphi$
iff $P(v) \approx^0 \sum\{P(w) \mid M, w \models \varphi\}$ and $M, v \models \varphi$
iff $v \in Max(\{w \in M \mid M, w \models \varphi\}, \leq^0)$. QED

This proposition is important because it shows that in a single-agent setting, updating in the BMS style amounts to perform belief revision in the AGM style as described in Section 4.2.2. This complements the correspondence established for the multi-agent case in Chapter3 between these two approaches to belief change.

Let us have a closer look at the mechanisms involved in this process. Suppose φ is false in every conceived world. When we revise by φ then the surprising worlds where φ is true and which have the least *global* degree become the conceived worlds. More interestingly, the *local* structure of these surprising worlds remains the same, that is to say their relative order of probability is the same before and after the revision. So we see here that the richness of our formalism enables not only a fine-grained account of the agent's epistemic state as we saw in Section 5.3.1 but also a fine-grained account of belief revision.

Unsurprisingly, our framework fulfills the AGM postulates.

Theorem 5.4.7 *If, as in the AGM theory, we restrict our attention to propositional beliefs then our update mechanism satisfies the eight AGM postulates for belief revision (see Section 4.2.2).*

PROOF. To check whether the AGM postulates are fulfilled, we first need to define the belief set, the expanded belief set and the revised belief set associated to a pd-model. We will deal with the propositional language $\mathcal{L}_0$ as in the AGM theory.

Definition 5.4.8 (Belief set, revision, expansion)
Let (M, w_a) be a pd-model and $\varphi \in \mathcal{L}_0$. We define

- the *belief set* $K = \{\psi \in \mathcal{L}_0 \mid M, w_a \models B\psi\}$,
- the *revision* of the belief set K by φ, $K * \varphi = \{\psi \in \mathcal{L}_0 \mid M \otimes A(\varphi), (w_a, a) \models B\psi\}$,
- the *expansion* of the belief set K by φ, $K + \varphi = \{\psi \in \mathcal{L}_0 \mid M, w_a \models B(\varphi \to \psi)\}$.

□

Now we can prove the theorem.

K*1 This postulate is clearly satisfied.

K*2 This postulate is satisfied, because we deal with propositional formulas which are persistent formulas (that is formulas which remain true after an announcement if true beforehand).

K*3 $M \otimes A(\varphi), (w_a, a) \models B\psi$

$\Leftrightarrow \sum\{P(w, a) \mid M \otimes A(\varphi), (w, a) \models \psi\} = 1$

$\Leftrightarrow \sum\{\frac{P(w)}{\sum\{P(v) | v \in M \text{ and } M, v \models \varphi\}}; w \in M \text{ and } M, w \models \varphi \text{ and } M, w \models \psi\} = 1$ because propositional formulas are persistent.

$\Leftrightarrow \sum\{P(w) \mid w \in M \text{ and } M, w \models \varphi \text{ and } M, w \models \psi\} = \sum\{P(w) \mid w \in M \text{ and } M, w \models \varphi\}$

$\Rightarrow^* \sum\{P(w) \mid w \in M \text{ and } M, w \models \varphi \to \psi\} = 1$ (1)

$\Leftrightarrow M, w_a \models B(\varphi \to \psi)$

K*4 $\neg\varphi \notin K$

$\Leftrightarrow M, w_a \models \neg B \neg\varphi$

$\Leftrightarrow$ there is a conceived world $w \in M$ such that $M, w \models \varphi$ (H).

We have to prove the other direction of $\Rightarrow^*$.

Formula (1) tells us that for all conceived worlds $w \in M$ $M, w \models \varphi \to \psi$. So,

$\sum\{P(w) \mid w \in M \text{ and } M, w \models \varphi\}$

$= \sum\{P(w) \mid w \in M$ and $M, w \models \varphi$ and w conceived $\}$ by (H), because if v is conceived and v' surprising then $P(v) + P(v') = P(v)$ (see Section 5.2)

$= \sum\{P(w) \mid w \in M$ and $M, w \models \varphi$ and $M, w \models \psi$ and w conceived $\}$ by (1)

$= \sum\{P(w) \mid w \in M$ and $M, w \models \varphi$ and $M, w \models \psi\}$ because there does exist a conceived world w satisfying $\varphi \wedge \psi$ by (H) and (1).

K*5 It is fulfilled because $K * \varphi \neq K_\perp$.

Indeed, otherwise $M \otimes A(\varphi), (w_a, a) \models B \perp$. Then $M \otimes A(\varphi)$ does not have conceived worlds, so $M \otimes A(\varphi)$ is not a pd-model. This is impossible because the announcement of φ is supposed to be truthful (i.e. $P^{w_a}(a) > 0$, see the footnote on p. 106) so the updated model should be a pd-model.

K*6 This postulate is clearly satisfied.

K*7 First note that $M \otimes A(\varphi), w_a \models B(\varphi' \to \psi) \Leftrightarrow \varphi \in K^*\varphi + \varphi'$

$M \otimes A(\varphi \wedge \varphi'), w_a \models B\psi$

$\Leftrightarrow$ if $M, w_a \models \varphi \wedge \varphi'$ then $M \otimes A(\varphi \wedge \varphi'), (w_a, a) \models B\psi$

$\Leftrightarrow M \otimes A(\varphi \wedge \varphi'), (w_a, a) \models B\psi$ because $M, w_a \models \varphi \wedge \varphi'$ (see the footnote on p. 106)

$\Leftrightarrow \sum\{P(w, a) \mid w \in M$ and $M, w \models \varphi \wedge \varphi'$ and $M, w \models \psi\} = 1$

$\Leftrightarrow \sum\{\frac{P(w)}{\sum\{P(v) | v \in M \text{ and } M, v \models \psi \wedge \varphi'\}} \mid w \in M$ and $M, w \models \varphi \wedge \varphi'$ and $M, w \models \psi\} = 1$

$\Leftrightarrow \sum\{P(w) \mid w \in M$ and $M, w \models \varphi \wedge \varphi'$ and $M, w \models \varphi\} = \sum\{P(w) \mid w \in M$ and $M, w \models \varphi \wedge \varphi'\}$

$\Leftrightarrow \sum\{\frac{P(w)}{\sum\{P(v) | v \in M \text{ and } M, v \models \varphi\}} \mid w \in M$ and $M, w \models \varphi \wedge \varphi'$ and $M, w \models \psi\} =$

$\sum\{\frac{P(w)}{\sum\{P(v) | v \in M \text{ and } M, v \models \varphi\}} \mid w \in M$ and $M, w \models \varphi \wedge \varphi'\}$

$\Leftrightarrow \sum\{P(w, a) \mid (w, a) \in M \otimes A(\varphi)$ and $M, w \models \varphi \wedge \varphi'$ and $M, w \models \psi\} = \sum\{P(w, a) \mid (w, a) \in M \otimes A(\varphi)$ and $M, w \models \varphi \wedge \varphi'\}$

$\Leftrightarrow \sum\{P(w, a) \mid (w, a) \in M \otimes A(\varphi)$ and $M \otimes A(\varphi), (w, a) \models \varphi'$ and $M \otimes A(\varphi), (w, a) \models \psi\} = \sum\{P(w, a) \mid (w, a) \in M \otimes A(\varphi)$ and $M \otimes A(\varphi), (w, a) \models \varphi'\}$

$\Rightarrow^{*'} \sum\{P(w, a) \mid (w, a) \in M \otimes A(\varphi)$ and $M \otimes A(\varphi), (w, a) \models \varphi' \to \psi\} = 1$ (1′)

$\Leftrightarrow M \otimes A(\varphi), (w_a, a) \models B(\varphi' \to \psi)$

$\Leftrightarrow \psi \in K^*\varphi + \varphi'$

K*8 $\neg\varphi' \notin K^*\varphi$

$\Leftrightarrow M \otimes A(\varphi), (w_a, a) \models \neg B \neg \varphi'$

$\Leftrightarrow$ there is a conceived world $(w, a) \in M \otimes A(\varphi)$ such that $M \otimes A(\varphi), (w, a) \models \varphi'$ (H′)

We have to prove the other direction of $\Rightarrow^{*'}$. (1′) tells us that for all conceived worlds $(w,a) \in M \otimes A(\varphi)\ M \otimes A(\varphi), (w,a) \models \varphi' \rightarrow \psi$. So,

$\sum\{P(w,a) \mid (w,a) \in M \otimes A(\varphi)$ and $M \otimes A(\varphi), (w,a) \models \varphi'\}$

$= \sum\{P(w,a) \mid (w,a) \in M \otimes A(\varphi)$ and $M \otimes A(\varphi), (w,a) \models \varphi'$ and (w,a) is conceived $\}$ by (H′), because of (H′) and if v is conceived and v' surprising then $P(v)+P(v') = P(v)$ (see Section 5.2).

$= \sum\{P(w,a) \mid (w,a) \in M \otimes A(\varphi)$ and $M \otimes A(\varphi), (w,a) \models \varphi'$ and $M \otimes A(\varphi), (w,a) \models \psi$ and (w,a) conceived $\}$ by (1′)

$= \sum\{P(w,a) \mid (w,a) \in M \otimes A(\varphi)$ and $M \otimes A(\varphi), (w,a) \models \varphi'$ and $M \otimes A(\varphi), (w,a) \models \psi\}$ by (H′).

So the other direction of $\Rightarrow^{*'}$ is proved.

QED

Adding knowledge to our formalism

First, note that the semantics for the belief operator B defined in Definition 5.3.4 could also be defined in terms of an accessibility relation D: if (M, w_a) is a pd-model then for all $w \in M$, $D(w) = \{v \in M \mid P(v)$ is real$\}$. Now we are going to give the counterparts in our formalism of the four extensions of the accessibility relation D for belief to the accessibility relation R for knowledge spelled out in the last paragraph of the previous section. In the sequel, $(M, w_a) = (W, P, V, w_a)$ is a pd-model and $w \in M$.

1. The first possible extension was that the accessibility relation for knowledge is just the reflexive closure of the accessibility relation for belief. Formally, this amounts to define $R(w)$ as

$$R(w) = D(w) \cup \{(w,w)\}.$$

   One can indeed check that R satisfies .4: for all w, v, if (wRv and $w \neq v$) then (for all u, (if wRu then uRv)) (see Section 2.2.1).

2. The second possible extension was already completely specified as follows: for all $w, v \in M$, wRv iff ((wDw and wDv) or (not wDw)). Equivalently, this amounts to state that

$$R(w) = \begin{cases} D(w) & \text{if } w \text{ is a conceived world} \\ W & \text{otherwise.} \end{cases}$$

One can indeed check that R satisfies .3.2: for all w, v, u, if (wRu and not uRw) then (wRv implies vRu) (see Section 2.2.1).

3. The third possible extension specifies that for all $w \in M$ and all $\varphi \in \mathcal{L}_0$, $M, w \models K\varphi$ iff ($M, w \models \varphi \wedge B\varphi$ and for all $\psi \in \mathcal{L}_0$, if $M, w \models \psi$ then $M \otimes A(\psi), (w, a) \models B\varphi$). One can easily show thanks to Proposition 5.4.6 that this amounts to define $R(w)$ as

$$R(w) = \{v \in M \mid v^0 \geq^0 w^0\}.$$

One can indeed check that R satisfies Weak Connectedness: for all w, v, u, if (wRv and wRu) then (vRu or $u = v$ or uRv) (see Section 2.2.1).

4. The last possible extension specifies that for all $w \in M$ and all $\varphi \in \mathcal{L}_0$, if ($M, w \models \varphi \wedge B\varphi$ and for all $\psi \in \mathcal{L}_0$, if $M, w \models \psi$ then $M \otimes A(\psi), (w, a) \models B\varphi$) then $M, w \models K\varphi$, but not necessarily the other way round. One can easily show that this amounts to constrain $R(w)$ as follows

$$D(w) \subseteq R(w) \subseteq \{v \in M \mid v^0 \geq^0 w^0\}.$$

In that case, the only semantic constraint we can impose on the accessibility relation for knowledge is Confluence: for all w, v, u, if (wRv and wRu) then there is z such that (vRz and uRz) (see Section 2.2.1).

The fourth possible extension is the one adopted by Stalnaker. The advantage of the third extension is that the specification of the accessibility relation R for knowledge is completely determined by the belief structure of the pd-model. This entails that the dynamics of knowledge is also completely determined by the dynamics of belief that we already defined. This is not the case for the other definitions where we would have to add some further conditions in the update product of Definition 5.5.3 in order to specify how to update the accessibility relations R. The problem would then be to determine such conditions.

5.5 Adding agents

A generalization of our formalism to the multi-agent case could consist in simply indexing the probabilities P^Γ of the generic event model by the agents $j \in G$; and indexing the probabilities P of the pd-model (resp. generic event model) by the agents $j \in G$ and possible worlds w (resp. possible events a). If we do so we have to allow these probability measures to take the value 0 (note that it was not the case for the single-agent case). $P_{j,w}(v) = 0$ would then mean that if the agent j is in world w then the world v is not relevant to describe her own epistemic state (and similarly for possible events). However, if we want to have a formalism equivalent to the first one, we have to add constraints to the $P_{j,w}$ (and $P_{j,a}$), namely

1. $P_{j,w}(w) > 0$ for all $w \in W$,
2. if $P_{j,w}(v) > 0$ then for all $u \in W$ $P_{j,w}(u) = P_{j,v}(u)$,

3. $\sum\{P_{j,w}(v) \mid v \in W\} = 1$.

Note that these constraints look similar to seriality (3), transitivity and euclidianity (2) constraints. Finally, the update product is the same except that the probabilities P have to be replaced by $P_{j,w}$ (or $P_{j,a}$) and the probabilities P^Γ by P_j^Γ.

This generalization could be criticized because we have to resort to the 3 constraints above. We propose below another equivalent generalization where we do not have to resort to additional constraints. Instead, we introduce a rough uncertainty relation R_j for each agent. Like for the single agent case we divide our task into three parts: the static part, the dynamic part and finally the update mechanism.

5.5.1 The static part

Definition 5.5.1 (Multi-agent proba-doxastic model)
A *multi-agent proba-doxastic model* is a tuple $M = (W, R, P, V, w_a)$ where:

1. W is a finite set of possible worlds;
2. w_a is the possible world corresponding to the actual world;
3. $R : G \to 2^{W \times W}$ is an *equivalence* relation defined on W for each agent j;
4. $P : G \to (W \to]0; 1])$ is a probability measure for each agent j defined on each equivalence class $R_j(w)$ such that
$$\sum\{P_j(v) \mid v \in R_j(w)\} = 1;$$
5. $V : \Phi \to 2^W$ is a valuation.

□

We introduce for each agent $j \in G$ an equivalence relation R_j on the set of worlds W, modeling agent j's rough uncertainty. This equivalence relation R_j partitions the model into equivalence classes which can each be considered as a 'single-agent' pd-model with probability measure P_j.

5.5.2 The dynamic part

Definition 5.5.2 (Multi-agent generic action model)
A *multi-agent generic action model* is a structure $A = (E, S, R, P, \{P^\Gamma \mid \Gamma \text{ is a maximal consistent subset of } S\}, \{Pre_a \mid a \in E\}, a_a)$ where:

1. E is a finite set of possible events;
2. a_a is the actual event;
3. S is a set of formulas of $\mathcal{L}_{St}$ closed under negation;

4. $R : G \to 2^{E \times E}$ is an *equivalence* relation defined on E for each agent j;

5. $P^{\Gamma} : G \to (E \to [0;1])$ is a probabilistic measure (assigning *real* numbers in [0;1]) for each agent j and each maximal consistent subset Γ of S defined on each equivalence class $R_j(a)$ such that
$$\sum\{P_j^{\Gamma}(b) \mid b \in R_j(a)\} = 1;$$
Besides, for each possible event a and agent j_0
$$\text{if } P_{j_0}^{\Gamma}(a) = 0 \text{ then } P_j^{\Gamma}(a) = 0 \text{ for all } j \in G \text{ (*)};$$

6. $P : G \to (E \to]0;1])$ is a probabilistic measure for each agent j defined on each equivalence class $R_j(a)$ and such that
$$\sum\{P_j(b) \mid b \in R_j(a)\} = 1;$$

7. $Pre_a : \Phi \to \mathcal{L}_{St}$ is a function indexed by each possible event a.

□

We proceed as in the static part by introducing for each agent $j \in G$ an equivalence relation R_j on the set of possible events E modeling agent j's rough uncertainty. This equivalence relation R_j partitions the multi-agent generic event model in equivalence classes which can each be considered as a 'single-agent' generic event model with probability measures P_j and P_j^{Γ}.

Moreover, we assume that the condition under which an event cannot physically take place in a world is common knowledge among agents. That is why we add in the definition of the (multi-agent) generic event model that for each possible event a and agent $j_0 \in G$, if $P_{j_0}^{\Gamma}(a) = 0$ then $P_j^{\Gamma}(a) = 0$ for all $j \in G$ (note that this constraint should also be added to our first proposal of generalization to the multi-agent case).

5.5.3 The update mechanism

Definition 5.5.3 (Update product)
Given a multi-agent pd-model $M = (W, R, P, V, w_a)$ and a multi-agent generic event model $A = (E, S, R, P, \{P^{\Gamma} \mid \Gamma \text{ is a m.c. subset of } S\}, \{Pre_a \mid a \in E\}, a_a)$, we define their *update product* to be the pd-model $M \otimes A = (W', R', P', V', w'_a)$, where:

1. $W' = \{(w, a) \in W \times E \mid P_j^w(a) > 0\}$;

2. $(v, b) \in R_j(w, a)$ iff $v \in R_j(w)$ and $b \in R_j(a)$;

3. We set
$$P'_j(a) = \frac{P_j(a) \cdot P_j^W(a)}{\sum\{P_j(b) \cdot P_j^W(b) \mid b \in E\}} \text{ where } P_j^W(a) = \sum\{P_j(v) \cdot P_j^v(a) \mid v \in W\};$$

Then

$$P'_j(w,a) = \frac{P_j(w)}{\sum\{P_j(v) \mid v \in W \text{ and } P^v_j(a) > 0\}} \cdot P'_j(a);$$

4. $V'(p) = \{(w,a) \in W' \mid M, w \models Pre_a(p)\}$;
5. $w'_a = (w_a, a_a)$.

□

The update mechanism is the same as for the single-agent case: the P operators just have to be indexed by the agents $j \in G$. However we also have to deal with the new component R. As in BMS, we set $(v,b) \in R_j(w,a)$ iff $v \in R_j(w)$ and $b \in R_j(a)$ because the uncertainty relations R_j for the multi-agent pd-model and the multi-agent generic event model are assumed to be independent from one another.

5.6 Comparisons

Comparison with Kooi's system. Kooi's dynamic probabilistic system [Kooi, 2003] is based on the static approach by Fagin and Halpern in [Fagin and Halpern, 1988; Fagin and Halpern, 1994]. Unlike in our system, probability is not meant only to model the agents' epistemic states. In that respect, his probability measures are defined relatively to each world without any constraint on them. Moreover he only deals with public announcement. But in this particular case our update mechanism is a bit different from his. The worlds of his initial model are the same as in his updated model, only the accessibility relations and probability distributions are changed, depending on whether or not the probability of the formula announced is zero in the initial model. However, our probabilistic update rule in this particular case boils down to the same as his for the worlds where the probability of the formula announced is different from zero. Finally, he does not consider events changing facts (he tackles this topic independently in [Kooi, 2007] and [van Ditmarsch *et al.*, 2005]).

Comparison with van Benthem's system. van Benthem's early system [van Benthem, 2003] is similar to ours in its spirit and goals. However he does not introduce the probabilities $P^W(a)$ and $P(a)$ but only a single $P^w(a)$. Hence, his probabilistic update rule is different. The intended interpretation of his $P^w(a)$ seems also to be different from ours if we refer to his example. Anyway, his discussion and comparison with the Bayesian setting in his Section 5 are still valid here.

A more elaborated version of his system which is very similar to ours has been developed independently by him, Kooi and Gerbrandy [van Benthem *et al.*, 2006a]. In this system, they have three kinds of probability which correspond in their terms to our three kinds of probability: prior probabilities on worlds (here $P(w)$), observation probabilities for events (here $P(a)$), and occurrence probabilities on events in worlds (here $P^w(a)$). Nevertheless, their probabilistic update rule is still different and does not comply to the Jeffrey update, contrary to ours. They also study some parameterized versions of their probabilistic rule and they

show that one of them actually complies to the Jeffrey update. They provide a sketch of a completeness proof via reduction axioms. However, they do not resort to infinitesimals to represent epistemic states and thus can neither express degrees of potential surprise nor allow for belief revision. Finally they do not consider events changing facts.

Comparison with Baltag and Smets' system. In [Baltag and Smets, 2007], Baltag and Smets propose a system which combines update logic, belief revision and probability, like our system. But instead of resorting to infinitesimals to deal with belief revision in a probabilistic setting, they resort to Popper-Renyi theory of conditional probabilities. On the one hand and despite this different primary formalism, their static part has similar features to ours. Indeed, our ranking defined in Definition 5.4.2 corresponds to what they call a priority relation (they even mention that this was originally called "ranking ordering" by different authors). Besides, they also use Stalnaker's non-standard notion of knowledge but in their system it is not intended to model the notion of knowledge but rather what they call 'safe belief', being a very strong belief. They even refine this notion by defining degrees of safety, these degrees belonging to $[0; 1]$ and a safe belief being a belief of degree of safety 1. They also analyze the role of this concept in game theory. Their notion of knowledge is instead modeled by a S5 modality: a formula is known by the agent if it is true in all the worlds of the pd-model. Their primary notion being conditional probability instead of standard probability, they also introduce a notion of conditional belief. On the other hand, their dynamic and update parts are different. The uncertainty about events is still assumed to be independent of the uncertainty about the world. So they still use rough preconditions and therefore cannot really account for the intricate logical dynamics present in the interpretation of events that we have studied. In consequence their probabilistic update rule is also different and is based on the principle that "beliefs of changes induce (and "encode") changes of beliefs". Finally, unlike our system, they are able to axiomatize their semantics but they do not deal with events changing facts.

Comparison with the situation calculus of Bacchus, Halpern and Levesque. Their system [Bacchus *et al.*, 1999] can be viewed as the counterpart of van Benthem's early system in the situation calculus except that they deal as well with events changing facts. Their probabilistic update rule is also the same as van Benthem's (modulo normalization). So what applies to van Benthem's early system applies here too. In particular, the logical dynamics present in the interpretation of an event are not explored.

Comparison with the observation systems of Boutilier, Friedman and Halpern. Their system [Boutilier *et al.*, 1998] deals with noisy observations. Their approach is semantically driven like ours. However they use a different formalism called observation system based on the notion of (ranked) interpreted system. On the one hand their system is more general because it incorporates the notion of time and a ranking of evolutions of states over time (called runs). On the other hand the only events they consider are noisy observations (which do not change facts of the situation). An advantage of our system is its versatility because we can represent many kinds of events. In that respect, their noisy observations can be modelled in our formalism using two possible events, the first corresponding to a truthful

observation and the second to an erroneous one. Then by a suitable choice of probabilities we can for example express, as they do, that the observation is "credible". However, their formalism seems to enable them to characterize formally more types of noisy observations. Finally, because we do not introduce the notions of time and history, our formalism is rather comparable to a particular case of their system called *Markovian* observation system. But nothing precludes us to introduce these notions as an extension of our system.

5.7 Conclusion

In order to represent as accurate as possible the agent's epistemic state, we have introduced a rich formalism based on hyperreal numbers (and which is an extension of Weydert's cumulative algebra). Our epistemic state representation includes both degrees of belief expressed by a subjective probability and degrees of potential surprise expressed by infinitesimals. We have seen that the richness of this formalism enabled genuine belief revision thanks to the existence of infinitesimals (and multiplicative inverse). By a closer look at this revision process, we could even notice some interesting and meaningful patterns due to the dual aspect (local and global) of this formalism. So, our formalism indirectly offers a new (probabilistic) approach to belief revision.

But other important logical dynamics were studied, namely the ones present in the process of interpreting an event. Starting from the observation that this interpretation hinges on two features, the actual perception of the event happening and our expectation of it to happen, we have proposed a way to model this phenomenon. Incidentally, note that in a sense our approach complements and reverses the classical view (as in belief revision theory) whereby only our interpretation of events affects our beliefs and not the other way round. For sake of generality, we have also taken into account in our formalism events that may change the facts of a situation.

Finally, we have reviewed different approaches to model the notion of knowledge and showed how they can be captured in our formalism in a straightforward way. This illustrates the richness and expressivity of our formalism.

Our formalism is semantically driven and it would be interesting to look for a completeness result, and in particular for reduction axioms. But because of its high expressivity, it is likely that we will not get decidability (and therefore, that such reduction axioms cannot be formulated). However this system can be of use as it is in several areas. Firstly, in game theory where the kinds of phenomena we studied are quite common. Secondly, in psychology if we want to devise realistic formal models of belief change. Finally, the logical dynamics we modeled could be used in artificial intelligence.

Chapter 6

Exploring the power of converse events

6.1 Introduction

In this chapter we follow the perfect external approach. Our first aim is to enrich the (dynamic) epistemic language with a modal operator expressing what was true before an event occurr*ed*. Our second aim is to propose a unified language which does not refer in its syntax to an event model as in the BMS formalism. Indeed, this event model can be viewed as a semantic object and it seems to us inappropriate to introduce it directly into the syntax of the language (although the way it is actually done in the BMS formalism is formally correct).

Semantics of events: products vs. accessibility relations. Expressing within the BMS formalism what was true before an event a occurr*ed*, i.e. to give semantics to the converse event a^- is not simple.

On the other hand, in PDL [Harel *et al.*, 2000], events are interpreted as transition relations on possible worlds, and not as restricted products of models as in BMS. Converse events a^- can then easily be interpreted by inverting the accessibility relation associated to a. The resulting logic is called the tense extension of PDL. To this we then add an epistemic accessibility relation. We call (tensed) Epistemic Dynamic Logic EDL the combination of epistemic logic and PDL with converse.[1]

A semantics in terms of transition relations is more flexible than the BMS product semantics: we have more options concerning the interaction between events and beliefs. In Section 6.2, we will propose an account for this delicate relationship by means of constraints on the respective accessibility relations: a no-forgetting and a no-learning constraint, and a constraint of epistemic determinism.

[1] EDL is related to Segerberg's Doxastic Dynamic Logic DDL [Segerberg, 1995; Segerberg, 1999]. But research on DDL focusses mainly on its relation with AGM theory of belief revision, and studies particular events of the form $+\varphi$ (expansion by φ), $*\varphi$ (revision by φ), and $-\varphi$ (contraction by φ).

Translating BMS into EDL. To demonstrate the power of our approach we will provide in Section 6.3 a translation from BMS to EDL. To do so, we will express the structure of an event model A by a nonlogical theory $\Gamma(A)$ of EDL, and prove that any formula φ is valid in BMS if and only if it is a logical consequence of $\Gamma(A)$ in EDL.

So, unlike BMS, we avoid to refer to a semantical structure (i.e. the BMS event model A) in the very definition of the language. Encoding the structure of a BMS event model A by a nonlogical theory $\Gamma(A)$ of EDL is done thanks to converse events. For example $[a]B_i(\langle a^-\rangle\top \vee \langle b^-\rangle\top)$ expresses that agent i perceives the occurrence of a as that of either a or b.

Finally, in Section 6.4, we conclude and compare our formalism with related works.

6.2 EDL: Epistemic Dynamic Logic with converse

6.2.1 The language $\mathcal{L}_{\mathsf{EDL}}$ of EDL

Just as for BMS (see Chapter 3), we suppose given sets of propositional symbols Φ and of agent symbols G, and a *finite* set of event symbols E.

Definition 6.2.1 (Language $\mathcal{L}_{\mathsf{EDL}}$)
The language $\mathcal{L}_{\mathsf{EDL}}$ is defined as follows

$$\mathcal{L}_{\mathsf{EDL}} : \varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \varphi' \mid B_j\varphi \mid [a]\varphi \mid [a^-]\varphi,$$

where p ranges over Φ, j over G and a over E.

The dual modal operators $\langle a\rangle$ and $\langle a^-\rangle$ are defined as follows: $\langle a\rangle\varphi$ abbreviates $\neg[a]\neg\varphi$; $\langle a^-\rangle\varphi$ abbreviates $\neg[a^-]\neg\varphi$. □

The formula $[a]\varphi$ reads " φ holds after every possible occurrence of event a". $[a^-]\varphi$ reads "φ held before a".

Note that the language $\mathcal{L}_A$ of Definition 3.2.5 is the set of those formulas of $\mathcal{L}_{\mathsf{EDL}}$ that do not contain the converse operator $[a^-]$. Note also that the epistemic language $\mathcal{L}$ of Definition 2.2.2 is the set of those formulas of $\mathcal{L}_{\mathsf{EDL}}$ that do not contain any dynamic operator, i.e. built from Φ, the Boolean operators and the B_j operators alone. For example $[a]B_j[a^-]\bot$ is an $\mathcal{L}_{\mathsf{EDL}}$-formula that is not in $\mathcal{L}_A$.

6.2.2 Semantics of EDL

When designing models of events and beliefs the central issue is to account for the interplay of these two concepts. In our PDL-based semantics this is done by means of constraints on the respective accessibility relations.

Definition 6.2.2 (EDL-model, no-forgetting, no-learning, epistemic determinism)
An *EDL-model* is a tuple $M = (W, R, \mathcal{R}, V)$ such that

- W is a set of possible worlds;

- $R : G \to 2^{W \times W}$ assigns an accessibility relation to each agent;
- $\mathcal{R} : E \to 2^{W \times W}$ assigns an accessibility relation to each possible event; and
- $V : \Phi \to 2^W$ is a valuation.

Moreover an EDL-model satisfies the constraints of *no-forgetting*, *no-learning* and *epistemic determinism*:

nf	If $v' \in \mathcal{R}_a \circ R_j(w)$ then there is $b \in E$ such that $v' \in R_j \circ \mathcal{R}_b(w)$.
nl	If $(\mathcal{R}_a \circ R_j \circ \mathcal{R}_b^{-1})(w) \neq \emptyset$ then $(R_j \circ \mathcal{R}_b)(w) \subseteq (\mathcal{R}_a \circ R_j)(w)$.
ed	If $v_1, v_2 \in \mathcal{R}_a(w)$ then $R_j(v_1) = R_j(v_2)$.

We write $\mathcal{R}_a^{-1}(v) = \{w \mid w \in \mathcal{R}_a^{-1}(v)\} = \{w \mid v \in \mathcal{R}_a(w)\}$. □

The *no-forgetting* principle says that if after an event agent j considers a world v' possible, then before this event agent j already considered possible that there was an event leading to this world (see Figure 6.1, left). So everything agent j considers possible after the performance of an event stems from what she considered possible before the event. This principle is a generalization of the perfect-recall principle [Fagin *et al.*, 1995].

To understand the principle *no-learning*, also known as no-miracle [van Benthem and Pacuit, 2006], assume that agent j perceives the occurrence of a as that of b_1, $b_2 \ldots$ or b_n. Then, informally, the *no-learning* principle says that *all* such alternatives resulting from occurrence of b_1, $b_2, \ldots$, b_n in j's alternatives before a are indeed alternatives after a. In a sense there is no-miracle: everything the agent was supposed to consider possible after the event is indeed considered possible after the event if this one actually takes place. Formally, assume that agent j perceives b as a possible alternative of a, i.e. $(\mathcal{R}_a \circ R_j \circ \mathcal{R}_b^{-1})(w) \neq \emptyset$. If at w world v' was a possible outcome of event b for j, then v' is possible for j at some $v \in \mathcal{R}_a(w)$ (see Figure 6.1, middle).

Finally, the *epistemic determinism* principle says that an agent's epistemic state after an event does not depend on the particular nondeterministic outcome. Formally, suppose we have $w\mathcal{R}_a v_1$ and $w\mathcal{R}_a v_2$. Then ed forces that the epistemic states at v_1 and v_2 are identical: $R_j(v_1) = R_j(v_2)$ (see Figure 6.1, right).

These constraints restrict the kind of events we consider. Our events are such that the epistemic state of an agent after the occurrence of an event depends only on the previous epistemic state of the agent and on how the event is perceived by the agent, and *not* on what is true in the world before or after the event. This feature of our events is somehow formally captured by Proposition 6.2.3 below: $R_j(w)$ is the epistemic state of the agent before the event and $A_{a,w} = \{b \in E \mid \mathcal{R}_a \circ R_j \circ \mathcal{R}_b^{-1}(w) \neq \emptyset\}$ specifies how the event a is perceived by the agent. For example an agent testing whether φ is the case is not an event dealt with by our formalism. Indeed the epistemic state of this agent after the test (the agent knowing whether φ is true) depends on the actual state of the world (whether φ is true or not). In this example the no-learning constraint might be violated. Another example of event which is not dealt with by our formalism is that of tossing a coin and looking at it. In this example, the epistemic state of the agent after the toss depends on the state of the world after the event,

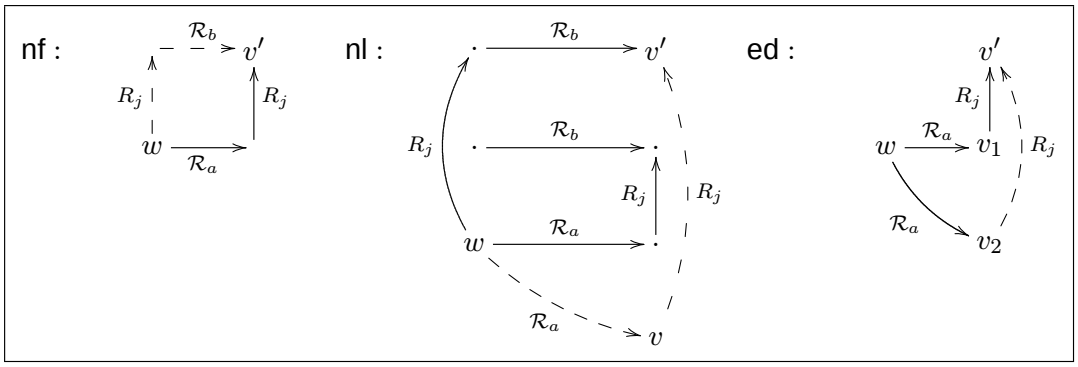


Figure 6.1: *no-forgetting*, *no-learning* and *epistemic determinism* constraints

i.e. whether the coin lands heads or tails up. Here the epistemic determinism constraint is violated. On the other hand, both public and private announcement are dealt with by our framework. More generally, any kind of announcement (public, private...) about any kind of information (epistemic, stating that an event just occurred...) is dealt with by our framework. Our events are sometimes called ontic events, feedback-free events or uninformative events [Herzig *et al.*, 2000; de Lima, 2007].

Proposition 6.2.3 *Let $M = (W, R, \mathcal{R}, V)$ be a tuple. M is an EDL-model, i.e. M satisfies* nf, nl, ed, *iff for all $j \in G$, all $w \in M$, all $a \in E$, all $w' \in \mathcal{R}_a(w)$,*

$$R_j(w') = \bigcup\{\mathcal{R}_b(v) \mid b \in A_{a,w}, v \in R_j(w)\}(*),$$

where $A_{a,w} = \{b \in E \mid \mathcal{R}_a \circ R_j \circ \mathcal{R}_b^{-1}(w) \neq \emptyset\}$.

PROOF.

- Assume M satisfies nf, nl and ed.
  - Let $v' \in R_j(w')$. Then $v' \in \mathcal{R}_a \circ R_j(w)$. So by nf there is $b \in E$ and $v \in R_j(w)$ such that $v' \in \mathcal{R}_b(v)$. So $\mathcal{R}_a \circ R_j \circ \mathcal{R}_b^{-1}(w) \neq \emptyset$ and $b \in A_{a,w}$. So $v' \in \cup\{\mathcal{R}_b(v) \mid b \in A_{a,w}, v \in R_j(w)\}$.
  - Let $v' \in \cup\{\mathcal{R}_b(v) \mid b \in A_{a,w}, v \in R_j(w)\}$. Then there is $b \in E$ such that $v' \in R_j \circ \mathcal{R}_b(w)$ and $\mathcal{R}_a \circ R_j \circ \mathcal{R}_b^{-1}(w) \neq \emptyset$. So by nl, $v' \in \mathcal{R}_a \circ R_j(w)$, i.e. there is $w'' \in \mathcal{R}_a(w)$ such that $v' \in R_j(w'')$. Then by ed, $v' \in R_j(w')$.
- Assume M satisfies (*).

  nf Assume that $v' \in \mathcal{R}_a \circ R_j(w)$. Then there is $w' \in \mathcal{R}_a(w)$ such that $v' \in R_j(w')$. By (*) there is $b \in A_{a,w}$ and $v \in R_j(w)$ such that $v' \in \mathcal{R}_b(v)$. So there is $b \in E$ such that $v' \in R_j \circ \mathcal{R}_b(w)$.

  nl Assume that $\mathcal{R}_a \circ R_j \circ \mathcal{R}_b^{-1}(w) \neq \emptyset$ and $v' \in R_j \circ \mathcal{R}_b(w)$. Then there is $v \in R_j(w)$ and $b \in A_{a,w}$ such that $v' \in \mathcal{R}_b(v)$. So $v' \in R_j(w')$ for all $w' \in \mathcal{R}_a(w)$, i.e. $v' \in \mathcal{R}_a \circ R_j(w)$.

ed is clearly fulfilled.

QED

Definition 6.2.4 (Truth conditions for $\mathcal{L}_{\mathsf{EDL}}$)
The semantics of $\mathcal{L}_{\mathsf{EDL}}$ is defined inductively as follows. Let M be an EDL-model and $w \in M$.

$$\begin{array}{lll} M, w \models \top & & \\ M, w \models p & \text{iff} & w \in V(p) \\ M, w \models \varphi \wedge \varphi' & \text{iff} & M, w \models \varphi \text{ and } M, w \models \varphi' \\ M, w \models B_j\varphi & \text{iff} & \text{for all } v \in R_j(w), M, v \models \varphi \\ M, w \models [a]\varphi & \text{iff} & \text{for all } v \in \mathcal{R}_a(w), M, v \models \varphi \\ M, w \models [a^-]\varphi & \text{iff} & \text{for all } v \in \mathcal{R}_a^{-1}(w), M, v \models \varphi. \end{array}$$

Truth of φ in a EDL-model M is written $M \models \varphi$ and is defined as: $M, w \models \varphi$ for every $w \in M$. Let Γ be a set of $\mathcal{L}_{\mathsf{EDL}}$-formulas. The (global) consequence relation is defined by:

$$\Gamma \models_{\mathsf{EDL}} \varphi \text{ iff for every EDL-model } M, \text{ if } M \models \psi \text{ for every } \psi \in \Gamma \text{ then } M \models \varphi.$$

□

For example we have

$$\{[b]\varphi, \langle a\rangle B_j\langle b^-\rangle\top\} \models_{\mathsf{EDL}} [a]B_j\varphi$$

and

$$\models_{\mathsf{EDL}} (B_j[b]\varphi \wedge \langle a\rangle B_j\langle b^-\rangle\top) \to [a]B_j\varphi. \text{ (*)}$$

Consider $\varphi = \bot$ in (*): $B_j[b]\bot$ means that perception of event b was unexpected by agent j, while $\langle a\rangle B_j\langle b^-\rangle\top$ means that j actually perceives a as b. By our no-forgetting constraint it follows that $[a]B_j\bot$. In fact, one would like to avoid agents getting inconsistent: in such situations some sort of belief revision should take place. We do not investigate this further here.

6.2.3 Completeness

Definition 6.2.5 (Proof system of EDL)
The logic EDL is defined by the multi-modal logic K for all the modal operators B_j, $[a]$ and $[a^-]$, plus the axioms schemes Conv_1, Conv_2, NF, NL and ED below:

$$\begin{array}{ll} \mathsf{Conv}_1 & \vdash_{\mathsf{EDL}} \varphi \to [a]\langle a^-\rangle\varphi \\ \mathsf{Conv}_2 & \vdash_{\mathsf{EDL}} \varphi \to [a^-]\langle a\rangle\varphi \\ \mathsf{NF} & \vdash_{\mathsf{EDL}} B_j \bigwedge_{a\in E} [a]\varphi \to \bigwedge_{a\in E} [a]B_j\varphi \\ \mathsf{NL} & \vdash_{\mathsf{EDL}} \langle a\rangle\hat{B}_j\langle b^-\rangle\top \to ([a]B_j\varphi \to B_j[b]\varphi) \\ \mathsf{ED} & \vdash_{\mathsf{EDL}} \langle a\rangle B_j\varphi \to [a]B_j\varphi \end{array}$$

□

Conv_1 and Conv_2 are the standard conversion axioms of tense logic and converse PDL. NF, NL and ED respectively axiomatize no-forgetting, no-learning and epistemic determinism.

We write $\Gamma \vdash_{\mathsf{EDL}} \varphi$ when φ is provable from the set of formulas Γ in this axiomatics.

One can then show that EDL is strongly complete:

Proposition 6.2.6 *For every set of $\mathcal{L}_{EDL}$-formulas Γ and $\mathcal{L}_{EDL}$-formula φ,*

$$\Gamma \models_{EDL} \varphi \textit{ iff } \Gamma \vdash_{EDL} \varphi.$$

PROOF. The proof follows from Sahlqvist's theorem [Sahlqvist, 1975]: all our axioms NF, NL, ED are of the required form, and match the respective constraints nf, nl, ed. QED

6.3 From BMS to EDL

In this section we show that BMS can be embedded into EDL. We do that by building a particular EDL-theory that encodes syntactically the structure of a given BMS event model A.

Definition 6.3.1 (Theory of an event model)
Let G be a finite set of agents, E a finite set of events and $A = (E, R, Pre)$ be an event model. The *theory of* A, written $\Gamma(A)$, is made up of the following non-logical axioms:

(1) $p \to [a]p$ and $\neg p \to [a]\neg p$, for every $a \in E$ and $p \in \Phi$;

(2) $\langle a\rangle\top \leftrightarrow Pre(a)$, for every $a \in E$;

(3) $[a]B_j\left((\langle a_1^-\rangle\top \vee \ldots \vee \langle a_n^-\rangle\top) \wedge ([b_1^-]\bot \wedge \ldots \wedge [b_n^-]\bot)\right)$,
where $a_1, \ldots, a_n$ is the list of all b such that $b \in R_j(a)$, and $b_1, \ldots, b_n$ is the list of all b such that $b \notin R_j(a)$;

(4) $\hat{B}_j Pre(b) \to [a]\hat{B}_j\langle b^-\rangle\top$, for every (a, b) such that $b \in R_j(a)$.

□

Axiom 1 encodes the fact that events do not change propositional facts of the world where they are performed (see definition of $V'(p)$ in Definition 3.2.3). Axiom 2 encodes the fact that an event a can occur in a world iff this world satisfies the precondition of event a (see the definition of W' in Definition 3.2.3). Axiom 3 encodes the Kripke structure of the event model. Axiom 4 encodes the definition of R'_j (see Definition 3.2.3).

Example 6.3.2 Consider that $G = \{A, B\}$ and $\Phi = \{p\}$. In Figure 6.2 we recall the event models A_1 and A_2 corresponding respectively to the public announcement of φ and the private announcement of φ to A, where $\varphi \in \mathcal{L}$. Here, $Pre(a) = \varphi$ in both models and $Pre(b) = \top$.

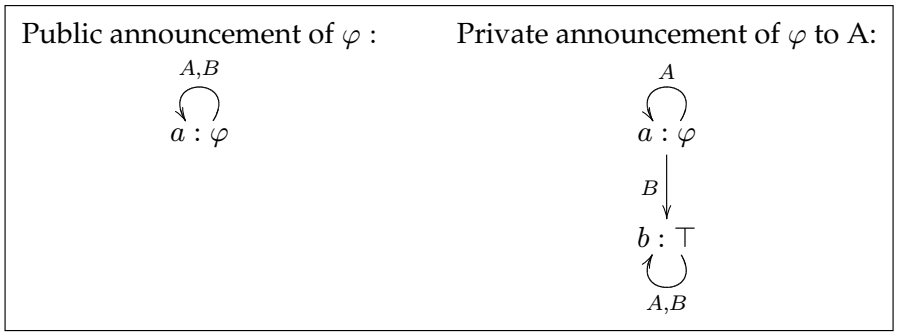


Figure 6.2: Event models for public announcement and private announcement

Applying Definition 6.3.1, we get

$\Gamma(A_1) = \{p \to [a]p$ and $\neg p \to [a]\neg p$, $\langle a\rangle\top \leftrightarrow \varphi$, $[a]B_A(\langle a^-\rangle\top)$, $[a]B_B(\langle a^-\rangle\top)$, $\hat{B_A}\varphi \to [a]\hat{B_A}\langle a^-\rangle\top\}$, $\hat{B_B}\varphi \to [a]\hat{B_B}\langle a^-\rangle\top\}$

and

$\Gamma(A_2) = \{p \to [a]p$ and $\neg p \to [a]\neg p$, $p \to [b]p$ and $\neg p \to [b]\neg p$, $\langle a\rangle\top \leftrightarrow \varphi$, $\langle b\rangle\top \leftrightarrow \top$, $[a]B_A(\langle a^-\rangle\top \wedge [b^-]\bot)$, $[a]B_B(\langle b^-\rangle\top \wedge [a^-]\bot)$, $[b]B_A(\langle b^-\rangle\top \wedge [a^-]\bot)$, $[b]B_B(\langle b^-\rangle\top \wedge [a^-]\bot)$, $\hat{B_A}\varphi \to [a]\hat{B_A}\langle a^-\rangle\top$, $\hat{B_A}\top \to [b]\hat{B_A}\langle b^-\rangle\top$, $\hat{B_B}\top \to [a]\hat{B_B}\langle b^-\rangle\top$, $\hat{B_B}\top \to [b]\hat{B_B}\langle b^-\rangle\top\}$ □

It turns out that the axiom of determinism is a logical consequence of $\Gamma(A)$ in EDL . This is comforting because the axiom of determinism is indeed valid in BMS .

Proposition 6.3.3 *Let A be an event model. For every $\mathcal{L}_A$-formula φ we have $\Gamma(A) \models_{EDL} \langle a\rangle\varphi \to [a]\varphi$.*

PROOF. Let $A = (E, R, Pre)$ be a given event model, and let M be an EDL-model such that $M \models \psi$ for every $\psi \in \Gamma(A)$. Assume $w_0\mathcal{R}_a v_0$ and $w_0\mathcal{R}_a u_0$ with $v_0 \neq u_0$. We are going to show that u_0 and v_0 are bisimilar.

Z^e is defined to be an epistemic bisimulation between models M_1 and M_2 if Z^e is a bisimulation between the restriction of these models to epistemic accessibility relations. Let $Z^e := \{(w, w) : w \in W\} \cup \{(v_0, u_0)\}$. We are going to show that Z^e is an epistemic bisimulation. To do so, we need to prove

1. $u_0 \in V(p)$ iff $v_0 \in V(p)$ for all $p \in \Phi$;
2. if $v_0 R_j v'$ then $u_0 R_j v'$;
3. if $u_0 R_j u'$ then $v_0 R_j u'$.

(1) is guaranteed by Definition 6.3.1 (1). (2) and (3) are guaranteed by epistemic determinism: ed makes that $R_j(u) = R_j(v)$.

Now from Z^e, we are going to build up a bisimulation. We proceed as follows.

$Z^0 = Z^e$;

$Z^{n+1} = \{(u_{n+1}, v_{n+1}) \mid u_n \mathcal{R}_a u_{n+1}$ and $v_n \mathcal{R}_a v_{n+1}$ for some $a \in E$ and $u_n Z^n v_n\}$;

$Z = \bigcup_{n\in\mathbb{N}} Z^n$.

We are going to show that Z is a bisimulation.

1. We first show that Z is an epistemic bisimulation.

   We prove by induction on n that every Z^n is an epistemic bisimulation.

   We have already proved that Z^0 is an epistemic bisimulation. Assume it is true for Z^n and $u_{n+1} Z^{n+1} v_{n+1}$. Then there are u_n, v_n such that $u_n Z^n v_n$, $u_n \mathcal{R}_a u_{n+1}$ and $v_n \mathcal{R}_a v_{n+1}$.

   (a) $u_n \in V(p)$ iff $v_n \in V(p)$ because Z^n is an epistemic bisimulation. So $u_{n+1} \in V(p)$ iff $v_{n+1} \in V(p)$ by Definition 6.3.1 (1).

   (b) Assume $u'_{n+1} \in R_j(u_{n+1})$. Then by nf, there are u'_n and b such that $u'_n \in R_j(u_n)$ and $u'_{n+1} \in \mathcal{R}_b(u'_n)$.
   Then there is $v'_n \in W$ such that $v'_n \in R_j(v_n)$ and $v'_n Z^n u'_n$ by induction hypothesis. But $M, u'_n \models Pre(b)$ and besides for all $\varphi \in \mathcal{L}^C$, $M, v'_n \models \varphi$ iff $M, u'_n \models \varphi$ because Z^n is an epistemic bisimulation by induction hypothesis. So $M, v'_n \models Pre(b)$.
   Then there is v'_{n+1} such that $v'_{n+1} \in \mathcal{R}_b(v'_n)$ by Definition 6.3.1 (2). So $v'_{n+1} \in R_j \circ \mathcal{R}_b(v_n)$.
   Besides $M, u_n \models \hat{B}_j Pre(b)$, so $M, v_n \models \hat{B}_j Pre(b)$ by induction hypothesis and because $\hat{B}_j Pre(b) \in \mathcal{L}^C$. So $M, v_n \models [a]\hat{B}_j\langle b^-\rangle\top$ by Definition 6.3.1 (4).
   But $M, v_n \models \langle a\rangle\top$, so $M, v_n \models \langle a\rangle\hat{B}_j\langle b^-\rangle\top$. So $(\mathcal{R}_a \circ R_j \circ \mathcal{R}_b^{-1})(v_n) \neq \emptyset$. So $(R_j \circ \mathcal{R}_b)(v_n) \subseteq (\mathcal{R}_a \circ R_j)(v_n)$ by nl. So there is $v''_{n+1} \in \mathcal{R}_a(v_n)$ such that $v'_{n+1} \in R_j(v''_{n+1})$. Then by ed, $v'_{n+1} \in R_j(v_{n+1})$.
   Besides $u'_n Z^n v'_n$ and $u'_{n+1} \in \mathcal{R}_b(u'_n), v'_{n+1} \in \mathcal{R}_b(v'_n)$.
   So by definition of Z^{n+1}, $u'_{n+1} Z^{n+1} v'_{n+1}$.

   So there is v'_{n+1} such that $v'_{n+1} \in R_j(v_{n+1})$ and $u'_{n+1} Z^{n+1} v'_{n+1}$

   (c) The case $v'_{n+1} \in R_j(v_{n+1})$ is similar.

   So for all $n \in \mathbb{N}$, Z^n is an epistemic bisimulation. Henceforth Z is also a bisimulation.

2. Now we are going to show that Z is a full bisimulation. Assume uZv for some $u, v \in W$. Then $uZ^n v$ for some $n \in \mathbb{N}$.

   (a) If $u' \in \mathcal{R}_a(u)$ then $M, u \models Pre(a)$ by Definition 6.3.1 (2). So $M, v \models Pre(a)$ because Z is an epistemic bisimulation and $Pre(a) \in \mathcal{L}^C$.
   So there is v' such that $v\mathcal{R}_a v'$. But then $u' Z^{n+1} v'$ by construction of Z^n. So $u' Z v'$.

   (b) Similarly we prove that if $v' \in \mathcal{R}_a(v)$ then there is u' such that $u' \in \mathcal{R}_a(u)$ and $u' Z v'$.

QED

Thanks to this lemma, we can now prove that for every formula φ of the language $\mathcal{L}_A$, $\models_{\mathsf{BMS}} \varphi$ if and only if $\Gamma(A) \models_{\mathsf{EDL}} \varphi$. We first prove two Propositions.

Proposition 6.3.4 *Let A be an event model, and let ψ be a formula from $\mathcal{L}_A$. If $\not\models_{BMS} \psi$ then $\Gamma(A) \not\models_{EDL} \psi$.*

PROOF. We have to prove that if there is an epistemic model M^s and a w in M^s such that $M^s, w \models \psi$ then M^s can be turned into an EDL-model M such that $M \models \Gamma(A)$, with a w' in M such that $M, w' \models \psi$.

The proof iteratively applies the product construction to the initial M^s as follows: we set $M^0 = M^s$, and

$$M^{n+1} = M^n \otimes_{\mathsf{EDL}} A = (W^{n+1}, R^{n+1}, \mathcal{R}^{n+1}, V^{n+1}),$$

where

- $W^{n+1} = W^n \cup \{(w, a) \mid w \in W^n \text{ and } M^s, w \models Pre(a)\}$;
- $R_j^{n+1} = R_j^n \cup \{((w_1, a_1), (w_2, a_2)) \mid w_1 R_j^n w_2 \text{ and } a_1 R_j a_2\}$;
- $\mathcal{R}_a^{n+1} = \mathcal{R}_a^n \cup \{(w, (w, a)) \mid w \in W^n\}$;
- $V^{n+1}(p) = V^n(p) \cup \{(w, a) \mid w \in W^n \text{ and } w \in V^n(p)\}$.

Note that we use $\otimes_{\mathsf{EDL}}$ to distinguish our product construction here from the BMS product that we write $\otimes_{\mathsf{BMS}}$ from now on to avoid confusion. Finally, we set $M^\infty = (W^\infty, R^\infty, \mathcal{R}^\infty, V^\infty)$, where $W^\infty = \bigcup_n W^n$, $V^\infty(p) = \bigcup_n V^n(p)$, $\mathcal{R}_a^\infty = \bigcup_n \mathcal{R}_a^n$, and $R_j^\infty = \bigcup_n R_j^n$.[2] We are going to prove that $M^\infty, w \models \varphi$. Then we will show that $M^\infty \models \Gamma(A)$. First we prove a lemma:

> **Lemma 6.3.5** *Let $k \geq 0$. $(M^s \otimes_{BMS} A)^k, (w, a) \leftrightarrows M^{k+1}, (w, a)$, where $(M^s \otimes_{BMS} A)^k$ is the result of the iteration process applied k times to the static model $M^s \otimes_{BMS} A$ and the event model A.*
>
> PROOF. We prove it by induction on k.
>
> $k = 0$: $(M^s \otimes_{\mathsf{BMS}} A)^0 = M^s \otimes_{\mathsf{BMS}} A$ and $M^1 = M^s \otimes_{\mathsf{EDL}} A$. Then by definition of $\otimes_{\mathsf{EDL}}$, we clearly have $(M^s \otimes_{\mathsf{BMS}} A)^0, (w, a) \leftrightarrows M^1, (w, a)$
>
> $k + 1$: $(M^s \otimes_{\mathsf{BMS}} A)^{k+1} = (M^s \otimes_{\mathsf{BMS}} A)^k \otimes_{\mathsf{EDL}} A$. Now $(M^s \otimes_{\mathsf{BMS}} A)^k, (w, a) \leftrightarrows M^{k+1}, (w, a)$ by induction hypothesis. So $(M^s \otimes_{\mathsf{BMS}} A)^k \otimes_{\mathsf{EDL}} A, (w, a) \leftrightarrows M^{k+1} \otimes_{\mathsf{EDL}} A, (w, a)$ because for any M, M' if $M, w \leftrightarrows M', w'$ then $M \otimes_{\mathsf{EDL}} A, w \leftrightarrows M' \otimes_{\mathsf{EDL}} A, w'$. Then $(M^s \otimes_{\mathsf{BMS}} A)^{k+1}, (w, a) \leftrightarrows M^{k+2}, (w, a)$.
>
> QED

Now we prove a second lemma:

> **Lemma 6.3.6** *For all $\varphi \in \mathcal{L}_A$, $M^s, w \models_{BMS} \varphi$ iff $M^\infty, w \models_{EDL} \varphi$*

[2] Note that is just as Yap's construction [Yap, 2006].

PROOF. For any formula φ we define the integer $\delta(\varphi)$ as the *maximum number of nested event operator occurrences* as follows:

- $\delta(p) = 0$
- $\delta(\varphi_1 \wedge \varphi_2) = \max(\delta(\varphi_1), \delta(\varphi_2)))$
- $\delta(\neg\varphi) = \delta(B_i\varphi) = \delta(\varphi)$
- $\delta([a]\varphi) = \delta([a^-]\varphi) = \delta(\varphi) + 1$

We set $\mathcal{P}(k)$: "For all $\varphi \in \mathcal{L}_A$ such that $\delta(\varphi) = k$, $M^s, w \models_{\mathsf{BMS}} \varphi$ iff $M^k, w \models_{\mathsf{EDL}} \varphi$", where M^s is the static model and M^k is the iteration of the product construction.

We prove $\mathcal{P}(k)$ for all k by induction on k.

$k = 0$: Then φ is epistemic so it works by definition of $\otimes_{\mathsf{EDL}}$.

$k + 1$: We prove it by induction on φ.

- $\varphi = [a]\varphi'$.
  $M^s, w \models_{\mathsf{BMS}} [a]\varphi'$
  iff if $M^s, w \models_{\mathsf{BMS}} Pre(a)$ then $M^s \otimes_{\mathsf{BMS}} A, (w, a) \models_{\mathsf{BMS}} \varphi'$
  iff if $M^s, w \models_{\mathsf{BMS}} Pre(a)$ then $(M^s \otimes_{\mathsf{BMS}} A)^k, (w, a) \models \varphi'$ by Induction Hypothesis because $\delta(\varphi') \leq k$,
  iff if $M^s, w \models_{\mathsf{BMS}} Pre(a)$ then $M^{k+1}, (w, a) \models_{\mathsf{EDL}} \varphi'$ by Lemma 6.3.5
  iff if $M^{k+1}, w \models_{\mathsf{EDL}} Pre(a)$ then $M^{k+1}, (w, a) \models_{\mathsf{EDL}} \varphi'$
  iff $M^{k+1}, w \models_{\mathsf{EDL}} [a]\varphi'$ by definition of $\otimes_{\mathsf{EDL}}$
  iff $M^{k+1}, w \models_{\mathsf{EDL}} \varphi$.
- $\varphi = \varphi_1 \wedge \varphi_2$ works by Induction Hypothesis.
- $\varphi = B_j\varphi'$ works as well.
- $\varphi = p$ is impossible because $k + 1 \geq 1$.

QED

Then we can easily prove that for all φ such that $\delta(\varphi) = k$, $M^k, w \models_{\mathsf{EDL}} \varphi$ iff $M^\infty, w \models_{\mathsf{EDL}} \varphi$. Then for all k, for all φ such that $\delta(\varphi) = k$, $M^s, w \models_{\mathsf{BMS}} \varphi$ iff $M^\infty, w \models_{\mathsf{EDL}} \varphi$

i.e. for all $\varphi \in \mathcal{L}_A$, $M^s, w \models_{\mathsf{BMS}} \varphi$ iff $M^\infty, w \models_{\mathsf{EDL}} \varphi$. In particular, because $M^s, w \models_{\mathsf{BMS}} \psi$, we have $M^\infty, w \models_{\mathsf{EDL}} \psi$.

It remains to prove that $M^\infty \models_{\mathsf{EDL}} \Gamma(A)$. Conditions (1) and (2) of Definition 6.3.1 are clearly fulfilled. As for condition (3), let $w \in W^\infty$, w' is such that $w\mathcal{R}_a w'$ iff $w' = (w, a)$. Now $(w, a)R_j u$ iff $u = (v, b)$ with $wR_j v$ and $aR_j b$ by definition of $\otimes_{\mathsf{EDL}}$. So for all u such that $(w, a)R_j u$, there are b and v such that $aR_j b$ and $v\mathcal{R}_b u$. This proves that $M^\infty, w \models_{\mathsf{EDL}} [a]B_j(\langle a_1^-\rangle\top \vee \ldots \vee \langle a_n^-\rangle\top)$ where $a_1, \ldots, a_n$ is the list of all b such that $aR_j b$. Finally, concerning condition (4), assume $M^\infty, w \models_{\mathsf{EDL}} \hat{B}_j Pre(b)$ and $w\mathcal{R}_a(w, a)$. Then there is v such that $wR_j v$ and $v\mathcal{R}_b(v, b)$. So by definition of $\otimes_{\mathsf{EDL}}$, because $aR_j b$, we have $(w, a)R_j(v, b)$. Hence $M^\infty, (w, a) \models_{\mathsf{EDL}} \hat{B}_j\langle b^-\rangle\top$ and finally $M^\infty, w \models_{\mathsf{EDL}} [a]\hat{B}_j\langle b^-\rangle\top$. QED

Proposition 6.3.7 *Let A be an event model, and let ψ be a formula from $\mathcal{L}_A$. If $\models_{\mathsf{BMS}} \psi$ then $\Gamma(A) \models_{\mathsf{EDL}} \psi$.*

PROOF. We take advantage of the complete axiomatization of BMS-validities given in [Baltag *et al.*, 1998; Baltag and Moss, 2004], and show that the BMS-axioms are EDL-valid, and that the BMS-inference rules preserve EDL-validity. As the inference rules of BMS and EDL are identical (i.e. modus ponens and necessitation) it is clear that the BMS-inference rules preserve EDL-theoremhood. It is straightforward to show that every instance of the BMS-axioms not involving dynamic operators is EDL-valid. So what remains is to prove that the BMS schemas

$$\begin{array}{ll} \mathsf{R1} & [a]p \leftrightarrow (Pre(a) \to p) \\ \mathsf{R2} & [a]\neg\varphi \leftrightarrow (Pre(a) \to \neg[a]\varphi) \\ \mathsf{R3} & [a]B_j\varphi \leftrightarrow (Pre(a) \to B_j[a_1]\varphi \wedge \ldots \wedge B_j[a_n]\varphi) \end{array}$$

where $a_1, \ldots, a_n$ is the list of all b such that aR_jb, are logical consequences of $\Gamma(A)$ in EDL.

R1 Axiom R1 can be proved by the nonlogical axioms (1) $p \to [a]p$ and (2) $\langle a\rangle\top \leftrightarrow Pre(a)$ of the theory $\Gamma(A)$ in Definition 6.3.1.

R2 For the left-to-right direction of R2 we have

$$\Gamma(A) \models_{\mathsf{EDL}} ([a]\neg\varphi \wedge Pre(a) \wedge [a]\varphi) \to \bot$$

because of the nonlogical axiom (2) $\langle a\rangle\top \leftrightarrow Pre(a)$ of Definition 6.3.1.

For the right-to-left direction, on the one hand we have $\Gamma(A) \models_{\mathsf{EDL}} \neg Pre(a) \to [a]\bot$ again by the nonlogical axiom (2) of Definition 6.3.1, and on the other hand $\Gamma(A) \models_{\mathsf{EDL}} \neg[a]\varphi \to [a]\neg\varphi$ by Proposition 6.3.3.

R3 For the left-to-right direction of R3, let M be an EDL-model such that $M \models_{\mathsf{EDL}} \Gamma(A)$ and suppose

$$M, w \models_{\mathsf{EDL}} [a]B_j\varphi \wedge Pre(a),$$

and suppose $M, w \models_{\mathsf{EDL}} \neg B_j[b]\varphi$ for some b such that aR_jb. So there must exist worlds w' and v' such that wR_jw', $w'\mathcal{R}_bv'$ and $M, v' \models \neg\varphi$. Therefore $M, w' \models Pre(b)$ by nonlogical axiom 6.3.1 (2), and $M, w \models_{\mathsf{EDL}} \hat{B}_j Pre(b)$. As aR_jb, our nonlogical axiom 6.3.1 (4) tells us that $M, w \models_{\mathsf{EDL}} \hat{B}_j Pre(b) \to [a]\hat{B}_j\langle b^-\rangle\top$, and hence $M, w \models_{\mathsf{EDL}} [a]\hat{B}_j\langle b^-\rangle\top$. As by hypothesis $M, w \models_{\mathsf{EDL}} Pre(a)$, by nonlogical axiom 6.3.1 (2) $(\mathcal{R}_a \circ R_j \circ \mathcal{R}_b^{-1})(w) \neq \emptyset$. By the constraint nl on EDL-models we have

$$(R_j \circ \mathcal{R}_b)(w) \subseteq (\mathcal{R}_a \circ R_j)(w),$$

i.e. $v' \in (\mathcal{R}_a \circ R_j)(w)$. As we have supposed that $M, w \models_{\mathsf{EDL}} [a]B_j\varphi$, we must have $M, v' \models_{\mathsf{EDL}} \varphi$, which is contradictory.

For the right-to-left direction of R3, we know that $\Gamma(A) \models_{\mathsf{EDL}} \neg Pre(a) \to [a]\bot$ again by the nonlogical axiom 6.3.1 (2), so it remains to prove that

$$\Gamma(A) \models_{\mathsf{EDL}} (B_j[a_1]\varphi \wedge \ldots \wedge B_j[a_n]\varphi) \to [a]B_j\varphi. (*)$$

where $a_1, \ldots, a_n$ is the list of all b such that aR_jb.

Suppose $M, w \models_{\mathsf{EDL}} B_j[a_1]\varphi \wedge \ldots \wedge B_j[a_n]\varphi$, and suppose $M, w \models_{\mathsf{EDL}} \neg[a]B_j\varphi$. The latter implies that there are worlds v and v' such that $w\mathcal{R}_avR_jv'$ and $M, v' \models_{\mathsf{EDL}} \neg\varphi$. By the constraint nf, there is $b \in E$ such that $v' \in R_j \circ \mathcal{R}_b(w)$.

Now, by the nonlogical axiom 6.3.1 (3) we have
$$[a]B_j\left((\langle a_1^-\rangle\top \vee \ldots \vee \langle a_n^-\rangle\top) \wedge ([b_1^-]\,\bot \wedge \ldots \wedge [b_n^-]\,\bot)\right),$$
where $a_1, \ldots, a_n$ is the list of all b such that $b \in R_j(a)$ and
$b_1, \ldots, b_n$ is the list of all b such that $b \notin R_j(a)$.
Hence $M, v' \models_{\mathsf{EDL}} (\langle a_1^-\rangle\top \vee \ldots \vee \langle a_n^-\rangle\top) \wedge ([b_1^-]\,\bot \wedge \ldots \wedge [b_n^-]\,\bot)$. So $b \in R_j(a)$. Then $M, w \models_{\mathsf{EDL}} B_j[b]\varphi$ by (*). So $M, v' \models_{\mathsf{EDL}} \neg\varphi$, which is contradictory.

QED

Putting these two results together we obtain the following key result:

Theorem 6.3.8 *Let A be an event model. Let φ be a formula from $\mathcal{L}_A$. Then*

$$\models_{\mathsf{BMS}} \varphi \quad \textit{iff} \quad \Gamma(A) \models_{\mathsf{EDL}} \varphi$$

It follows that

$$\vdash_{\mathsf{BMS}} \varphi \quad \text{iff} \quad \Gamma(A) \vdash_{\mathsf{EDL}} \varphi$$

This thus provides a new axiomatization of BMS-validities. This new axiomatization is just made of $\Gamma(A)$ together with the axiomatization of EDL .

Remark 6.3.9 In [Aucher and Herzig, 2007], the constraint of no-forgetting and condition (3) of Definition 6.3.1 were replaced by the following ones

nf′ if $v(\mathcal{R}_a \circ R_i \circ \mathcal{R}_b^{-1})v'$ then vR_iv'

(3)′ $\vdash_{\mathsf{BMS}} [a]B_i\varphi \leftrightarrow (Pre(a) \rightarrow B_i[a_1]\varphi \wedge \ldots \wedge B_i[a_n]\varphi)$
where $a_1, \ldots, a_n$ is the list of all b such that $b \in R_j(a)$.

Neither do EDL models satisfy nf′, nor the other way round. Hence the version of EDL in [Aucher and Herzig, 2007] cannot be compared with our present version. If we moreover assume that event models are serial then we obtain the same results as here. Here we do not need this last assumption and our condition (3) describes more accurately than (3)′ the structure of event models. Our constraint nf is also a better generalization of the principle of perfect-recall than nf′. □

6.4 Conclusion and related work

We have presented an epistemic dynamic logic EDL whose semantics differs from the BMS semantics. We have shown that BMS can be embedded into EDL. This result allows to conclude that EDL is an interesting alternative to Baltag et al.'s logic, that allows to talk about agents' perception of events just in the same way as BMS does. However, EDL is more expressive than BMS because it allows to talk about past events. Another of its advantages is

that EDL allows for incomplete beliefs about the event taking place and can still draw inferences from this incomplete description of the event, while in BMS the event model has to specify everything. So in a sense EDL seems more versatile than BMS to describe events.

On the other hand, the power of event models is not completely exploited in the BMS approach. Indeed, as we said in Chapter 3, the philosophy of the BMS approach is to represent events in the same way as situations are represented in epistemic logic by means of epistemic models. But unlike an epistemic model, an event model does not have a genuine valuation to describe possible events. An obvious extension of the BMS formalism would be to add a valuation to event models in order to describe possible events more precisely. Then we could define epistemic languages for event models completely identical to the various epistemic languages we already defined for epistemic models, except that the propositional letters of these languages would describe possible events instead of possible worlds. This would allow to express things about events that are *currently* taking place, and not only to express things before or after the occurrence of events as in EDL. This would also allow to update/revise events by other events which is a phenomenon that often occurs in everyday life.[3] It is not possible to model such phenomena in EDL because the accessibility relations for events are set once and for all.

Another approach studying information change over time is Epistemic Temporal Logic ETL [Parikh and Ramanujam, 2003] (or equivalently interpreted systems [Fagin *et al.*, 1995]). In this approach, the notion of belief is still present but the notion of event is replaced by the notion of time, and their models are tree-like models representing the possible evolutions of a situation over time. So their models are somehow similar to our EDL models in the sense that the specifications of time (instead of events) and beliefs are on the same formal level (unlike the BMS formalism). The connection of ETL with the BMS formalism is made by van Benthem, Pacuit and Liu in [van Benthem and Pacuit, 2006] and [van Benthem and Liu, 2004] but more especially with Gerbrandy in [van Benthem *et al.*, 2007]. In this last paper, they also introduced converse events and independently proposed for their tree-like models perfect-recall and no-miracle principles. Their perfect-recall principle corresponds to our constraint nf′ in Remark 6.3.9, and their no-miracle principle is $\neg C_G \neg\langle a\rangle \hat{B}_j \langle b^-\rangle \top \to ([a]B_j\varphi \to B_j[b]\varphi)$ which is almost identical to our no-learning principle. Another approach is that of [van Ditmarsch *et al.*, 2007a] where the authors show how to translate a BMS formula satisfied in an epistemic model into an ETL formula satisfied in an interpreted system. So their approach is less general than ours because it only deals with the model checking problem. Still in

[3] For example, assume that Bob wants to know whether the coin is heads or tails up and starts to open the box to look at the coin, Ann suspecting nothing about it, and Bob knowing that. This can be modelled by a first event model. However, while Bob opens the box to look at the coin Ann notices him but Bob does not notice that Ann has noticed him and he still believes that she did not notice anything. This second event, temporally included in the first event, can also be modelled by a second event model. However its preconditions deal with what is true in the first event model and thus are expressed in the language of the first event model (the precondition for Ann observing that Bob cheats is that Bob does cheat, which is expressible in the language of the first event model). Then we could update the first event model by the second completely similarly to the way we already update situations by events in the BMS formalism, except that the preconditions would be expressed in the language of the first event model. This would yield a third event model in which Bob is cheating and Ann knows it but Bob believes that Ann does not know it.

the ETL paradigm but starting from the BMS formalism, Yap [Yap, 2006] and Sack [Sack, 2008; Sack, 2007] introduce a 'yesterday' temporal modal operator to the BMS language expressing what was true before the last event; Sack gets a complete characterization. To prove completeness Sack [Sack, 2007] also introduces a separate component expressing that an event just occurred but this is not a converse *modal* operator like ours. However he does introduce a converse modal operator for public announcement logic but does not provide a completeness proof for it [Sack, 2008].

Another approach embedding the BMS formalism to a formalism that also deals with events and beliefs on the same formal level is proposed by van Eijck et col. in [van Eijck, 2004; van Benthem *et al.*, 2006b]. They map the BMS formalism to (epistemic) propositional dynamic logic (refining a similar result for *automata* propositional dynamic logic [van Benthem and Kooi, 2004]). However they do not resort to converse events and translate directly event models into a transformation on PDL programs.

Chapter 7

Conclusion and further research

In this thesis, we have investigated several logical models of belief change and belief representation by stressing the importance of choosing a modeling point of view. In that respect we first identified the three possible modeling points of view, proceeding by successive dichotomies: the internal, perfect external and imperfect external approaches.

From Chapter 2 to 4, we focused on the internal approach. In Chapter 2 we provided an internal version of epistemic logic by introducing the notions of multi-agent possible world and internal model and proved a completeness result for this semantics. In Chapter 3 we then added dynamics to our internal approach and proposed an internal version of dynamic epistemic logic as viewed by BMS. We also studied in which case seriality of accessibility relations is preserved during an update. In Chapter 4 we first showed that generalizing belief revision theory to a multi-agent setting amounts to study private announcement. Then we proposed a way to deal with belief revision when the private announcement is incoherent with the agent's beliefs by generalizing AGM belief revision theory to the multi-agent case. Finally, we provided an example of a revision operation based on a degree of similarity between multi-agent possible worlds and applied it to our 'coin' example.

However, it still remains to show that the internal logic Int is $PSPACE$-complete for $N = 2$ and provide a complete axiomatization for the second notion of validity. But the main open problem for the internal approach remains to find constructive revision mechanisms for any kind of event, and not only for private announcement. This would enable artificial agents to cope with any kind of unexpected event. Besides, this would also indirectly enable us to update and revise external models by any kind of event, as we said in Section 4.5. In fact the formal asymmetry between the formalisms that we proposed for the perfect external and internal approach would be resolved if we could resort in the formalisms for the (perfect) external approach to the revision mechanisms designed for the internal approach. Finally, another line of research would be to define the internal version of other (external) logics than epistemic logic such as probabilistic logic, possibilistic logic... in order to deal more accurately with the representation of uncertainty and ignorance.

In Chapter 5 we followed the (perfect) external approach and introduced a rich formal-

ism using probabilities and hyperreal numbers. This enabled us to model accurately epistemic states of human agents including what would surprise them, and also to model accurately how human agents interpret events and revise their beliefs. We also reviewed different approaches to the notion of knowledge and showed how some of them could be captured in our formalism. So our formalism is general in the sense that it covers many aspects of belief change and belief (and knowledge) representation.

Two issues related to our external approach deserve further research. First, our formalism still suffers from the logical omniscience problem which is a real problem for the kind of applications we have in mind (particularly in psychology). An improvement of our formalism would be to address this issue. Second, the notion of surprising world needs conceptual and theoretical foundation. As this touches the core of our ideas, we will address that matter below, separately.

Finally, in Chapter 6, we also followed the (perfect) external approach and enriched the dynamic epistemic language with a converse operator. We then proposed a logic EDL which specifies more explicitly than BMS the interactions between events and beliefs by means of axioms. Finally we showed that we can embed BMS into EDL by translating the structure of an event model as a non-logical EDL-theory. A natural continuation of this work could be to study the decidability of EDL or to add to the language a common belief operator.

Finally, let us have a closer look at the notion of surprising world that we introduced in Chapter 5 to perform belief revision. The ontological status of these surprising worlds is subject to debate. Indeed, we assumed that the agent is not aware of them. So are they somewhere in her mind but below a certain 'activation threshold' for the agent to be aware of them? In that case, there would be a problem if the overall number of possible worlds was infinite, which is the case if there is an infinite number of propositional letters or if we are in a multi-agent setting. Indeed, from a psychological point of view it is difficult to accept that human agents can have infinite structures in their mind. Even if we assumed that the agent only has a finite number of them in her mind we would still have to give reasons why some are in her mind and some are not. In particular, we would need to motivate philosophically why the possible world corresponding to the actual world has always to be in the agent's mind (in other words, why the actual world has to be a conceived world or a surprising world). Indeed, the agent could have to revise by formulas which are true only in the actual world and we would face a technical problem if this actual world was not in the model. So, are the surprising worlds absent from the agent's mind and created by the agent only when she needs to revise her beliefs? But in order to create these new worlds, the agent would need means to do so. This could be for example thanks to a distance (or equivalently a degree of similarity) between (multi-agent) possible worlds like in Section 4.4. Indeed, this distance can easily be present in the agent's mind and orders implicitly all the (multi-agent) possible worlds, even if all of them are not actually present in the agent's mind. It seems that the surprising worlds are just a way to bypass a deeper problem which consists in defining a generic object present in the agent's mind like a distance which would implicitly order all the (multi-agent) possible worlds, like in the internal approach. The connection of the external approach with the internal approach is even more salient when viewed from this perspective. In our formalism this ordering of possible worlds is given from the start by

the probability values of these surprising worlds. In a sense, one could even consider these surprising worlds as a technical 'trick' (even if they indirectly also allow to model an aspect of the notion of surprise). So it seems that the real issue, even for the external approach, is to find constructive revision mechanisms for the internal approach in order to cope with any kind of unexpected event (and not only for the case of private announcement as we did).

Abstract

Representing an epistemic situation involving several agents depends very much on the modeling point of view one takes. For example, in a poker game the representation of the game will be quite different whether the modeler is a poker player playing in the game or the card dealer who knows perfectly what the players' cards are. One of the main contributions of this thesis is to systematically distinguish the different modeling points of view and their respective formalisms. Classically, in epistemic logic, the modeler is somebody external to the situation who has a perfect knowledge of it (like the card dealer). We call this modeling approach the *external approach*. Another possibility is that the modeler is an agent involved in the situation who interacts with the other agents (like the poker player). We call this modeling approach the *internal approach*. In this thesis, we focus on these two modeling approaches.

The internal approach has not been studied so far in the logical literature. So we first propose an internal version of epistemic logic (that we axiomatize) and we set some formal links between the internal and external approaches. Then we add change to the picture and propose an internal version of the BMS dynamic epistemic logic very much in the same spirit as we did for epistemic logic. Doing so we also provide conditions under which seriality of models is preserved during an update. This logical study of the internal approach and its link with the external approach is another contribution of this thesis.

Then we show how this new internal approach allows for a straightforward generalization of AGM belief revision theory to a multi-agent setting. We first observe, thanks to our internal version of dynamic epistemic logic, that generalizing AGM belief revision theory to a multi-agent setting amounts to study private announcement. Then we generalize the theorems of AGM theory to the multi-agent case. Afterwards we propose rationality postulates in the AGM style in order to better specify formally that we study private announcement. Finally we provide an example of revision operation satisfying one of these postulates. Generalizing AGM belief revision theory to a multi-agent setting is also one of the main contributions of this thesis.

Afterwards, we turn our attention to the external approach for which we propose a rich and general formalism based on the BMS one. This rich formalism uses probability to model degrees of belief, and infinitesimals to model degrees of potential surprise. This allows to model accurately how human agents interpret events and revise their beliefs. Formalizing these intricate logical dynamics and providing an expressive framework to model accurately epistemic states of agents is another contribution of this thesis. Finally we review various axioms proposed to characterize the notion of knowledge and its interaction with the notion of belief, and show how they can be captured in our rich formalism.

Eventually, we propose an alternative to the BMS formalism for the external approach where events are simply represented as accessibility relations between possible worlds, unlike the BMS formalism where they are represented as event

models. This allows to define easily a converse event modal operator and to specify more explicitly the interactions between events and beliefs by means of constraints on the accessibility relations for beliefs and events. Our contribution here is to propose such constraints and to show that the BMS formalism can be embedded in ours by translating the structure of an event model as a non-logical theory.

Résumé

La modélisation d'une situation épistémique qui fait intervenir plusieurs agents dépend beaucoup du point de vue que l'on adopte vis à vis de la situation. Par exemple, dans un jeu de poker, la représentation du jeu sera bien différente selon que le modélisateur est l'un des joueurs ou bien celui qui distribue les cartes et qui sait parfaitement quelles sont les cartes que les joueurs possèdent. Une contribution essentielle de cette thèse est de distinguer clairement ces différents points de vue de modélisation et leurs formalismes respectifs. En logique épistémique, le modélisateur est quelqu'un d'externe à la situation et qui en a une connaissance parfaite (comme la personne qui distribue les cartes dans le jeu de poker). Nous appelons cette approche l'*approche externe*. Une autre possibilité est celle où le modélisateur est l'un des agents qui interagit avec les autres agents (comme l'un des joueurs de poker). Nous appelons cette approche l'*approche interne*. Dans cette thèse, nous nous focalisons sur ces deux approches.

L'approche interne n'a pas été étudiée en logique jusqu'à maintenant. Dans un premier temps nous proposons une version interne de la logique épistémique (que l'on axiomatise) et nous établissons des liens formels entre les approches externes et internes. Ensuite nous introduisons du dynamisme à notre formalisme et proposons une version interne de la logique épistémique dynamique de BMS dans le même esprit que la version interne de la logique épistémique.

Ensuite nous montrons que cette nouvelle approche interne permet de généraliser de façon immédiate la théorie de révision des croyances au cas multi-agent. On observe d'abord, grâce à notre version interne de la logique épistémique dynamique, que généraliser la théorie de révision des croyances d'AGM au cas multi-agent revient à étudier les annonces privées. Ensuite on généralise les théorèmes de la théorie d'AGM au cas multi-agent. Après cela, nous proposons des postulats de rationalité dans le style d'AGM afin de mieux spécifier formellement que nous étudions les annonces privées. Finalement nous proposons un exemple d'opération de révision qui satisfait un de ces postulats. Généraliser la théorie de révision des croyances au cas multi-agent est aussi une des contributions principales de cette thèse.

Ensuite nous nous concentrons sur l'approche externe pour laquelle nous proposons un formalisme riche et général basé sur celui de BMS. Ce riche formalisme utilise les probabilités pour modéliser les degrés de croyance, et les infinitésimaux pour modéliser les degrés de surprise potentielle. Cela permet de modéliser précisément comment des agents humains interprètent les événements et révisent leurs croyances. Formaliser ces mechanismes logiques complexes et fournir un système expressif pour modéliser avec précision les états epistémiques des agents est une autre contribution de cette thèse. Finalement nous passons en revue différents axiomes proposés dans la littérature pour caractériser la notion de connaissance et sa relation avec la notion de croyance, et montrons comment ceux-ci peuvent être capturés dans notre riche formalisme.

Finalement, nous proposons une alternative au formalisme de BMS (pour

l'approche externe) où les événements sont simplement représentés par des relations d'accessibilité entre mondes possibles, à la différence du formalisme de BMS où ils sont représentés par des modèles d'événements. Cela permet de définir facilement un opérateur modal d'événement inverse et de spécifier plus explicitement les intéractions entre croyances et événements à l'aide de contraintes sur les relations d'accessibilité respectives. Notre contribution est ici de proposer de telles contraintes et de montrer que le formalisme de BMS peut être plongé dans le notre en traduisant la structure d'un modèle d'événement en une théorie non-logique.

Bibliography

[Adams, 1975] Ernest Adams. *The Logic of Conditionals*, volume 86 of *Synthese Library*. Springer, 1975. 94

[Alchourrón *et al.*, 1985] Carlos Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50(2):510–530, 1985. 2, 18, 57

[Arló Costa and Levi, 1996] Horacio Arló Costa and Isaac Levi. Two notions of epistemic validity (epistemic models for Ramsey's conditionals). *Synthese*, 109:217–262, 1996. 25

[Aucher and Herzig, 2007] Guillaume Aucher and Andreas Herzig. From DEL to EDL: exploring the power of converse events. In *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU), Hammamet*, volume 4724 of *LNCS*, pages 199–209. Springer, 2007. 4, 138

[Aucher, 12 16 May 2008] Guillaume Aucher. Internal models and private multi-agent belief revision. In Muller Padgham, Parkes and Parsons, editors, *Proceedings of Autonomous Agents and Multi-agent Systems (AAMAS 2008)*, pages 721–727, Estoril, Portugal, 12-16 May 2008. 4

[Aucher, 2004] Guillaume Aucher. A combined system for update logic and belief revision. In Mike Barley and Nikola K. Kasabov, editors, *PRIMA 2004*, volume 3371 of *LNCS*, pages 1–17. Springer, 2004. Revised Selected Papers. 103, 115

[Aucher, 2007] Guillaume Aucher. Interpreting an action from what we perceive and what we expect. *Journal of Applied Non-Classical Logics*, 17(1):9–38, 2007. 4

[Aumann, 1976] Ronald Aumann. Agreeing to disagree. *Annals of Statistics*, 4(6):1236–1239, 1976. 12

[Aumann, 1977] Ronald Aumann. Game theory. In J. Eatwell, M. Milgate, and P. Newman, editors, *Game Theory*, The New Palgrave, pages 1–54, Macmillan, 1977. 12

[Ayer, 1956] Alfred Jules Ayer. *The Problem of Knowledge*. Penguin books, London, 1956. 110

[Bacchus *et al.*, 1999] Fahiem Bacchus, Joseph Halpern, and Hector Levesque. Reasoning about noisy sensors and effectors in the situation calculus. *Artificial Intelligence*, 111(1-2):171–208, July 1999. 94, 124

[Balbiani and Herzig, 2007] Philippe Balbiani and Andreas Herzig. Talkin'bout Kripke models. In Torben Braüner and Jørgen Villadsen, editors, *International Workshop on Hybrid Logic 2007 (Hylo 2007)*, Dublin, 06/08/07-10/08/07 2007. 78

[Balbiani *et al.*, 2007] Philippe Balbiani, Alexandru Baltag, Hans Van Ditmarsch, Andreas Herzig, Tomohiro Hoshi, and Tiago Santos De Lima. What can we achieve by arbitrary announcements? A dynamic take on Fitch's knowability. In Dov Samet, editor, *Theoretical Aspects of Rationality and Knowledge (TARK), Brussels, 25-JUN-07-27-JUN-07*, pages 42–51, http://www.uclouvain.be/pul.html, 2007. Presses universitaires de Louvain. 113

[Balbiani, 2007] Philippe Balbiani. Manuscript, unpublished. 2007. 79

[Baltag and Moss, 2004] Alexandru Baltag and Larry Moss. Logic for epistemic programs. *Synthese*, 139(2):165–224, 2004. 9, 17, 35, 38, 40, 94, 103, 137

[Baltag and Smets, 2007] Alexandru Baltag and Sonja Smets. Probabilistic dynamic belief revision. In Johan van Benthem, Shier Ju, and Frank Veltman, editors, *A Meeting of the Minds: Proceedings of the Workshop on Logic, Rationality and Interaction*, volume 8 of *Computing Series*, London, 2007. College Publications. 124

[Baltag *et al.*, 1998] Alexandru Baltag, Larry Moss, and Slawomir Solecki. The logic of common knowledge, public announcement, and private suspicions. In I. Gilboa, editor, *Proceedings of the 7th conference on theoretical aspects of rationality and knowledge (TARK98)*, pages 43–56, 1998. 2, 35, 137

[Baltag *et al.*, 1999] Alexandru Baltag, Larry Moss, and Slawomir Solecki. The logic of public announcements, common knowledge and private suspicions. Technical report, Indiana University, 1999. 55

[Barwise and Moss, 1997] John Barwise and Larry Moss. *Vicious Circles*. CSLI Publications, Stanford, 1997. 15

[Battigalli and Bonanno, 1999] Pierpaolo Battigalli and Giacomo Bonanno. Recent results on belief, knowledge and the epistemic foundations of game theory. *Research in Economics*, 53:149–225, 1999. 5, 17

[Ben-Naim, 2006] Jonathan Ben-Naim. Lack of finite characterizations for the distance-based revision. In *Proceedings of Knowledge Representation (KR 2006)*, pages 239–248, 2006. 77

[Benferhat *et al.*, 2002] Salem Benferhat, Didier Dubois, Henri Prade, and Mary-Anne Williams. A practical approach to revising prioritized knowledge bases. *Studia Logica*, 70(1):105–130, 2002. 57

[Blackburn *et al.*, 2001] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*, volume 53 of *Cambridge Tracts in Computer Science*. Cambridge University Press, 2001. 5, 9, 10, 12, 14, 15, 25, 78

[Blume *et al.*, 1991] Larry Blume, Adam Brandenburger, and Eddie Dekel. Lexicographic probabilities and choice under uncertainty. *Econometrica*, 59(1):61–79, JAN 1991. 115

[Boh, 1993] Ivan Boh. *Epistemic Logic in the later Middle Ages*. Routledge, London, 1993. 5

[Bonanno, 1996] Giacomo Bonanno. On the logic of common belief. *Mathematical Logic Quaterly*, 42(3):305–311, 1996. 13

[Booth and Nittka, 2007a] Richard Booth and Alexander Nittka. A method for reasoning about other agents' beliefs from observations. *Texts in Logic and Games*, 2007. accepted for publication. 18

[Booth and Nittka, 2007b] Richard Booth and Alexander Nittka. Reconstructing an agent's epistemic state from observations about its beliefs and non-beliefs. *Journal of Logic and Computation*, 2007. accepted for publication. 18

[Boutilier *et al.*, 1998] Craig Boutilier, Joseph Halpern, and Nir Friedman. Belief revision with unreliable observations. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI 1998)*, pages 127–134, 1998. 124

[Cantwell, 2005] John Cantwell. A formal model of multi-agent belief interaction. *Journal of Logic, Language and Information*, 14(4):397–422, 2005. 10

[Castañeda, 1964] Hector-Neri Castañeda. Review of 'knowledge and belief'. *Journal of Symbolic Logic*, 29:132–134, 1964. 32

[Chopra and Parikh, 1999] Samir Chopra and Rohit Parikh. An inconsistency tolerant model for belief representation and belief revision. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 192–199, 1999. 75

[Cohen and Levesque, 1990] Philip Cohen and Hector Levesque. Intention is choice with commitment. *Artificial intelligence*, 42:213–261, 1990. 19

[Cover and Thomas, 1991] Thomas Cover and Joy Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications. John Wiley & Sons Inc., 1991. 2

[Darwiche and Pearl, 1997] Adnan Darwiche and Judea Pearl. On the logic of iterated belief revision. *Artificial Intelligence*, 89(1-2):1–29, 1997. 57

[de Lima, 2007] Tiago de Lima. *Optimal Methods for Reasoning About Actions and Plans in Multi-agent Systems*. PhD thesis, Université Paul Sabatier, Toulouse, 2007. 130

[Dubois and Fargier, 2004] Didier Dubois and Hélène Fargier. A unified framework for order-of-magnitude confidence relation. In M. Chickering and J. Halpern, editors, *Twentieth Conference in Artificial Intelligence*, pages 138–145, 2004. Banff, Canada. 115

[Dubois and Prade, 1991] Didier Dubois and Henri Prade. Possibilistic logic, preferential model and related issue. In *Proceedings of the 12th International Conference on Artificial Intelligence (IJCAI)*, pages 419–425. Morgan Kaufman, 1991. 115

[Duc, 2001] Ho Ngoc Duc. *Resource-Bounded Reasoning about Knowledge*. PhD thesis, University of Leipzig, 2001. 32

[Fagin and Halpern, 1988] Ronald Fagin and Joseph Halpern. Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34:39–76, 1988. 32, 123

[Fagin and Halpern, 1994] Ronald Fagin and Joseph Halpern. Reasoning about knowledge and probability. *Journal of the ACM (JACM)*, 41(2):340–367, 1994. 123

[Fagin *et al.*, 1995] Ronald Fagin, Joseph Halpern, Yoram Moses, and Moshe Vardi. *Reasoning about knowledge*. MIT Press, 1995. 1, 5, 10, 12, 17, 112, 129, 139

[Gärdenfors and Rott, 1995] Peter Gärdenfors and Hans Rott. *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume Volume 4, Epistemic and temporal reasoning, chapter Belief Revision, pages 35–132. Clarendon Press, Oxford, 1995. 2

[Gärdenfors, 1988] Peter Gärdenfors. *Knowledge in Flux (Modeling the Dynamics of Epistemic States)*. Bradford/MIT Press, Cambridge, Massachusetts, 1988. 2, 19, 25, 58, 59, 64, 116

[Georgeff and Rao, 1991] Michael Georgeff and Anand Rao. Asymmetry thesis and side-effect problems in linear time and branching time intention logics. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence (IJCAI-91)*, pages 498–504, (Sydney, Australia), 1991. 19

[Gerbrandy and Groeneveld, 1997] Jelle Gerbrandy and Willem Groeneveld. Reasoning about information change. *Journal of Logic, Language and Information*, 6:147–196, 1997. 2, 35

[Gerbrandy, 1999] Jelle Gerbrandy. *Bisimulation on Planet Kripke*. ILLC dissertation series, Amsterdam, 1999. 35

[Gettier, 1963] Edmund Gettier. Is justified true belief knowledge? *Analysis*, 25:121–123, 1963. 110

[Grove, 1988] Adam Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17:157–170, 1988. 57, 67

[Halpern and Moses, 1992] Joseph Halpern and Yoram Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54:311–379, 1992. 12

[Halpern, 2003] Joseph Halpern. *Reasoning about Uncertainty*. MIT Press, Cambridge, Massachussetts, 2003. 1

[Harel *et al.*, 2000] David Harel, Dexter Kozen, and Jerzy Tiuryn. *Dynamic Logic*. MIT Press, 2000. 127

[Harsanyi, 1967 1968] John Harsanyi. Games with incomplete information played by Bayesian players. Parts i-iii. *Management Science*, 14:159–182,320–334,486–502, 1967-1968. 10

[Heifetz and Mongin, 2001] Aviad Heifetz and Philippe Mongin. Probability logic for type space. *Games and Economic Behaviour*, 35:31–53, 2001. 10

[Herzig *et al.*, 2000] Andreas Herzig, Jérôme Lang, Dominique Longin, and Thomas Polacsek. A logic for planning under partial observability. In *AAAI/IAAI*, pages 768–773, 2000. 130

[Herzig *et al.*, 2004] Andreas Herzig, Jérôme Lang, and Pierre Marquis. Revision and update in multiagent belief structures. In *5th Conference on Logic and the Foundations of Game and Decision Theory (LOFT6)*, Leipzig, July 2004. 56

[Hintikka, 1962] Jaakko Hintikka. *Knowledge and Belief, An Introduction to the Logic of the Two Notions*. Cornell University Press, Ithaca and London, 1962. 1, 5, 32, 111

[Katsuno and Mendelzon, 1991] Hirofumi Katsuno and Alberto Mendelzon. On the difference between updating a knowledge base and revising it. In *Proceedings of Knowledge Representation*, pages 387–394, 1991. 57, 76

[Katsuno and Mendelzon, 1992] Hirofumi Katsuno and Alberto Mendelzon. Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52(3):263–294, 1992. 65, 66, 72, 76

[Keisler, 1986] H. Jerome Keisler. *Elementary Calculus: an Approach Using Infinitesimals*. Prindle and Weber & Schmidt, 1986. Online edition on the website http://www.math.wisc.edu/ keisler/calc.html. 94

[Kooi, 2003] Barteld Kooi. Probabilistic dynamic epistemic logic. *Journal of Logic, Language and Information*, 12(4):381–408, 2003. 123

[Kooi, 2007] Barteld Kooi. Expressivity and completeness for public update logics via reduction axioms. *Journal of Applied Non-Classical Logics*, 17(2):231–253, 2007. 35, 107, 123

[Kraus and Lehmann, 1988] Sarit Kraus and Daniel Lehmann. Knowledge, belief and time. *Theoretical Computer Science*, 58:155–174, 1988. 111

[Lehmann *et al.*, 2001] Daniel Lehmann, Menachem Magidor, and Karl Schlechta. Distance semantics for belief revision. *Journal of Symbolic Logic*, 66(1):295–317, 2001. 77

[Lehmann, 1996] Daniel Lehmann. Generalized qualitative probability: Savage revisited. In E. Horvitz and F. Jensen, editors, *Twelfth Conference on Uncertainty in Artificial Intelligence*, pages 381–388, Portland, Oregon, August 1996. Morgan Kaufmann. 115

[Lehrer and Paxson, 1969] Keith Lehrer and Thomas Paxson. Knowledge: Undefeated justified true belief. *The Journal of Philosophy*, 66:225–237, 1969. 113

[Lenzen, 1978] Wolfgang Lenzen. *Recent Work in Epistemic Logic*. Acta Philosophica 30. North Holland Publishing Company, 1978. 5, 19, 101, 102, 111, 112

[Lenzen, 1979] Wolfgang Lenzen. Epistemologische betractungen zu [S4;S5]. *Erkenntnis*, 14:33–56, 1979. 112

[Levesque, 1984] Hector Levesque. A logic of implicit and explicit knowledge. In *AAAI-84*, pages 198–202, Austin Texas, 1984. 32

[Lewis, 1969] David Lewis. *Convention, a Philosophical Study*. Harvard University Press, 1969. 12

[Lewis, 1973] David Lewis. *Counterfactuals*. Basil Blackwell, Oxford, 1973. 57

[Li and Vitányi, 1993] Ming Li and Paul Vitányi. *An introduction to Kolmogorov complexity and its applications*. Graduate texts in computer science. Springer-Verlag, Berlin, second edition, 1993. 2

[Lismont and Mongin, 1994] Luc Lismont and Philippe Mongin. On the logic of common belief and common knowledge. *Theory and Decision*, 37(1):75–106, 1994. 13

[Lomuscio, 1999] Alessio Lomuscio. *Knowledge Sharing among Ideal Agents*. PhD thesis, University of Birmingham, 1999. 10

[Lorini and Castelfranchi, 2007] Emiliano Lorini and Cristiano Castelfranchi. The cognitive structure of surprise: looking for basic principles. *Topoi: An International Review of Philosophy*, 26(1):133–149, 2007. 102

[Lycan, 2006] William Lycan. *Epistemology Futures*, chapter On the Gettier Problem Problem, pages 148–168. Oxford University Press, 2006. 110

[Meyer and van der Hoek, 1995] John-Jules Ch. Meyer and Wiebe van der Hoek. *Epistemic Logic for AI and Computer Science*. Cambridge University Press, Cambridge, 1995. 1, 5

[Meyer *et al.*, 2000] Thomas Andreas Meyer, Willem Adrian Labuschagne, and Johannes Heidema. Refined epistemic entrenchment. *Journal of Logic, Language and Information*, 9(2):237–259, 2000. 57

[Meyer *et al.*, 2001] John-Jules Ch. Meyer, Frank de Boer, Rogier van Eijk, Koen Hindriks, and Wiebe van der Hoek. On programming KARO agents. *Logic Journal of the IGPL*, 9(2), 2001. 19

[Moore, 1985] Robert Moore. A formal theory of knowledge and action. In J.R. Hobbs and R.C. Moore, editors, *Formal Theories of the Commonsense World*, pages 319–358. Ablex, Norwood, NJ, 1985. 35

[Nittka, 2008] Alexander Nittka. *A Method for Reasoning about other Agents' Beliefs from Observations*. PhD thesis, University of Leipzig, 2008. 18

[Parikh and Ramanujam, 2003] Rohit Parikh and Ramaswamy Ramanujam. A knowledge based semantics of messages. *Journal of Logic, Language and Information*, 12(4):453–467, 2003. 139

[Plaza, 1989] Jan Plaza. Logics of public communcations. In M. L. Emrich, M. Z. Pfeifer, M. Hadzikadic, and Z. W. Ras, editors, *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*, pages 201–216, 1989. 35

[Pratt, 1976] Vaughan Pratt. Semantical considerations on floyd-hoare logic. In *Proceedings of the 17th IEEE Symposium on the Foundations of Computer Science*, pages 109–121, 1976. 38

[Rao and Georgeff, 1991] Anand Rao and Michael Georgeff. Modeling rational agents within a BDI-architecture. In R. Fikes and E. Sandewall, editors, *Proceedings of Knowledge Representation and Reasoning (KR & R-91)*, pages 473–484. Morgan Kaufmann Publishers, 1991. 19

[Renardel De Lavalette, 2004] Gerard R. Renardel De Lavalette. Changing modalities. *Journal of Logic and Computation*, 14(2):251–275, 2004. 107

[Ruan, 2004] Ji Ruan. Eploring the update universe. Master's thesis, ILLC, University of Amsterdam, The Netherland, 2004. 55

[Sack, 2007] Joshua Sack. Logic for update products and steps into the past. Submitted to Annals of Pure and Applied Logic, 2007. 140

[Sack, 2008] Joshua Sack. Temporal languages for epistemic programs. *Journal of Logic, Language and Information*, 17(2):183–216, 2008. 140

[Sahlqvist, 1975] Henrik Sahlqvist. Completeness and correspondence in the first and second order semantics for modal logics. In Stig Kanger, editor, *Proceedings of the 3rd Scandinavian Logic Symposium 1973*, number 82 in Studies in Logic. North Holland, 1975. 132

[Segerberg, 1995] Krister Segerberg. Belief revision from the point of view of doxastic logic. *Bulletin of the IGPL*, 3:534–553, 1995. 127

[Segerberg, 1999] Krister Segerberg. Two traditions in the logic of belief: bringing them together. In Hans Jürgen Ohlbach and Uwe Reyle, editors, *Logic, Language and Reasoning: essays in honour of Dov Gabbay*, volume 5 of *Trends in Logic*, pages 135–147. Kluwer Academic Publishers, Dordrecht, 1999. 127

[Spohn, 1988] Wolfgang Spohn. A general non-probability theory of inductive reasoning. In R. Schachter, T. Levitt, L. Kanal, and J. Lemmer, editors, *Uncertainty in Artificial Intelligence 4*, pages 149–158. North-Holland, 1988. 115

[Stalnaker, 2006] Robert Stalnaker. On logics of knowledge and belief. *Philosophical studies*, 128:169–199, 2006. 112, 113

[Steiner and Studer, 2007] David Steiner and Thomas Studer. Total public announcements. In *Symposium on Logical Foundations of Computer Science (LFCS'07)*, pages 498–511, 2007. 91

[Steiner, 2006] David Steiner. A system for consistency preserving belief change. In Sergei Artemov and Rohit Parikh, editors, *ESSLLI'06: Proceedings of the European Summer School in Logic, Language and Information, Workshop on Rationality and Knowledge*, pages 133–144. Association for Logic, Language and Information, 2006. 91

[Tallon *et al.*, 2004] Jean-Marc Tallon, Jean-Christophe Vergnaud, and Shmuel Zamir. Communication among agents: A way to revise beliefs in KD45 kripke structures. *Journal of Applied Non-Classical Logics*, 14(4):477–500, 2004. 26

[van Benthem and Kooi, 2004] Johan van Benthem and Barteld Kooi. Reduction axioms for epistemic actions. In R. Schmidt, I. Pratt-Hartmann, M. Reynolds, and H. Wansing, editors, *AiML-2004: Advances in Modal Logic*, number UMCS-04-9-1 in Technical Report Series, pages 197–211, University of Manchester, 2004. 140

[van Benthem and Liu, 2004] Johan van Benthem and Fenrong Liu. Diversity of agents in games. *Philosophia Scientiae*, 8(2), 2004. 139

[van Benthem and Pacuit, 2006] Johan van Benthem and Eric Pacuit. The tree of knowledge in action: Towards a common perspective. In *Advances in Modal Logic*, pages 87–106, 2006. 129, 139

[van Benthem *et al.*, 2006a] Johan van Benthem, Jelle Gerbrandy, and Barteld Kooi. Dynamic update with probability. Technical report, ILLC, march 2006. 123

[van Benthem *et al.*, 2006b] Johan van Benthem, Jan van Eijck, and Barteld Kooi. Logics of communication and change. *Information and Computation*, 204(11):1620–1662, 2006. 140

[van Benthem *et al.*, 2007] Johan van Benthem, Jelle Gerbrandy, and Eric Pacuit. Merging frameworks for interaction: DEL and ETL. In Dov Samet, editor, *Theoretical Aspect of Rationality and Knowledge (TARK XI)*, pages 72–82, Brussels, June 2007. 139

[van Benthem, 1989] Johan van Benthem. Semantic prallels in natural language and computation. In H.D. Ebbinghaus, J. Fernandez-Prida, M. Garrido, D. lascar, and M.R. Artalejo, editors, *Logic Colloquium '87*. North-Holland, Amsterdam, 1989. 35

[van Benthem, 2003] Johan van Benthem. Conditional probability meets update logic. *Journal of Logic, Language and Information*, 12(4):409–421, 2003. 94, 103, 123

[van Benthem, 2006] Johan van Benthem. "One is a Lonely Number": logic and communication. In Z. Chatzidakis, P. Koepke, and W. Pohlers, editors, *Logic Colloquium'02*, volume 27 of *Lecture Notes in Logic*. 2006. Association for Symboloc Logic. 15

[van der Hoek, 1993] Wiebe van der Hoek. Sytems for knowledge and belief. *Journal of Logic and Computation*, 3(2):173–195, 1993. 111

[van Ditmarsch and Kooi, 2006] Hans van Ditmarsch and Barteld Kooi. The secret of my success. *Synthese*, 151:201–232, 2006. 42

[van Ditmarsch *et al.*, 2003] Hans van Ditmarsch, Wiebe van der Hoek, and Barteld Kooi. Concurrent dynamic epistemic logic. In V. F. Hendricks, K. F. Jorgensen, and S. A. Pedersen, editors, *Knowledge Contributors*, volume 322 of *Synthese Library Series*, pages 105–143. Kluwer Academic Publisher, 2003. 35

[van Ditmarsch *et al.*, 2004] Hans van Ditmarsch, Wiebe van der Hoek, and Barteld Kooi. Public announcements and belief expansion. In *Advances in Modal Logic*, pages 335–346, 2004. 62

[van Ditmarsch *et al.*, 2005] Hans van Ditmarsch, Wiebe van der Hoek, and Barteld Kooi. Dynamic epistemic logic with assignment. In Frank Dignum, Virginia Dignum, Sven Koenig, Sarit Kraus, Munindar P. Singh, and Michael Wooldridge, editors, *Autonomous Agents and Multi-agent Systems (AAMAS 2005)*, pages 141–148. ACM, 2005. 35, 107, 123

[van Ditmarsch *et al.*, 2007a] Hans van Ditmarsch, Ji Ruan, and Wiebe van der Hoek. Model checking dynamic epistemics in branching time. In *Formal Approaches to Multi-agent Systems 2007 (FAMAS 2007)*, Durham UK, 2007. 139

[van Ditmarsch *et al.*, 2007b] Hans van Ditmarsch, Wiebe van der Hoek, and Barteld Kooi. *Dynamic Epistemic Logic*, volume 337 of *Synthese library*. Springer, 2007. 2

[van Ditmarsch, 2000] Hans van Ditmarsch. *Knowledge Games*. ILLC dissertation series DS-2000-06, Amsterdam, 2000. 35

[van Ditmarsch, 2002] Hans van Ditmarsch. Descriptions of game actions. *Journal of Logic, Language and Information (JoLLI)*, 11:349–365, 2002. 35

[van Eijck, 2004] Jan van Eijck. Reducing dynamic epistemic logic to pdl by program transformation. Technical Report SEN-E0423, CWI, 2004. 140

[van Linder *et al.*, 1998] Bernd van Linder, Wiebe van der Hoek, and John-Jules Ch. Meyer. Formalising abilities and opportunities of agents. *Fundamenta Informaticae*, 34(1-2):53–101, 1998. 19

[Veltman, 1996] Frank Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25:221–261, 1996. 35

[Voorbraak, 1993] Frans Voorbraak. *As Far as I know. Epistemic Logic and Uncertainty*. PhD thesis, Utrecht University, 1993. 19, 112

[Weydert, 1994] Emil Weydert. General belief measure. In Ramon López de Mántaras and David Poole, editors, *Tenth Conference on Uncertainty in Artificial Intelligence*, pages 575–582. Morgan Kaufmann, 1994. 114

[Wooldridge, 2000] Michael Wooldridge. *Reasoning About Rational Agents*. MIT Press, June 2000. 19

[Yap, 2006] Audrey Yap. Product update and looking backward. prepublications PP-2006-39, ILLC, 2006. 135, 140

Index

(A, a_a), 36
(M, w), 6, 18
A, 36
$A(\varphi)$, 73
A_0, 73
B_j, 7
C_{G_1}, 13
$Cn(.)$, 58
E, 36
$Ext(.)$, 28
G, 6
K, 58
$K * \varphi$, 58
$K + \varphi$, 58
$K_\perp$, 58
M, 6
$M, w \leftrightarrows M', w'$, 14
$M, w \leftrightarrows_n M', w'$, 78
$M, w \leftrightarrows_{A_1} M', w'$, 74
$M, w \models \varphi$, 7
$M \models \varphi$, 7
$Mod(\varphi)$, 66
N, 6
O, 15
P, 98, 102
$P(\varphi) = \varepsilon$, 102
$P(\varphi) \geq x$, 101
P^W, 106
P^w, 103
P^Γ, 102
Pre, 36
Pre_a, 102
R_j, 6
$St(.)$, 94
Sup^k, 78
U, 15
Y, 16
$[a]$, 39, 128
$[a^-]$, 128
$\Gamma \models_{\mathsf{L}} \varphi$, 9
Φ, 6
$\circ$, 65
$\delta(\varphi)$, 136
$\delta^n(a)$, 43
$\delta_M(w)$, 15
$\delta_M^{A_1}(w)$, 75
$\hat{B}_j$, 7
$\leq_\psi$, 66, 80
$\mathbb{V}$, 94, 95
$\mathcal{A}$, 48
$\mathcal{L}$, 7, 23
$\mathcal{L}^C$, 13
$\mathcal{L}^U$, 15
$\mathcal{L}_0$, 58
$\mathcal{L}_A$, 39
$\mathcal{L}_{A_1}$, 74
$\mathcal{L}_{St}$, 101
$\mathcal{L}_{\neq Y}$, 77
$\mathcal{L}^C_{\neq Y}$, 67
$\mathcal{L}_{\mathsf{EDL}}$, 128
$\mathcal{M}$, 21
$\mathcal{M}, w \models \varphi$, 23
$\mathcal{P}(A)$, 41
$\mathcal{R}_a$, 129
$\mathcal{R}_a^{-1}$, 129
$\mathcal{W}_G$, 68

$\models_{\mathsf{EDL}}$, 131
$\models_{\mathsf{C}} \varphi$, 9
$\models_{\mathsf{L}} \varphi$, 9
$\otimes$, 37, 106
$\otimes_1$, 49
$\otimes_{\mathsf{BMS}}$, 135
$\otimes_{\mathsf{EDL}}$, 135
$\sigma(.)$, 79
.2, 11
.3.2, 11
.3, 11
.4, 11
4, 11
5, 11
D, 11
$\mathsf{KD45}_{\mathsf{G}}$, 8, 11
K_{G}, 8
$\mathsf{S4.2}_{\mathsf{G}}$, 8, 11
$\mathsf{S4.3.2}_{\mathsf{G}}$, 8, 11
$\mathsf{S4.3}_{\mathsf{G}}$, 8, 11
$\mathsf{S4.4}_{\mathsf{G}}$, 8, 11
$\mathsf{S4}_{\mathsf{G}}$, 8, 11
$\mathsf{S5}_{\mathsf{G}}$, 8, 11
$\Gamma(A)$, 132
ε, 94
$\vdash_{\mathsf{L}} \varphi$, 10
$deg(\varphi)$, 7
$form(\mathcal{M})$, 66
$m(.)$, 79
p, 7
$s^k(.)$, 79
.3.2, 8
.4, 8
Int, 28
ed, 129
nf, 129
nl, 129
pf, 73

agent, 1
AGM, 2
  expansion, 58–61
  postulates, 2
    expansion, 58
    revision, 64
  revision, 64–67
  theory, 58–61, 64–73
artificial intelligence, 2, 5, 17, 125

belief base, 65
belief revision, 2, 56, 109, 116, 131
belief set, 58
bisimulation
  for epistemic models, 14
  for event models, 55
BMS, 3
  axioms, 39
  formalism, 36–40
  language, 39

cognitive psychology, 3, 17, 93, 125
coin example, 9, 36, 38
common belief, 12–13
completeness, 11
conceived events, 104
conceived worlds, 98–99, 115, 116
confluence, 8, 11, 120
consequence
  epistemic, 9
  global, 9, 131
  internal
    global, 25
    local, 25
  local, 9
consistent formula, 10
conviction, 101

DDL, 127
decidability, 12, 13
degree of an epistemic formula, 7
degree of similarity, 79
determinism, 56, 133

EDL
  axioms, 131
  decidability, 142
  model, 128
  semantics, 131
epistemic determinism, 129

epistemic logic, 5–15
    complexity, 12, 13
epistemic model, 6
ETL, 139
euclidicity, 8, 11, 40, 111, 112, 121
event model, 36
    associated to agent j, 54
expansion, 58
external approach, 3, 16–19
external model, 26

faithful assignment, 66

game theory, 3, 17, 125
generated submodel, 14, 42
generic event model, 102
Gettier problem, 110, 113

height, 14
higher-order beliefs, 1, 7, 35
Hintikka, 5, 111
hyperreal numbers, 94

imperfect external approach, 16–19, 56
infinitesimal numbers, 94
internal approach, 3, 16–18
internal event model
    of type 1, 48
    of type 2, 48
internal logic, 28
    axioms, 28
    complexity, 31
    decidability, 31
internal model
    of type 1, 19
    of type 2, 21
interpreted systems, 10, 112, 124, 139

knowledge, 110–113

language
    $\mathcal{L}$, 7, 23
    $\mathcal{L}_0$, 58
    $\mathcal{L}^C$, 13
    $\mathcal{L}^U$, 15
    $\mathcal{L}_A$, 39
    $\mathcal{L}^C_{\neq Y}$, 67
    $\mathcal{L}^C_{A_1}$, 74
    $\mathcal{L}_{St}$, 101
    $\mathcal{L}_{\neq Y}$, 77
    $\mathcal{L}_{\mathsf{EDL}}$, 128
logical omniscience problem, 32, 142

model associated to agent j, 26
Moore sentence, 25, 42
multi-agent possible event, 47
multi-agent possible world, 18

negative introspection, *see* euclidicity
no-forgetting, 129
no-learning, 129
no-miracle, 129, 139

ontic events, 56, 130

pd-model, 98
PDL, 38, 127, 140
perfect external approach, *see* external approach
perfect-recall, 129, 139
positive introspection, *see* transitivity
private announcement, 2, 36
proof system, 10
provable formula, 10

ranking, 114–115, 124
reduction axioms, 40, 125
reflexivity, 8, 40, 111
representation theorems, 66, 68
revision operation, 58

satisfiability, 9
    external, 26
    internal, 25
seriality, 8, 11, 111, 121
    preservation, 40–47, 51, 56, 91
situation calculus, 94, 124
soundness, 11
surprising events, 104
surprising worlds, 98–99, 115, 116, 142

transitivity, 8, 11, 40, 111, 121

true belief, 11

universal modality, 15
update product, 37, 105
    preservation, 55
updated model
    of type 1, 49
    of type 2, 50

validity, 9
    external, 26
    internal, 25
    negative, 25
    positive, 25
valuation, 6, 139

weak belief, 101
weak connectedness, 8, 11, 120

Des Perspectives sur les Croyances et le Changement

GUILLAUME AUCHER

Dans cette thèse, nous proposons des modèles logiques pour la représentation des croyances et leur changement dans un cadre multi-agent, en insistant sur l'importance de se fixer un point de vue particulier pour la modélisation. A cet égard, nous distinguons deux approches différentes: l'approche externe, où le modélisateur est quelqu'un d'externe à la situation; l'approche interne, où le modélisateur est l'un des agents. Nous proposons une version interne de la logique épistémique dynamique (avec des modèles d'événements), ce qui nous permet de généraliser facilement la théorie de la révision des croyances d'AGM au cas multi-agent. Ensuite, nous modélisons les dynamismes logiques complexes qui sous-tendent notre interprétation des événements en introduisant des probabilités et des infinitésimaux. Finalement, nous proposons un formalisme alternatif qui n'utilise pas de modèle d'événement mais qui introduit à la place un opérateur d'événement inverse.

Mots clés: Logique épistémique, Logique dynamique, Logique épistémique dynamique, Révision des croyances, Représentation des connaissances, Systèmes multi-agent.

Cette thèse, présentée et soutenue à Toulouse le 9 juillet 2008, a été réalisée sous la direction de Hans van Ditmarsch (Nouvelle Zélande) et d'Andreas Herzig (France). L'auteur a obtenu le grade de docteur en informatique de l'Université d'Otago et de l'Université de Toulouse.

University of Otago - Department of Computer Science
PO Box 56, DUNEDIN 9054, New Zealand

.

Université de Toulouse - Institut de Recherche en Informatique de Toulouse
118 route de Narbonne, 31062 TOULOUSE Cedex 9, France

Perspectives on Belief and Change

GUILLAUME AUCHER

In this thesis, we propose logical models for belief representation and belief change in a multi-agent setting, stressing the importance of choosing a particular modeling point of view. In that respect, we distinguish two approaches: the external approach, where the modeler is somebody external to the situation; the internal approach, where the modeler is one of the agents. We propose an internal version of dynamic epistemic logic (with event models) which allows us to generalize easily AGM belief revision theory to the multi-agent case. Afterwards, we model the complex logical dynamics underlying the interpretation of events by adding probabilities and infinitesimals. Finally we propose an alternative without using event models by introducing instead a converse event operator.

Keywords: Epistemic logic, Dynamic logic, Dynamic epistemic logic, Belief revision, Knowledge representation, Multi-agent systems.

This thesis, presented and defended at Toulouse on the 9^{th} of July 2008, was performed under the supervision of Hans van Ditmarsch (New Zealand) and Andreas Herzig (France). The author obtained the degree of Doctor of Philosophy in Computer science of the University of Otago and Docteur en Informatique de l'Université de Toulouse.

University of Otago - Department of Computer Science
PO Box 56, DUNEDIN 9054, New Zealand

.

Université de Toulouse - Institut de Recherche en Informatique de Toulouse
118 route de Narbonne, 31062 TOULOUSE Cedex 9, France