PLAYING WITH KNOWLEDGE AND BELIEF

Virginie Fiutek

PLAYING WITH KNOWLEDGE AND BELIEF

ILLC Dissertation Series DS-2013-02

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Science Park 107
1098 XG Amsterdam
phone: +31-20-525 6051
fax: +31-20-525 5206
e-mail: illc@uva.nl
homepage: http://www.illc.uva.nl/

These investigations were supported in part by the Philosophy Department at the University of Groningen and in part by the ILLC at the University of Amsterdam.

Cover design by Sébastien Magnier.
Printed and bound by Ipskamp Drukkers.

ISBN: 978–94–6191–942–7

PLAYING WITH KNOWLEDGE AND BELIEF

Academisch Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof.dr. D. van den Boom
ten overstaan van een door het college voor
promoties ingestelde commissie, in het openbaar
te verdedigen in de Aula der Universiteit
op donderdag 12 december 2013, te 13.00 uur

door

Virginie Fiutek

geboren te Grande-Synthe, Frankrijk.

Promotor: Prof.dr. F. Veltman
Co-promotor: Dr. S. Smets

Overige leden: Prof.dr. J.F.A.K van Benthem
Prof.dr. M. Marion
Prof.dr.mr. H. Prakken
Prof.dr. S. Rahman
Prof.dr. G. Sandu
Prof.dr. M.J.B Stokhof

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

to all those who have trusted me so far

and continue to support me

Contents

Acknowledgments

I would like to take this opportunity to acknowledge a large number of people.

First and foremost, I would like to thank my daily supervisor Sonja Smets for her constant and unfailing support. Thanks for your confidence in me. I really enjoyed to work with you on all the interesting topics in my thesis.

Writing this thesis would not have been possible without the help of Alexandru Baltag. Unfortunately you are not one of my official co-supervisors, but unofficially you are definitely one! Thanks for all of your advice and all of these discussions in nice pubs!

I thank both Jeanne Peijnenburg and Frank Veltman for accepting to be involved in the process as the official main promotors of my thesis project. Jeanne Peijnenburg was my first official promotor at the University of Groningen and later this role was taken over by Frank Veltman at the University of Amsterdam, I am thankful to both for their official role in this.

I can honestly say that without Shahid Rahman, I would not be where I am today. He made me discover the wonderful world of Logic and also supported me when I applied for the PhD position in Groningen. Thank you for all of this!

I am extremely grateful to the members of my thesis committee, Johan van Benthem, Mathieu Marion, Henry Prakken, Shahid Rahman, Gabriel Sandu, Martin Stokhof for accepting to read through the manuscript of this thesis and for giving very useful comments on it.

I also want to thank my colleagues: Ben who was part of this adventure from the very beginning, Bryan who made my office in Groningen a very nice place to be, Sujata for the many meetings and discussions we had together in Groningen, Olivier for giving me the opportunity to give a guest lecture on Dialogical Logic in Munich, Kohei and Jort who always spontaneously helped me with my Lira tasks, Nina who patiently answered my numerous questions, Joshua who read the thesis and gave me useful comments and Zoé for all the discussions in french which made the ILLC more like home.

I really must thank Jenny Batson, Karin Gigengack, Tanja Kassenaar, and Peter van Ormondt, for all their valuable help with administrative matters allowing me to save time and focus on my thesis.

The one thing I know is that I would not be here if my family and friends had not been there from the start. Thank you to Cristina, Matthieu, Nicolas, Laurent, Juliele and all the others I have not explicitly mentioned, here or elsewhere, I will never forget you. It would take too long to list everything for which I am deeply grateful to Sébastien, so I will just say: thank you for everything! I know I can always rely on my sister and brother, when things go right and when they go wrong, thank you for that. Finally, I will never be able to thank my parents enough for their unconditional support and love.

Amsterdam Virginie Fiutek
October, 2013.

Chapter 1

Introduction

The title of this book may be puzzling. "Playing with knowledge and belief" sounds somewhat flippant and flimsy. Actually, this book is a very serious study of belief change. In everyday life, people are constantly changing their minds. In the hyper connected world we live in, we receive lots of information coming from different sources. These pieces of information cause changes in the way we look at the world. Roughly speaking, we can distinguish between two attitudes a person (from now we will call such a person "an agent") has towards some proposition describing the state of the world (here "world" is taken to be a very large notion including the agent's environment as well as her own mental state). The first attitude is called knowledge while the second attitude is called belief. An agent may know a given proposition p, or she may believe this proposition p. There is a common understanding that the notion of knowledge is stronger than the notion of belief. Indeed the notion of knowledge involves the idea of factivity: if an agent knows p then p is true while the notion of belief does not. If an agent believes p, it does not imply that p is true. The agent may hold wrong beliefs.

In this book we are interested in how the knowledge and the beliefs of an agent[1] evolve when she learns new information. On the one hand, knowledge changes monotonically. Once an agent knows a proposition, she will never give up her knowledge about this proposition but she may learn and come to know some other propositions when she is given new information. So the knowledge of an agent can only increase over time. On the other hand, beliefs do not change monotonically. The agent may come to give up or revise her beliefs. Now we distinguish between two types of information an agent can receive: factual information and higher-order information. An agent receives a piece of factual information when she is informed of some facts about her environment, about the world she lives in. Thus if she is informed that the Earth revolves around

[1]The agent we consider is taken to be an ideal agent who is able to remember the past. Thus she remembers all the information she once had. In short we summarize this by saying that our agent has "perfect recall".

the Sun, she has received a piece of factual information. She receives a piece of higher-order information if this incoming information is about her own knowledge or beliefs. For example, if she is informed that she *believes* that the Earth revolves around the Sun, she has received a piece of higher-order information. We point out that from now on, we only consider an unchanging factual world. This means that we deal with a world where the facts do not change over time contrary to the mental state of the agent. In this world, the agent only receives information about a static, fixed factual world or about her own mental state.

Philosophy has long considered the notion of knowledge as a topic of key interest. Plato already investigates the concepts of knowledge and belief in dialogues such as *Meno*, *Theaetetus* and *Gorgias*. The specific branch of philosophy dealing with the nature of knowledge and belief is called epistemology. Formal epistemology has recently been developed to model and reason about knowledge and beliefs with formal methods (mathematically based techniques). In the literature, there already exist many different settings to model and formalise belief change. We can distinguish between probabilistic settings and purely qualitative settings. Among the last ones, the setting of modal logic has recently led to a new approach to model belief change. Some modal settings adopt a single agent perspective while others focus on a multi-agent perspective. Single agent settings only deal with the knowledge and beliefs of one single agent, disregarding the existence of other agents in their models. On the contrary multi-agent settings are interested in the interaction between the knowledge and beliefs of several agents (for example, what agents believe about the beliefs of the other agents...). In this thesis we focus only on the single agent setting. We can further distinguish between two different approaches in this thesis, in particular we focus on those with a dynamic dimension and those with a temporal dimension. Dynamic settings can be divided into two branches depending on whether the dynamics is internal as in the setting of Dynamic Doxastic Logic (abbreviated as *DDL*) or external as in the setting of Dynamic Epistemic Logic (abbreviated as *DEL*). As the later chapters show, these two dynamic settings use the formal tools of dynamic modal logic to model belief change. Among the approaches with a temporal dimension, we will focus later on Giacomo Bonanno's setting: his temporal setting uses temporal (modal) logic to model belief change. Other approaches such as the dialogical approach have an argumentative dimension: argumentative settings use argumentation (in dialogues or games) to model belief change. In the last chapter we will focus on a dialogical approach to reason about belief revision.

In this thesis we set three different goals. First we expose a brief overview of some of the different existing settings of belief change so we can later refer back to them. Next, we develop Soft Dynamic Epistemic Logic in three different ways:

1. we study belief contraction in the setting of Dynamic Epistemic Logic. We define three belief contracting operations as operations on total plausibility models, assigning them epistemic actions and axiomatizing them in *DEL* style.

2. we introduce justification models as a new formalisation of belief, evidence and justification. This generalized setting will be used to explore new solutions to some known epistemic issues.

3. we use this new setting to define a game semantics allowing to determine if an agent really defeasibly knows some given proposition (or only believes it).

Finally, we connect Soft Dynamic Epistemic Logic to two of the above mentioned approaches:

1. we compare the internal dynamics of Dynamic Doxastic Logic (*DDL*) with the external dynamics of Dynamic Epistemic Logic (*DEL*);

2. we connect the temporal setting of Bonanno to model belief change with a dialogical approach to logic, providing an argumentative study of belief revision.

We choose to focus on Soft Dynamic Epistemic Logic since it is a powerful approach to model and reason about belief change and can be applied to tackle epistemological issues. To achieve our goals, we will need a tool which should be flexible enough to be easily adapted with respect to what properties of belief change we want to capture. We choose such a tool from the literature, using a family of spheres in Lewis-Grove style.[2] Thus we use nested sphere systems in Chapter 3, we use sphere-based justification models in Chapter 4 and hypertheories in Chapter 6. Finally, we are looking for some logical interactive tools to solve epistemological problems. We use a game semantics in Chapter 5, to solve the Gettier problem and we use the dialogical approach to logic in Chapter 7 to provide an argumentative study of temporal belief revision logic. Throughout this thesis, apart from Chapters 2 and 6, we restrict ourselves to *finite* models.

Through this book, we are looking for answers to some key questions.

– How can we model the result of changing our beliefs in the light of new information? Should we model belief revision in a static way or in a dynamic way? How can we model the dynamics of belief revision?

  The first formalisations of the notion of belief [41] were provided using static Kripke models. Indeed these Kripke models are static since the set

[2]We introduce Grove spheres in Section 2.2.

of possible worlds, the accessibility relation and the valuation remain fixed, unchanged. The main challenge was then to find a way to represent the mechanism of belief revision because these Kripke models are not fit for this purpose. The author in [74, 75, 76] proposed another type of models to internalize the dynamics of belief revision in which the model does not change in itself but the structure of the model does; while the authors of [8, 3, 10] provide richer structures than Kripke models as well as model changing operations on this new type of models, in this way the fundamental dynamic mechanism of belief revision is external to the models. The authors in [6, 8] also proposed an extension of Kripke models with an operation for hypothetical belief revision. Indeed it is interesting to describe what an agent would come to believe if she would receive a new piece of information. Here the purpose is not to capture the actual change of beliefs induced by a new piece of information but to capture what the agent would believe (after revision) about the state of the world as it is *before* the agent receives the information.

– How do the beliefs of an agent behave over time? How do beliefs evolve in a temporal setting? How are the revised beliefs of an agent different if he receives two different pieces of information? How can we compare these revised beliefs? What does this tell us about the rationality of the agent?

  Since beliefs change over time, many authors point out how important it is to formally represent this temporal parameter in belief revision theory [18, 21, 17]. This allows us to study the interaction between beliefs and information change over time. While the family of dynamic logics mentioned above describes local changes of models, temporal doxastic logics provide a global view of all the possible evolutions of the beliefs of an agent inside one global model. In other words, the model unfolds possible histories of informational processes involving belief change processes.

– How powerful are the different approaches of belief revision? How close or different are the different settings of belief revision? How can we connect the existing formalisms?

  Many different settings have already been provided to model belief revision, each of them comes with its own particular dimension. The authors in [27, 12, 13] have very recently started to compare some of the different settings. It is interesting to first compare the numerous existing frameworks instead of only creating new ones. The approach of dynamic logics for belief change has been compared with the approach of doxastic temporal logics in [27, 13]. Moreover, the approach of Dynamic Doxastic Logic and the approach of dynamic logics for belief change have been compared in [12].

– When can we say that an agent knows a proposition and not only believes this proposition? When can we say that a belief is justified? How can the setting of justification of beliefs be formalised?

  The first question has been asked for a long time and many have already tried to answer it. Thus the notions of knowledge and belief have been connected to the notion of justification since Plato. The evidence framework of [16] has been provided to formalise the notion of justification of beliefs. The models of [8, 3, 10] can also be used to deal with justifiable beliefs. However these settings seem too specific: either they satisfy all the AGM postulates[3] but they only consider consistent pieces of evidence ([8, 3, 10]), or they allow non compatible pieces of evidence but then the setting does not satisfy all the AGM postulates ([16]). Hence, it is important to design a new framework general enough to deal with justification as foundation of belief and knowledge such that this new setting allows for inconsistent evidence and satisfies the AGM postulates.

– What would an argumentative study of belief revision logic amount to? How are the notions of beliefs and information interpreted in an argumentative framework? How can we define a particular belief revision policy in such a framework?

  Public Announcement Logic has been quite recently studied in an argumentative setting [56]. The meaning of the public announcement operator is here reconstructed in terms of choice (more precisely, burden of choice) instead of truth. Both [56] and [57] explore this particular use of public announcement operators in the context of legal reasoning. Belief dynamics in an argumentative setting is also worthy of investigation. Such an argumentative setting can provide some new insights on the relation between beliefs and information.

[3]We introduce the AGM postulates in Section 2.1.

Structure and sources of the book

The content of this book is organized in three parts and 8 chapters. We provide here a brief overview of these chapters.

Part I consists of the chapter 2.
Chapter 2 is a background chapter, presenting some existing logics of belief change.

Part II consists of the chapters 3, 4 and 5. In these chapters, we develop Soft Dynamic Epistemic Logic in three different ways.
Chapter 3 This chapter introduces three operations of *AGM*-friendly versions of belief contraction. We define these three contracting operations as operations on plausibility models, we associate to them epistemic actions in *DEL* style and axiomatize them in *DEL* style.
Chapter 4 We introduce justification models as a general framework which represents the information and justification an agent has.
Chapter 5 This chapter provides a game semantics to formalise the notion of defeasible knowledge of Keith Lehrer. Our game formally determines if an agent defeasibly knows a proposition or merely believes but does not know this proposition.

Part III consists of the chapters 6 and 7. In these chapters, we connect Soft Dynamic Epistemic Logic with other existing settings.
Chapter 6 is based on: A. Baltag, V. Fiutek and S. Smets. *DDL* as an "Internalization" of Dynamic Belief Revision. In *Krister Segerberg on Logic of Action.* Outstanding Contributions to Logic, Springer, 2014. This chapter compares two main frameworks of belief revision namely, Dynamic Doxastic Logic and Dynamic Epistemic Logic. More precisely, full *DDL* is studied from the perspective of soft *DEL* in order to show that the *DDL* approach is at least as powerful as the *DEL* approach. We provide several versions of *DDL* internalizing different belief revision operations, as well as several operations of expansion and contraction.
Chapter 7 is based on: V. Fiutek, H. Rückert and S. Rahman. A Dialogical Semantics for Bonanno's System of Belief Revision. *Construction.* College Publications, London, pages 315-334, 2010. And on: V. Fiutek. A Dialogical Approach of Iterated Belief Revision. *Logic of Knowledge. Theory and Applications.* C. Gómez, S. Magnier, and F. Salguero, (eds.), College Publications, London, 2012. This chapter provides an argumentative study of a belief revision logic. Indeed we provide a dialogical reconstruction of the (last version of) branching-time belief revision logic of Bonanno. We first define the dialogical setting and then provide the language and rules of our dialogical system of belief revision. We focus on the dialogical interpretation of the notions of belief and information.

Chapter 8 summarizes the results achieved in this book and states some open questions.

Part I

General background

Chapter 2

Logics of belief change

In this chapter we provide a brief overview of some of the literature that we rely on in the following chapters. In the next sections, we introduce a number of different settings so we can later refer back to them. Note that we have chosen to restrict ourselves to present you in this chapter only with the necessary information needed for this thesis, without presenting all the details and interesting philosophical discussions surrounding each one of the different approaches.

2.1 *AGM* theory of belief revision

The first formal approach provided to deal with belief revision is the *AGM* theory called after the authors Alchourrón, Gärdenfors and Makinson. Since 1985, [1] is the main reference in the traditional literature on belief revision. The authors investigate how we can model the mechanism of rational belief change by analyzing the rationality constraints that are (or should be) imposed upon belief revision. The *AGM* approach adopts a syntactic view of the beliefs of an agent and provides a clear list of postulates that capture rational belief change.

In the *AGM* approach to belief revision, beliefs are modelled as sets of propositional formulas of the language $\mathcal{L}$ which is defined as follows:

2.1.1. DEFINITION. Let $\mathcal{L}$ be the set of formulas of a propositional language based on a given countable set of atomic sentences, and defined via:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \psi$$

2.1.2. DEFINITION. For a subset $Th \subseteq \mathcal{L}$, we denote $[Th]$ as the deductive closure of Th. The deductively closed sets of formulas are called "theories". Let $\bot$ denote the inconsistent theory (containing all formulas).

The revised beliefs form a new set of formulas (a new theory) obtained as the result of an operation which takes as input a theory and a formula. If the initial beliefs are represented by a deductively closed set of formulas (a theory) and the new information is a formula in the same language $\mathcal{L}$, then the result should again form a deductive closed set of sentences. This operation is called revision and satisfies 8 rationality postulates in the AGM theory of belief revision. We provide these revision postulates in Definition 2.1.3 but before we do this, let us mention that there are two other operations which are immediately relevant for our discussion. These operations are called expansion and contraction. Similar as for revision, belief contraction is also regulated by means of a series of postulates in AGM which we give below in Definition 2.1.4. Formally belief expansion can be defined as follows: the expansion $[Th] + \varphi$ of a theory $[Th]$ with a formula φ is given by, $[Th] + \varphi = \{\psi \mid [Th] \cup \{\varphi\} \vdash \psi\}$ which in turn will correspond to a series of expansion postulates (see [1]).

2.1.3. Definition. Suppose that $\mathbb{T}$ is the set of all theories. The belief revision operator $*$ is a map from $\mathbb{T} \times \mathcal{L}$ to $\mathbb{T}$ ($* : \mathbb{T} \times \mathcal{L} \longrightarrow \mathbb{T}$), satisfying 8 postulates.

($*$1. type)	$[Th] * \varphi$ is a theory;
($*$2. success)	$\varphi \in [Th] * \varphi$;
($*$3-4. upper and lower bound)	if $\neg\varphi \notin [Th]$, then $[Th] * \varphi = [Th] + \varphi$;
($*$5. triviality)	$[Th] * \varphi = \bot$ iff $\vdash \neg\varphi$;
($*$6. extensionality)	if $\vdash \varphi \leftrightarrow \psi$, then $[Th] * \varphi = [Th] * \psi$;
($*$7-8. iterated $*$3-4.)	if $\neg\psi \notin [Th] * \varphi$, then $[Th] * (\varphi \wedge \psi) = ([Th] * \varphi) + \psi$.

Postulate ($*$1. type) guarantees that the result of a revision is a theory (the revised belief set is deductively closed). Postulate ($*$2. success) states that after revision with φ, φ is believed (information is believed). Postulates ($*$3-4. upper and lower bound) express minimality of revision: revising with φ and expanding with φ is the same if φ is consistent with the theory. Postulate ($*$5. triviality) guarantees that only a new inconsistent piece of information can produce an inconsistent theory. Postulate ($*$6. extensionality) states that if two formulas are propositionally equivalent then the result of a revision with one of the two formulas is the same as the result of a revision with the other one. Postulates ($*$7-8. iterated $*$3-4.) generalise postulates ($*$3-4. upper and lower bound) in the case of iterated revision. Postulates ($*$7-8. iterated $*$3-4.) state that if ψ is consistent with the result of a revision with φ, then the result of a revision with $\varphi \wedge \psi$ is the same as the result of first revising with φ and then expanding with ψ.

2.1.4. Definition. Suppose that $\mathbb{T}$ is the set of all theories. The contraction revision operator $\div$ is a map from $\mathbb{T} \times \mathcal{L}$ to $\mathbb{T}$ ($\div : \mathbb{T} \times \mathcal{L} \longrightarrow \mathbb{T}$), satisfying 8 postulates.

(÷1. type)	$[Th] \div \varphi$ is a theory;
(÷2. success)	if $\nvdash \varphi$, then $[Th] \div \varphi \nvdash \varphi$;
(÷3. inclusion)	$[Th] \div \varphi \subseteq [Th]$;
(÷4. vacuity)	if $[Th] \nvdash \varphi$, then $[Th] \div \varphi = [Th]$;
(÷5. extensionality[1])	if $\vdash \varphi \leftrightarrow \psi$, then $[Th] \div \varphi = [Th] \div \psi$;
(÷6. recovery)	$[Th] \subseteq ([Th] \div \varphi) + \varphi$;
(÷7. conjunctive inclusion)	if $[Th] \div (\varphi \wedge \psi) \nvdash \varphi$,
	then $[Th] \div \varphi \subseteq [Th] \div \varphi$;
(÷7. conjunctive overlap)	if $[Th] \div \varphi \cap [Th] \div \psi \subseteq [Th] \div (\varphi \wedge \psi)$

Postulate (÷1. type) guarantees that the result of a contraction is a theory (the contracted belief set is deductively closed). Postulate (÷2. success) states that after a contraction with φ, φ is removed from the theory except if φ is a tautology. Postulate (÷3. inclusion) states that no new belief is added to the contracted belief set. Postulate (÷4. vacuity) says that if the theory does not imply φ, in that case contraction has no effect. Postulate (÷5. extensionality) guarantees that if two formulas are propositionally equivalent then the result of a contraction with one of the two formulas is the same as the result of a contraction with the other one. Postulate (÷6. recovery) states that the result of an expansion with φ of a theory resulting from a contraction with φ is the original theory (the original theory is recovered). Postulate (÷7. conjunctive inclusion) states that if φ is removed after a contraction with $\varphi \wedge \psi$, all formulas that have to be removed in order to remove φ also have to be removed in order to remove $\varphi \wedge \psi$. Postulate (÷8. conjunctive overlap) states that all formulas that do not have to be removed in order to remove φ or in order to remove ψ, do not have to be removed in order to remove $\varphi \wedge \psi$.

The operations of expansion, contraction and revision can be related to each other via the so-called Levi Identity. Starting from a given contraction operator (and expansion operator), one can define a revision operation. Levi defines the operation of revision as an operation of contraction followed by an operation of expansion : $[Th] * p = ([Th] \div \neg p) + p$ [49] also, this is called the Levi identity. The converse is also true: from a revision operator, one can define a contraction operator. Harper defines the operation of contraction as: $[Th] \div p = [Th] \cap ([Th] * \neg p)$ also called Harper identity [40]. Both identities have given rise to many discussions but for the purpose of this thesis we do not develop this point further.

While the above theory is set in a syntactic framework, recent developments show that a semantic approach to belief revision theory can give more insight in belief revision scenarios. We will turn to the semantics of belief change in the remaining sections of this background chapter.

2.2 Grove spheres

One of the first semantic approaches to provide a model for belief revision theory is given in terms of the so-called Grove spheres [39]. Grove adapted the spheres systems which Lewis uses to model counterfactual conditionals [52]. Semantically, the belief state of an agent is modelled using families of sets of sets called spheres. This type of model is built up from sets of possible worlds and so defines propositions as sets of possible worlds. The propositions believed by an agent (constituting her belief set) form a central sphere which is surrounded by concentric spheres, each of them representing a degree of similarity to the central sphere. The motivation behind this type of model is that an agent who receives a new piece of information has several ways to modify her belief set in order to incorporate the new information. The doxastic state of an agent contains much more than just an actual belief set, it also includes the predisposition the agent has for belief change.

We provide an example of a family of spheres in Lewis-Grove style in Figure 2.1. This system of spheres represents the actual belief state of an agent such that the central sphere corresponds to her belief set. In this example, the dots represent states (or possible worlds). The formula φ is satisfied in the states s and w. The possible worlds in the central sphere are the states s and v. Thus in this example, the agent considers φ possible (she does not believe φ nor $\neg\varphi$).

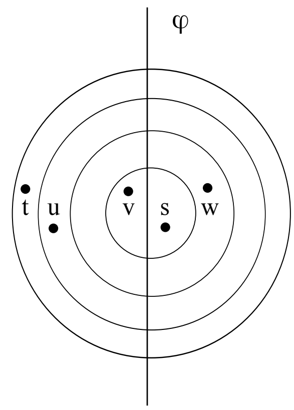

Figure 2.1: Example of a sphere system

Note that the actual state of affairs is not made explicit in the drawing but can in principle correspond to any of the states in Figure 2.1. If the beliefs of the agent happen to be true, then one of the states in the central sphere will correspond to the actual state of affairs.

2.3 Dynamic Doxastic Logic

In 1995, Krister Segerberg proposes a so-called Dynamic Doxastic Logic (*DDL*) [74]. Syntactically, he uses *Propositional Dynamic Logic*-style operators $[\pi]\psi$ whose standard intended interpretation is "ψ holds after the agent performs action π". Note that the origin of Propositional Dynamic Logic (*PDL*) traces back to Fischer and Ladner [32] who propose *PDL* as a modal logic to reason about programs. As a formal language for *PDL*, we include both formulas φ and programs π. In *PDL* we introduce dynamic modalities $[\pi]$ such that the formula $[\pi]\varphi$ describes the properties holding after the execution of some program π. It means "after executing program π, φ holds".

If we take belief changing operations as the programs of *PDL*, we get the core of the idea behind the syntax of *DDL*. Segerberg introduces three different dynamic doxastic operators namely, one operator for expansion $[+]$, one operator for contraction $[-]$ and the last one for revision $[*]$. Indeed he wants to recast AGM theory as a dynamic doxastic logic.

Segerberg provides different settings to study belief change distinguishing between basic *DDL* and full *DDL*. Basic *DDL* restricts revision operators to purely Boolean formulas that is, formulas that do not contain any modal operators. Contrary to full *DDL*, this setting only deals with agents revising their beliefs about the basic facts of the world, not about their own beliefs.

Semantically, Segerberg uses the work of Lewis and Grove [52, 39] (see Section 2.2). Depending on the conditions he imposed on this system of spheres, it is called a hypertheory or an onion. The semantics of onions and hypertheories will be analyzed in detail in Chapter 6.

Segerberg provides several corresponding axiomatizations together with soundness and completeness results [75, 77, 78, 79, 73, 80].

Some of the logics in the literature have first been developed to deal with knowledge change and have subsequently been extended to deal with belief change.

2.4 Public Announcement Logic

The formal study of epistemic logic has been initiated by Georg Henrik von Wright in [81]. Later on, Jaako Hintikka developed von Wright's ideas using modal logic in [41]. We introduce the basic epistemic logic.

Syntactically, a modal operator K is added to the language of propositional logic such that $K\varphi$ means "the agent knows φ". The language of epistemic logic $\mathcal{L}_K$ is defined as follows.

2.4.1. DEFINITION. Let Φ be a set of propositional atoms such that p ranges over Φ.

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \psi \mid K\varphi$$

The knowledge operator gives an epistemic interpretation to the standard necessity operators of modal logic. One can add further notions of group knowledge (distributed knowledge, common knowledge operator and so on), but in this thesis we will restrict all settings to the single agent case.

Semantically, we use Kripke models. As standard in modal logic, we introduce Kripke semantics by first specifying the frames.

2.4.2. DEFINITION. We introduce an *epistemic frame* to be a Kripke frame $(S, \sim)$ where S is a set of states (or possible worlds) and $\sim \subseteq S \times S$ is an equivalence relation.

2.4.3. DEFINITION. An *epistemic model* $\mathcal{M}$ is a Kripke model based on an epistemic frame and is obtained by adding a valuation $V : \Phi \to \mathcal{P}(S)$ assigning to each atomic sentence the set of states in which the sentence is true.

2.4.4. DEFINITION. The valuation map can be extended so that the truth of an arbitrary formula is defined as:

$\mathcal{M}, i \vDash p$ iff $i \in V(p)$.

$\mathcal{M}, i \vDash \neg\varphi$ iff $\mathcal{M}, i \nvDash \varphi$.

$\mathcal{M}, i \vDash \varphi \wedge \psi$ iff $\mathcal{M}, i \vDash \varphi$ and $\mathcal{M}, i \vDash \psi$.

$\mathcal{M}, i \vDash K\varphi$ iff $\big(j \in S$ such that $i \sim j$ implies $\mathcal{M}, j \vDash \varphi\big)$.

An axiomatic system for the basic epistemic logic K is given by the following axioms and rules:

a) All propositional tautologies.

b) $S5$ axioms for K.

$$K(\varphi \to \psi) \to (K\varphi \to K\psi)$$

$$K\varphi \to \varphi$$

$$K\varphi \to KK\varphi$$

$$\neg K\varphi \to K\neg K\varphi$$

c) Rules of inference.

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

$$\frac{\varphi}{K\varphi} \quad (K\text{-Necessitation})$$

The first axiom for K states that if an agent knows φ and knows that $\varphi \to \psi$, then the agent must also know ψ. The second axiom for K states that knowledge is truthful, that is, if an agent knows φ, φ is true. The third and fourth axioms for K respectively state that knowledge is positively and negatively introspective. This means that if an agent knows φ, then he knows that he knows φ and if he does not know φ, he knows that he does not know φ.

Public Announcement Logic, abbreviated as *PAL*, traces back to the work of [64]. In this logic we can express the epistemic change triggered by (truthful) public announcements via dynamic modal operators. In particular, a public announcement of a formula φ is expressed in the language via a dynamic operator (i.e. a modality) labelled by φ. Semantically we model these announcements via so-called model transformers which, as a consequence of the truthful announcement of φ, restrict the epistemic state of the agent such that all worlds where φ does not hold are eliminated from the original model. These model transformers can be studied more generally for several different kinds of events (including false and private announcements) and in essence they form the core ingredient of the logical systems we investigate in the setting of Dynamic Epistemic Logic.

2.4.5. DEFINITION. The language for *PAL* $\mathcal{L}_{K[]}$ is built up from a countable set of propositional atoms Φ, the usual propositional connectives, a unary modal operator K and a dynamic modal operator $[!\varphi]\varphi$. Let p range over the atomic propositions in Φ.

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \psi \mid K\varphi \mid [!\varphi]\psi$$

The dual of the dynamic modality $\neg[!\varphi]\neg\varphi$ is defined as $\langle !\varphi\rangle\varphi$.

The intended interpretation of the announcement operators is as follows:

$[!\varphi]\psi$: after a truthful announcement of φ, it holds that ψ

2.4.6. DEFINITION. The semantic clauses of the public announcement operators are the following:

$\mathcal{M}, i \vDash [!\varphi]\psi$ iff $\big(\mathcal{M}, i \vDash \varphi$ implies $\mathcal{M}|\varphi, i \vDash \psi\big)$.

$\mathcal{M}, i \vDash \langle !\varphi\rangle\psi$ iff $\big(\mathcal{M}, i \vDash \varphi$ and $\mathcal{M}|\varphi, i \vDash \psi\big)$.

where $\mathcal{M}|\varphi = \langle S', \sim', V' \rangle$ is defined as:

- $S' := \{i' \in S \mid (\mathcal{M}, i') \models \varphi\}$
- $\sim' := \{(i', j') \in S' \times S' : (i, j) \in \sim\}$
- $V'_p := \{i' \in S' : i \in V_p\}$

$\mathcal{M}|\varphi$ is the model obtained after the public announcement of φ: it is the model $\mathcal{M}$ restricted to all the worlds where φ holds such that the valuation and the equivalence relations between the remaining worlds do not change.

For a recent overview of the complete axiomatization of *PAL*, see [29]. In the absence of group modalities such as "common knowledge", one of the powerful and attractive features of this logic is the fact that a complete axiomatization for it will follow directly from the axioms and rules of the basic epistemic logic K by applying the following Reduction axioms:

a) Atomic permanence.

$$[!\varphi]p \leftrightarrow (\varphi \to p)$$

b) Announcement and negation.

$$[!\varphi]\neg\psi \leftrightarrow (\varphi \to \neg[!\varphi]\psi)$$

c) Announcement and conjunction.

$$[!\varphi](\psi \wedge \chi) \leftrightarrow ([!\varphi]\psi \wedge [!\varphi]\chi)$$

d) Announcement and knowledge.

$$[!\varphi]K\psi \leftrightarrow (\varphi \to K[!\varphi]\psi)$$

e) Announcement composition.

$$[!\varphi][\psi]\chi \leftrightarrow [!\varphi \wedge [!\varphi]\psi]\chi$$

Using the Reduction Laws, the public announcement operator can be step by step "pushed through" all other operators and at the end, completely eliminated using the Reduction Law for atomic formulas.

2.5 Soft Dynamic Epistemic Logic

Epistemic doxastic logic has been introduced by Jaako Hintikka in [41] as the logic of knowledge and belief. Syntactically, a modal operator K and a modal operator B are added to the language of propositional language such that $K\varphi$ means "the agent knows φ" and $B\varphi$ means "the agent believes φ".

Syntax The language of epistemic logic $\mathcal{L}_{EL}$ is defined as follows.

2.5.1. DEFINITION. Let Φ be a set of propositional atoms such that p ranges over Φ.

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \psi \mid K\varphi \mid B\varphi$$

The knowledge and belief operators give a specific interpretation to the standard necessity operators of modal logic, respectively an epistemic interpretation and a doxastic interpretation. Similar as before, also to this language one can add further notions of group knowledge or group belief (distributed knowledge, common knowledge operator, common belief operator and so on), but in this thesis we will restrict all settings to the single agent case.

Semantics We interpret $\mathcal{L}_{EL}$ in terms of plausibility models [8, 10].

2.5.2. DEFINITION. A pointed *plausibility model* $\mathcal{M}$ is a tuple

$$(S, \leq, V, s_0)$$

where:

- S is a set of states (possible worlds),
- $\leq$ is a well-founded pre-order on S,
- V is a propositional valuation and
- s_0 is an actual state of affairs.

A pre-order is a binary relation that is reflexive and transitive. A well-founded pre-order is a pre-order such that there is no infinite descending chain of states in S, i.e. there is no infinite sequence $s_1 > s_2 > s_3 > \ldots$, with all $s_i \in S$. A well pre-order is a well-founded pre-order that is also connected, that is, for all s and t, either $s \leq t$ or $t \leq s$. We explicitly distinguish between partial plausibility models (plausibility models with a non connected pre-order) and total plausibility models (plausibility models with a connected pre-order). The notation $s \leq t$ denotes that state s is at least as plausible as t for the agent. We write $s < t$ iff $s \leq t$ but $t \nleq s$ and call this the "strict" plausibility relation.

Given a pointed (total or partial) plausibility model $\mathcal{M}$, we identify a "proposition" with any set $P \subseteq S$ of states in $\mathcal{M}$. We write $\mathcal{M}, s \vDash P$ (that is, the proposition P is true at state s in the model $\mathcal{M}$) iff $s \in P$ in $\mathcal{M}$. We define the "always true" $\top$ and "always false" $\bot$ propositions as standard: $\bot := \varnothing, \top := S$. All the operations on sets can be "lifted" to propositions, so that we have a natural meaning of the negation of a proposition $(\neg P) := S \setminus P$, conjunction of propositions $(P \wedge R) := P \cap R$ etc. Besides, for propositions $P, Q \subseteq S$ we define the set of most plausible states given that a certain proposition P is true: $best\,P = Min_{\leq}P := \{s \in P :$ there is no $t < s$ for any $t \in P\}$. We define $best := best\,S$ as the set of most plausible states in the given state space. Propositions provide an interpretation to the well-formed formulas of the language $\mathcal{L}_{EL}$. This is done as usual by lifting the valuation of the atomic facts from basic propositions to complex ones.

We provide an example of a total plausibility model[2] in Figure 2.2. In this example, P is satisfied in the states x, v, t and s. The most plausible states are s and u. The state t is more plausible than the state v, while s and u are equiplausible as well as the states x and w.

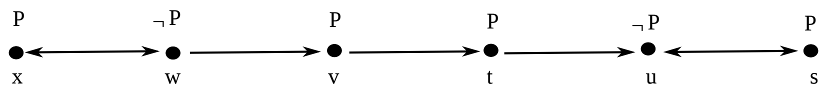

Figure 2.2: Example of a total plausibility model

We use our earlier notation to formally introduce a notion of "irrevocable" knowledge (K) and a notion of belief (B).

Irrevocable knowledge Irrevocable knowledge is formally defined as truth in all possible worlds. Given a (total or partial) plausibility model, we set:

$$KP := \{s \in S : P = S\}$$

where KP is read as "the agent (irrevocably) knows P".

We provide an example of irrevocable knowledge in Figure 2.3. In this example, the proposition P is satisfied in all the states of the model. Hence in this example, the agent irrevocably knows P, i.e. KP is true in all states of the model.

Belief Belief is defined as what is true in the most plausible worlds. Given a (partial or total) plausibility model, we set:

$$BP := \{s \in S : best \subseteq P\}$$

where BP is read as "the agent believes P".

[2]Note that we do not draw the reflexive and transitive arrows in any of our drawings of total or partial plausibility models.

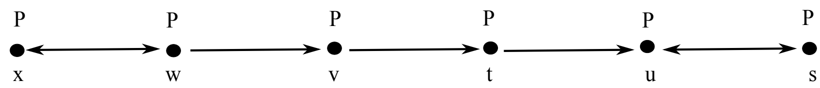

Figure 2.3: Example of irrevocable knowledge

We provide an example of belief in Figure 2.4. In this example, the proposition P is satisfied in the states x, v, u and s. The most plausible states are s and u. Hence in this example, the agent believes P, i.e. BP is true in all states of the model.

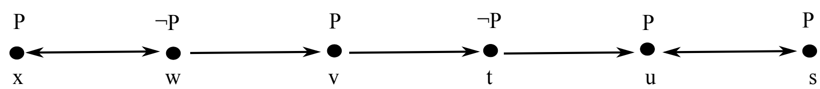

Figure 2.4: Example of belief

One can extend the language of $\mathcal{L}_{EL}$ with three other doxastic operators capturing the notions of conditional belief, strong belief and defeasible knowledge as done in [8].

Note that while the definitions of irrevocable knowledge, belief, conditional belief and strong belief are not state-dependent, the definition of defeasible knowledge is state-dependent. In other words, if an agent knows (believes, conditional believes or strongly believes) P at a state s in a given model $\mathcal{M}$, she also knows (believes, conditional believes or strongly believes) P at another state t in $\mathcal{M}$. But if an agent defeasibly knows P at s in $\mathcal{M}$, she does not necessarily defeasibly know P at any t in $\mathcal{M}$.

Conditional belief Conditional belief is defined as what is true in the most plausible worlds within a given subset Q of the state space (satisfying the condition Q). Given a (total or partial) plausibility model, we set:

$$B^Q P := \{s \in S : bestQ \subseteq P\}$$

where $B^Q P$ is read as "the agent believes P conditional on Q" and means that, if the agent would receive some further (certain) information Q (to be added to what he already knows) then she would believe that P was the case[3].

[3]We refer back to the past tense for the interpretation of conditional belief because of the existence of Moore sentences [59]. We present the Moore paradox in more details in Chapter 6.

We provide an example of conditional belief in Figure 2.5. In this example, the proposition P is satisfied in the states x, v and s and the proposition Q is satisfied in the states x and s. The most plausible state satisfying the condition Q is s. Hence in this example, the agent believes P conditional on Q, i.e. $B^Q P$ is true in all states of the model.

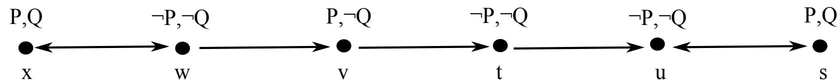

Figure 2.5: Example of conditional belief

2.5.3. PROPOSITION. *Knowledge of a sentence P can be said to be "irrevocable" iff P is known conditional on any information. Formally in a given (partial or total) plausibility model $\mathcal{M}$ we have:*

$$s \vDash KP \textit{ iff } s \vDash B^Q P \textit{ for all } Q$$

.

2.5.4. PROOF. – In the direction from left to right, we start from a given (partial or total) plausibility model $\mathcal{M}$ in which KP is true at s. Now we have to prove that $s \vDash B^Q P$ for all Q. As we know that $s \vDash KP$, this means that $t \vDash P$ for all states t in $\mathcal{M}$. Hence, $bestQ \subseteq P$ for all Q that is, it is the case that $s \vDash B^Q P$ for all Q.

– In the direction from right to left we assume as given a model (partial or total) plausibility model $\mathcal{M}$ and a state s such that $s \vDash B^Q P$ for all Q. Let Q be the singleton $\{t\}$ for any $t \in S$. Then $bestQ \subseteq P$, that is P is true at t for any $t \in S$. Hence, $s \vDash KP$.

□

Strong belief The next attitude is called strong belief [8] and is given by the following definition in a (total or partial) plausibility model:

$$SbP := \{s \in S : P \neq \varnothing \text{ and } t < w \text{ for all } t \in P \text{ and all } w \notin P\}$$

where SbP is read as "the agent strongly believes P". P is a strong belief, held at a state s, iff P is epistemically possible and moreover all epistemically possible P-states are strictly more plausible than all epistemically possible non-P states.

We provide an example of strong belief in Figure 2.6. In this example, the proposition P is satisfied in the states t, v, u and s. Hence in this example, the agent strongly believes P, i.e. SbP is true in all states of the model.

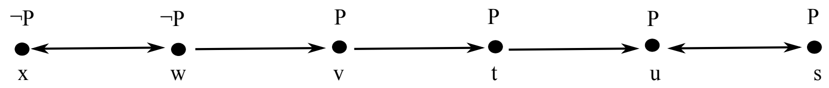

Figure 2.6: Example of strong belief

Defeasible knowledge The next attitude has been called defeasible knowledge in [8] and is given by the following definition in a partial plausibility model[4]:

$$s \models K_D P \text{ iff } \forall t (t \not> s \Rightarrow t \in P)$$

where $K_D P$ is read as "the agent defeasibly knows P". An agent defeasibly knows P in a state s iff P is true in all the worlds that are at least as plausible as s including the non comparable worlds.

In partial plausibility models, K_D is interpreted as the Kripke modality for the relation $\not>$. Thus in this setting, defeasible knowledge is not positively introspective. The notion of defeasible knowledge in partial plausibility models has not yet been studied in detail in the literature. There does not yet exist a complete axiomatization of that notion on partial plausibility models.

We introduce an example of defeasible knowledge in non-connected (partial) plausibility models. In Figure 2.7 we consider a partial plausibility model such that the states s and t, and the states u and v are not comparable. In this model, the agent believes $P \vee Q$ since $P \vee Q$ is true in the most plausible worlds (that is, in s and t). The agent does not defeasibly know $\neg Q$ at v since although $\neg Q$ is true at v and s, it is not true at t.

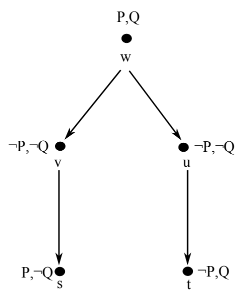

Figure 2.7: Partial plausibility model

In the special case of a total plausibility model, we have:

[4]We use here the notation $\Rightarrow$ for the implication in the meta-language.

2.5.5. Proposition. *In a total plausibility model,*

$$K_D P = \{s \in S : t \leq s \textit{ implies } t \in P\}$$

In a total plausibility model, P is defeasibly known at a state s iff all the states that are at least as plausible as s are P-states. This way of defining K_D in total epistemic plausibility models, uses the converse of $\leq$ as the accessibility relation. This operator satisfies the conditions of an $S4$-type Kripke modality[5].

We provide an example of defeasible knowledge in a total plausibility model, in Figure 2.8. In this example, the proposition P is satisfied in the states v, u and s. This example emphasizes that the defeasible knowledge of an agent is state-dependent. At state s, the agent defeasibly knows P whereas at state v she does not defeasibly know P (since the state t is more plausible than the state v and satisfies $\neg P$).

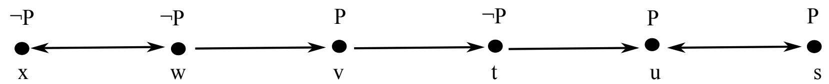

Figure 2.8: Example of defeasible knowledge

In a total plausibility model, knowledge of a sentence P can be said to be defeasible iff P is believed conditional on any true information.

2.5.6. Proposition. *Given a total plausibility model $\mathcal{M}$, observe that:*

$$s \vDash K_D P \textit{ iff } s \vDash B^Q P \textit{ for all } Q \textit{ such that } s \vDash Q$$

2.5.7. Proof.

- In the direction from left to right, we start from a given total plausibility model $\mathcal{M}$ in which $K_D P$ is true at s. Now we have to prove that $s \vDash B^Q P$ for all Q such that $s \vDash Q$. That is, we want to show that $bestQ \subseteq P$ for all Q such that $s \vDash Q$. Let $Q \subseteq S$ be such that $s \in Q$ and let $w \in bestQ$. Then we have $w \leq s$. As we know that $s \vDash K_D P$, this means that $t \vDash P$ for all states $t \leq s$. So $w \in P$. Hence, $bestQ \subseteq P$ for all Q such that $s \vDash Q$, that is it is the case that $s \vDash B^Q P$ for all Q such that $s \vDash Q$.
- In the direction from right to left we assume as given a model $\mathcal{M}$ and a state s such that $s \vDash B^Q P$ for all Q such that $s \vDash Q$. We have to show that $s \vDash K_D P$. Let t be such that $t \leq s$ and let Q be the set $\{s, t\}$. Then

[5] See below the axioms $S4$ for K_D.

$t \in bestQ$. Since $s \models B^Q P$ for all Q such that $s \models Q$, $bestQ \subseteq P$. Hence, $t \models P$ that is, $s \models K_D P$.

□

This alliance between $K_D P$ and a conditional belief in P, tells us something about how stable the belief in P really is in the light of new information. The stability account of knowledge characterizes the difference between knowledge and belief in terms of stability or robustness under belief revision when new true information is received [71, 8].

Irrevocable and defeasible knowledge The main difference between these two types of knowledge is how they persist when new information is received. While irrevocable knowledge is immune to change, that is, it persists whatever (true or false) information is received, defeasible knowledge is only persistent under new *true* pieces of information.

Justifiable beliefs All the beliefs in a given (total or partial) plausibility model $\mathcal{M}$ are justifiable beliefs. This is a consequence of the following proposition stating that an agent believes P at s iff it is entailed by some strong belief in F and $F \subseteq P$:

2.5.8. Proposition. *Given a (total or partial) plausibility model $\mathcal{M}$,*

$$s \models BP \textit{ iff } \exists F \textit{ such that } s \models K(F \to P) \wedge SbF$$

2.5.9. Proof. – In the direction from left to right, we start from a given (total or partial) plausibility model $\mathcal{M}$ in which BP is true at s. Now we have to prove a conjunction, which means that we show that both conjuncts are true. In particular we have to show that there exists a proposition F such that in state s and model $\mathcal{M}$, it is the case that $K(F \to P)$ and we have to show that in this model at state s it is the case that SbF holds. Let us take $F = \{t \mid t \in best\}$. As we know that $s \models BP$ this means that $best \subseteq P$. From this it follows that $F \subseteq P$, or in other words that $F \to P$. This implication is true in all states of the model $\mathcal{M}$, hence at s in $\mathcal{M}$ it is the case that $K(F \to P)$. To show $\mathcal{M}, s \models SbF$, we have to show two things: that $F \neq \varnothing$ and that all F-worlds are strictly more plausible than all $\neg F$-worlds. The first condition is satisfied because $F = best\, S$ and if S is non-empty then $best\, S$ is non-empty. The second condition is satisfied because for any state t if t is in $best\, S$ then from the definition of $best\, S$ it follows that t is strictly more plausible then any state $v \notin best\, S$.

– In the direction from right to left we assume as given a (total or partial) plausibility model $\mathcal{M}$ and a state s such that $s \models K(F \to P)$ and $s \models SbF$. We have to show that $s \models BP$. Because F is a strong belief, $F \neq \varnothing$ and every state $t \leq s$ must satisfy F. Note that from the first conjunct it follows that $F \subseteq P$ in all states of the model, hence all states $t \leq s$ must satisfy P. This means that the most plausible states will be P-states and hence BP is true in all states of the model, and hence also in state s.

□

We now introduce the notion of correctly justifiable belief. A belief is correctly justifiable if it has a correct (i.e. true) justification. So by adding to the previous equivalence in Proposition 2.5.8 the condition that the justification has to be truthful and restricting ourselves to total plausibility models, we obtain the notion of defeasible knowledge as the closure under logical consequence of true strong belief in the total plausibility model $\mathcal{M}$ that is, an agent defeasibly knows P at s iff it is entailed by some true strong belief in F and $s \in F \subseteq P$.

2.5.10. Proposition. *Given a total plausibility model $\mathcal{M}$:*

$$s \vDash K_D P \text{ iff } \exists F \text{ such that } s \vDash F \wedge K(F \rightarrow P) \wedge SbF$$

2.5.11. Proof. – In the direction from left to right we assume as given a model $\mathcal{M}$ and a state s such that $s \vDash K_D P$. Now we have to show that there exists a proposition F such that the following three conjuncts hold in state s and model $\mathcal{M}$, i.e.$s \vDash K(F \rightarrow P)$, $s \vDash SbF$ and $s \vDash F$. Let us take $F = \{t \mid t \leq s\}$ and $F \subseteq P$. Since $s \vDash K_D P$, $t \in P$ for every $t \leq s$ and hence we can indeed require that $F \subseteq P$. From the choice of F it follows that $s \in F$ and hence $s \vDash F$. It also follows that $F \rightarrow P$ is true in all states of the model, hence we have $\mathcal{M}, s \vDash K(F \rightarrow P)$. We now still have to show that $F \neq \varnothing$ and that all F-worlds are strictly more plausible than all $\neg F$-worlds. By the choice of F and by the assumption that $s \vDash K_D P$, any state $t \leq s$ will satisfy both$P$ and F, hence F is non-empty when P is non-empty and the second fact follows from the choice of F. Hence F is a strong belief in this model and in particular also in state s.

– In the direction from right to left we assume as given a model $\mathcal{M}$ and a state s such that there exists an F for which the following conjunction holds: $\mathcal{M}, s \vDash K(F \rightarrow P) \wedge SbF$. Now have to show that P is defeasible knowledge at state s in $\mathcal{M}$. From the first conjunct it follows that$s \in F$ and $F \subseteq P$. Since $s \in SbF$ (second conjunct) then for any $t \leq s$ it follows that t satisfies F and hence also P (because $F \subseteq P$). As all states which are at least as plausible as s satisfy P, it follows that P is defeasible knowledge.

□

One could wonder why would the strong beliefs yield a good notion of justification? The answer to this is that they establish good evidence because one can think of them as originating from highly trusted sources. In order words, justification can be looked at as the pieces of information that have generated or created the agent's plausibility structure on states via a prior process of belief dynamics[6].

Axiomatization Baltag and Smets provide a sound and complete proof system of the logic of irrevocable knowledge and defeasible knowledge (defined as an $S4$-type Kripke modality) [8]. The language of this logic is called $\mathcal{L}_{KK_D}$[7]. In this logic, belief and conditional belief are derived operators:

$$B^{\varphi}\psi := \neg K \neg \varphi \rightarrow \neg K \neg (\varphi \wedge K_D(\varphi \rightarrow \psi)$$

$$B\varphi := B^{\top}\psi$$

2.5.12. PROPOSITION. *Given a semantics for $K_D\varphi$ and $B^{\varphi}\psi$ in plausibility models, we can prove the following semantic equivalence.*

$$B^{\varphi}\psi \iff \neg K \neg \varphi \rightarrow \neg K \neg (\varphi \wedge K_D(\varphi \rightarrow \psi))$$

2.5.13. PROOF. – In the direction from left to right, we start from a given total plausibility model $\mathcal{M}$ in which $B^{\varphi}\psi$ is true at s. Suppose $\neg K \neg \varphi$ is also true at s. We have to prove that $\neg K \neg (\varphi \wedge K_D(\varphi \rightarrow \psi))$ is true at s. So we have to prove that there exists $t \in \mathcal{M}$ such that $t \vDash \varphi$ and $t \vDash K_D(\varphi \rightarrow \psi)$. Then we have to prove that $t \vDash \varphi$ and that for all $w \leq t$, $w \vDash \varphi \rightarrow \psi$. According to our assumption, $s \vDash \neg K \neg \varphi$. So, there exists a state t such that $t \vDash \varphi$. Let $t \in best\varphi$. Since we know that $s \vDash B^{\varphi}\psi$, $best\varphi \subseteq \psi$. Then for all $w \leq t$, $w \vDash \varphi \rightarrow \psi$. And we are done.

– In the direction from right to left, we start from a given total plausibility model $\mathcal{M}$ in which $\neg K \neg \varphi \rightarrow \neg K \neg (\varphi \wedge K_D(\varphi \rightarrow \psi))$ is true at s. We have to show that $s \vDash B^{\varphi}\psi$, that is $best\varphi \subseteq \psi$. As we know, $s \vDash \neg K \neg \varphi \rightarrow \neg K \neg (\varphi \wedge K_D(\varphi \rightarrow \psi))$. Then either:

1. $s \nvDash \neg K \neg \varphi$ or
2. $s \vDash \neg K \neg (\varphi \wedge K_D(\varphi \rightarrow \psi))$.

In the first case, $s \vDash B^{\varphi}\psi$ trivially. In the second case, there exists $t \in \mathcal{M}$ such that $t \vDash \varphi$ and that for all $w \leq t$, $w \vDash \varphi \rightarrow \psi$. Let $t \in best\varphi$. Then $best\varphi \subseteq \psi$. And we are done.

□

[6]We refer to Chapter 4 for more details.
[7]We refer to this language in Chapter 3.

Proof system In addition to the rules and axioms of propositional logic, the proof system for the logic KK_D includes the following rules and axioms:

a) $S5$ axioms for K.

b) $S4$ axioms for K_D.

$$K_D(\varphi \to \psi) \to (K_D\varphi \to K_D\psi)$$
$$K_D\varphi \to \varphi$$
$$K_D\varphi \to K_D K_D\varphi$$

c) $KP \to K_D P$

d) $K(P \vee K_D Q) \wedge K(Q \vee K_D P) \to KP \vee KQ$

e) Necessitation for K and K_D.

As we will refer later to the axiom d (see Chapter 4), we call this axiom *Totality.*

The logic for irrevocable knowledge and defeasible knowledge in partial plausibility models has not yet been axiomatized, this is still an open problem.[8]

Dynamics of information One can think of many ways to change the beliefs of an agent according to the information she receives. She can receive *hard information* or she can receive *soft information.* Hard information is a piece of information that is unrevisable and irrevocable since it has been received from an infallible source while soft information is a piece of information that is potentially revisable since it has been revised from a fallible source.

– Receiving "hard" information φ corresponds to what in the *DEL* literature [11, 3, 29] is called an update $!\varphi$ and in the Belief Revision literature is known as a "radical revision" (or irrevocable revision) with φ. This operation changes the model by eliminating all the $\neg\varphi$-worlds. The result of this elimination is a submodel only consisting of φ-worlds.

– A second, softer kind of revision, is given by the *DEL* operation of lexicographic upgrade $\Uparrow \varphi$ [10, 8], known in the Belief Revision literature as "moderate revision" (or lexicographic revision). This changes the model by making all φ-worlds become more plausible than all $\neg\varphi$-worlds.

[8] Note that in [14] van Benthem, Fernández-Duque and Pacuit provide a full axiomatization for the modality $[\leq]$ in partial pre-ordered models, but this modality does not capture defeasible knowledge (except only in total models).

– Finally, the *DEL* operation of conservative upgrade $\uparrow\varphi$ [10, 8] is known as "conservative revision" (or natural revision) in the Belief Revision literature. This changes the model by making the most plausible φ-worlds become the most plausible overall (while leaving everything else unchanged).

Syntactically, dynamic operators are added to $\mathcal{L}_{EL}$ to express the dynamics of information. The resulting language is called $\mathcal{L}_{DEL}$ and is defined as follows.

2.5.14. DEFINITION. Let Φ be a set of propositional atoms such that p ranges over Φ.

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \psi \mid K\varphi \mid B\varphi \mid B^{\varphi}\psi \mid Sb\varphi \mid K_D\varphi \mid [!\varphi]\psi \mid [\Uparrow\varphi]\psi \mid [\uparrow\varphi]\psi$$

The dual of $[!\varphi]$ is $\langle !\varphi\rangle$. Similarly, the dual of $[\Uparrow\varphi]$ is $\langle\Uparrow\varphi\rangle$ and the dual of $[\uparrow\varphi]$ is $\langle\uparrow\varphi\rangle$.

From traditional *DEL* to soft *DEL* While traditional Dynamic Epistemic Logic only deals with hard information, soft Dynamic Epistemic Logic deals with both hard and soft information.

Hard information The semantic clause for the $!\varphi$ operator is given in Definition 2.4.6.

Van Benthem provides a complete axiomatization for the logic of conditional belief under public announcements in [10].

Note that it is now possible to repackage the above given characterization of the K_D operator using the tools of update operations on models. For ontic (i.e. non-doxastic) facts p which do not refer to the beliefs of an agent (hence when p is a Boolean formula) we state the following proposition:

2.5.15. PROPOSITION. *If p is an atomic formula (i.e. an ontic fact), $s \vDash K_Dp$ iff $s \vDash [!\varphi]Bp$ for every $\varphi \in \mathcal{L}_{DEL}$.*

2.5.16. PROOF. – In the direction from left to right we assume as given a model $\mathcal{M}$ and a state s such that $s \vDash K_Dp$. Now we have to show that $s \vDash [!\varphi]Bp$ for every φ. By Proposition 2.5.6, we have $s \vDash B^{\varphi}p$ for all φ such that $s \vDash \varphi$. It means that $best\varphi \subseteq \parallel p \parallel_{\mathcal{M}}$ for all φ such that $s \vDash \varphi$. We know that in $\mathcal{M}|\varphi$, $\leq' = \leq \cap (S' \times S')$ and that $\parallel p \parallel_{\mathcal{M}|\varphi} = \parallel p \parallel_{\mathcal{M}} \cap \varphi$ (see Definition 2.4.6). For all φ such that $s \vDash \varphi$, we then have $best'\varphi = bestS' \subseteq \parallel p \parallel_{\mathcal{M}|\varphi} = \parallel p \parallel_{\mathcal{M}} \cap \varphi$. So for all φ such that $s \vDash \varphi$, $s \vDash_{\mathcal{M}|\varphi} Bp$. Hence, $s \vDash [!\varphi]Bp$ for every φ.

– In the direction from right to left we assume as given a model $\mathcal{M}$ and a state s such that $s \models [!\varphi]Bp$ for every φ. Take the sentence $\varphi := \neg K_D p$. Then by our assumption, we have $s \models [!\neg K_D p]Bp$. By the above semantics for $[!\varphi]\psi$, this is equivalent to the following statement:

$$(*) \text{ if } \mathcal{M}, s \models \neg K_D p \text{ then } \mathcal{M}|\neg K_D p, s \models Bp$$

So we have two cases:

 – either $\mathcal{M}, s \not\models \neg K_D p$, i.e. $\mathcal{M}, s \models K_D p$ and we are done,
 – or $\mathcal{M}, s \models \neg K_D p$. Then by $(*)$, we get $\mathcal{M}|\neg K_D p, s \models Bp$. Then $best \parallel \neg K_D p \parallel_{\mathcal{M}} \subseteq \parallel p \parallel_{\mathcal{M}|\neg K_D p} \subseteq \parallel p \parallel_{\mathcal{M}}$. Let $t \in best \parallel \neg K_D p \parallel_{\mathcal{M}}$. By the semantics of $K_D\varphi$, there exists $t' \leq t$ such that $t' \in \parallel \neg p \parallel_{\mathcal{M}}$. Since $t' \leq t$, $t' \in best \parallel \neg K_D p \parallel_{\mathcal{M}} \subseteq \parallel p \parallel_{\mathcal{M}}$. But this contradicts $t' \in \parallel \neg p \parallel_{\mathcal{M}}$. Hence, $s \models K_D p$.

□

The use of the update operator $!\varphi$ in Proposition 2.5.15 shows that $K_D p$ corresponds to a belief that is persistent under truthful learning of new (true) facts φ but can be defeated by false information.

Soft information The semantic clause for $\Uparrow \varphi$ operator is:

$$\mathcal{M}, s \models [\Uparrow \varphi]\psi \text{ iff } \mathcal{M}| \Uparrow \varphi, s \models \psi$$

such that $\mathcal{M}| \Uparrow \varphi$ is the result of applying a lexicographic upgrade operation with φ on $\mathcal{M}$, which is defined as follows.

2.5.17. Definition. $\mathcal{M}| \Uparrow \varphi$ is defined as $< S, \leq', V >$ in which $t \leq' s$ iff:

– $t \models \varphi$ and $s \models \neg\varphi$ or

– $t \leq s$.

The semantic clause for $\uparrow \varphi$ operator is:

$$\mathcal{M}, s \models [\uparrow \varphi]\psi \text{ iff } \mathcal{M}| \uparrow \varphi, s \models \psi$$

such that $\mathcal{M}| \uparrow \varphi$ is the result of applying a conservative upgrade operation with φ on $\mathcal{M}$, which is defined as follows.

2.5.18. Definition. $\mathcal{M}| \uparrow \varphi$ is defined as $< S, \leq', V >$ in which $t \leq' s$ iff:

– $t \in best \parallel \varphi \parallel_{\mathcal{M}}$ or

– $t \leq s$.

Van Benthem provides a complete axiomatization for the dynamic logic of lexicographic upgrade and the dynamic logic of conservative upgrade in [10].

Original epistemic models Note that originally, the *DEL* semantics was given in terms of epistemic models that includes standard $S5$ or $S4$-Kripke models $< S, \sim, V >$ such that $\sim$ is an equivalence relation on S. These types of models are designed to model knowledge and work well with update operations modelling knowledge change but they are not ideal to model beliefs (as beliefs are supposed to be fallible). However problems arise when we try to adapt the standard *DEL* models to deal with belief revision. Traditional *DEL* fails to model belief revision when the accessibility relation is serial, transitive and euclidean ($KD45$) or reflexive and transitive ($S4$). Van Benthem provides a perfect example of this failure in [10].

Consider the doxastic model illustrated in Figure 2.9 with two possible worlds s and t such that only the actual state s satisfies the proposition p and also satisfies $B\neg p$. This is a $KD45$ model such that there is *no* reflexive arrow in state s while there is a reflexive arrow in state t.

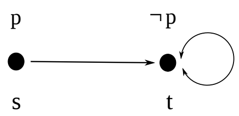

Figure 2.9: Example of the failure of standard doxastic models for belief revision

A piece of hard information $!p$ would change this model into the one-world model s with an empty doxastic accessibility relation. Indeed the state s has no outgoing arrows. In other words, the agent's beliefs become inconsistent!

Taking the conditional belief operator as basic, Baltag and Smets get back to a semantics which is close to *PDL*, they call this Conditional Doxastic Logic.

2.6 Conditional Doxastic Logic

In 2006, Alexandru Baltag and Sonja Smets provide a Kripke-model based, qualitative, multi-agent version of the classical Belief Revision theory, which is called the logic of conditional beliefs [6]. The Kripke-style models for the logic of conditional beliefs is cast in pure qualitative terms. The qualitative description is given in terms of conditional doxastic maps, which can be seen as labelled accessibility relations in a given Kripke model. Every accessibility relation is labelled by propositions that capture the new incoming information with which the agent can revise her beliefs. These maps are then used in a natural way to give an interpretation to the conditional doxastic belief operators in the object language of the logic. Here we only consider the single-agent case in which the agent receives new factual information only.

Syntax The language of Conditional Doxastic Logic (*CDL*) is defined as follows.

2.6.1. DEFINITION. The formal language $\mathcal{L}_{CDL}$ is built up from a countable set of propositional atoms, the usual connectives and a conditional belief operator. Let p range over a set of propositional atoms:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \psi \mid B^{\varphi}\psi$$

The intended interpretation of $B^{\varphi}\psi$ is "the agent believes that ψ conditional on φ".

We can define a knowledge and belief modality as an abbreviation:

$$K\varphi := B^{\neg\varphi}\bot$$
$$B\varphi := B^{\top}\varphi$$

Semantics We interpret $\mathcal{L}_{CDL}$ in the states of conditional doxastic models.

2.6.2. DEFINITION. A *conditional doxastic frame* is a tuple $(S, \{B^P\}_{P\subseteq S})$, where S is a non-empty set of worlds, and for each proposition $P \subseteq S$, we have a map $B^P : S \longrightarrow \mathcal{P}(S)$, satisfying the following conditions[9]: for each $w \in S$,

- if $w \in P$, then $B^P(w) \neq \varnothing$
- if $Q \cap B^P(w) \neq \varnothing$, then $B^P(w) \neq \varnothing$
- if $w' \in B^P(w)$, then $B^Q(w) = B^Q(w')$
- $B^P(w) \subseteq P$
- if $B^P(w) \cap Q \neq \varnothing$, then $B^{P\cap Q}(w) = B^P(w) \cap Q$

The first condition states that beliefs are consistent if the information received is true. The second one states that beliefs are consistent as long as new information is not in contradiction with the old information. Condition 3 states that beliefs are introspective. The fourth one states that the new information is believed. Condition 5 states that revision must be minimal in the sense the agent keeps as much as possible of her previous beliefs when she receives a new piece of information.

In [6], Baltag and Smets went on to show that any such conditional doxastic frame is equivalent to a semantic *AGM* theory[10] over a *KB*-frame.

[9]Note that, for each $P \subseteq S$, B^P can be equivalently defined as a binary relation over S, satisfying the corresponding conditions.

[10]Originally, the *AGM* theory is set in a syntactic framework.

2.6.3. Definition. A *KB-frame* is a tuple, (S, B, K) where S is a non-empty set, and the two maps $K : S \longrightarrow \mathcal{P}(S)$, $B : S \longrightarrow \mathcal{P}(S)$, satisfy the following conditions:

- $w \in K(w)$;
- if $w' \in K(w)$, then $B(w) = B(w'), K(w) = K(w')$;
- $B(w) \subseteq K(w)$
- $B(w) \neq \varnothing$

The first condition expresses the truthfulness of knowledge. The second one expresses full introspection (an agent knows what she knows/believes and what not). The third one says that the agent believes everything she knows, and the fourth one says that beliefs are consistent.

We now provide a semantic version of the AGM theory over a KB-frame. To this end, we need to consider the semantic counterparts of different syntactic notions, eg. theories, sentences and others. An S-theory is taken to be a set of states and an S-sentence is also considered to be a set of states. Note that each S-theory $A \subseteq S$ gives rise to a deductive closed set of sentences $Th_A = \{\varphi \in \mathcal{L} \mid i \vDash_S \varphi \text{ for all } i \in A\}$. We also assume that the belief sets form S-theories. The inconsistent theory $\perp$ can be represented by the empty set $\varnothing \subseteq S$. The deductive closure of the union of two syntactic theories corresponds to the intersection of the respective semantic theories that is, sets of states. Thus, an expansion $A + Y$ of a semantic theory $A \subseteq S$ with a semantic sentence $Y \subseteq S$ is given by the intersection $A \cap Y$.

2.6.4. Definition. Given a KB-frame $\langle S, B, K\rangle$, let $\mathbb{T} \subseteq \mathcal{P}(S)$ be a family of S-theories. Now we define an operation $* : \mathbb{T} \times \mathcal{P}(S) \longrightarrow \mathbb{T}$ be such that, for all $Y \subseteq S$, we have that:

($\mathbb{T}$1.) $B(i) \in \mathbb{T}$ for each $i \in S$.

($\mathbb{T}$2.) $\varnothing \in \mathbb{T}$.

($\mathbb{T}$3.) if $A \in \mathbb{T}$, then for all $i, j \in A$, $B(i) = B(j)$.

($*$1.) $A * Y \in \mathbb{T}$;

($*$2.) $A * Y \subseteq Y$;

($*$3-4.) $A * S = S$;

($*$5.) $A * Y = \varnothing$ iff $K(A) \cap Y = \varnothing$;

($*$6.) if $Y = Z$, then $A * Y = A * Z$;

($*$7-8.) if $(A * Y) \cap Z \neq \varnothing$, then $A * (Y \cap Z) = (A * Y) \cap Z$.

We should mention here that postulate ($*5$) is modified from the original syntactic *AGM* version. Since an agent's beliefs about her beliefs or knowledge are certain, they should not be revised.

2.6.5. FACT. Baltag and Smets went on to show that any such conditional doxastic frame is equivalent to a semantic *AGM* theory over a *KB*-frame.

2.6.6. DEFINITION. A *conditional doxastic model* is a Kripke model whose underlying frame is a conditional doxastic frame.

Axiomatics Baltag and Smets provide a sound and complete proof system for *CDL* in [6].

Dynamics Baltag and Smets also investigate belief update, focusing on public announcements. The mechanism of update is the same as the one in *DEL*: an update with $!P$ changes the model by eliminating all the $\neg P$-worlds. The result of this elimination is a submodel only consisting of P-worlds.

The syntax of *CDL* is then extended with dynamic modalities $\langle !P \rangle$. Reduction axioms for public announcements are added to the axioms of *CDL* to obtain a sound and complete proof system.

We now move on to a setting dealing explicitly with evidence and justification for belief.

2.7 Evidence Logic

In [16], Johan van Benthem and Eric Pacuit provide a semantic approach to evidence. They develop a very interesting extension of *DEL* aimed to deal with evidential dynamics. Their evidence models are based on the well-known neighbourhood semantics for modal logic, in which the neighbourhoods are interpreted as evidence sets: pieces of evidence (possibly false, possibly mutually inconsistent) possessed by the agent. Indeed, the evidence setting deals with inconsistent evidence. Thus an agent can have several non compatible pieces of evidence. It means that in their evidence sets, some subsets (representing pieces of evidence) are disjunct sets (their intersection is empty). The plausibility relation that can be induced on states in evidence models is not a total pre-order. That means that not all states are comparable. As a result, the action of belief revision in neighbourhood models does not satisfy the *AGM* postulates (in particular the postulates ($*$7-8. iterated $*$3-4.) in the Definition 2.1.3 will fail).

2.7.1. DEFINITION. A pointed *evidence model* $\mathcal{M}$ is a tuple $(S, E, \|\cdot\|, s_0)$ consisting of:

- a non-empty set of worlds S,
- an evidence relation $E \subseteq S \times \mathcal{P}(S)$,
- a standard valuation function $\|\cdot\|$
- an actual state of affairs s_0.

A pointed uniform evidence model is an evidence model with a constant function E.

2.7.2. Definition. The *collection of evidence sets* is defined as

$$E(s) = \{X | sEX, X \subseteq S\}$$

and we impose two constraints on the evidence function:

$$\text{(Cons) For each state s, } \varnothing \notin E(s)$$

$$\text{(Triv) For each state s, } S \in E(s)$$

These constraints ensure that no evidence set is empty and that the universe S is itself an evidence set.

In this framework, the combination of different evidence sets does not necessarily yield consistent evidence. Indeed for any two evidence sets X and Y, X and Y may be disjoints sets that is, $X \cap Y = \varnothing$.

2.7.3. Definition. Van Benthem and Pacuit introduce the notion of *s-scenario*: a s-scenario is a maximal collection $\chi \subseteq E(s)$ that has the *finite intersection property* (f. i. p.). χ has the *f. i. p.* if for each finite subfamily $X \subseteq \chi$, $\bigcap X \neq \varnothing$. A s-scenario relative to P is a maximal collection $\chi \subseteq E(s)$ that has the *P-finite intersection property* (f. i. p.). χ has the *P-f. i. p.* if for each finite subfamily $X \subseteq \chi^P$, $\bigcap X \neq \varnothing$ with $\chi^P = \{Y \cap P \mid Y \in \chi\}$.

Epistemic and doxastic notions We can now formally define the notions of irrevocable knowledge (K), belief (B), conditional belief (B^-), evidence ($\boxminus$) and conditional evidence ($\boxminus^-$) in evidence models.

2.7.4. Definition. Irrevocable knowledge is formally defined as truth in all possible worlds, i.e.

$$KP := \{s \in S : P = S\}$$

Belief is defined as:

$$BP := \{s \in S : \text{ for some } s\text{-scenario } \chi \subseteq E(s) \text{ and } \forall t \in \bigcap \chi (t \in P)\}$$

Conditional belief is defined as:

$$B^Q P := \{s \in S : \text{ for some } s\text{-scenario } \chi \subseteq E(s) \text{ and } \forall t \in \bigcap \chi^Q (t \in P))\}$$

Evidence is defined as:

$$\boxminus P := \{s \in S : sEX \text{ for } \varnothing \neq X \subseteq S \text{ and } \forall t \in X(t \in P)\}$$

Conditional evidence is defined as:

$$\boxminus^Q P := \{s \in S : sEX \text{ for } \varnothing \neq X \cap \|Q\| \subseteq S \text{ and } \forall t \in X \cap \|Q\|(t \in P)\}$$

The interpretation of $\boxminus P$ is the agent has evidence for P.

Note that these definitions come from [16]. When the evidence model $\mathcal{M}$ is not finite, these definitions can lead to inconsistent beliefs, that is, the agent might believe a contradiction ($B\bot$). There are several solutions for this problem:

– the first solution is the solution we adopt in Chapter 6 where we change the definition of belief to make belief globally consistent;

– the second solution is the solution adopted by van Benthem, Fernández-Duque and Pacuit in [14] where they introduce the notion of flateness. An evidence model $\mathcal{M}$ satisfies the axiom D, that is, a flat evidence frame is serial. Throughout most of this thesis (in Part II and Chapter 7), we adopt a more neutral solution, we assume finite models[11].

Dynamics of evidence When dealing with evidence dynamics, van Benthem and Pacuit suggest two kinds of operations modifying evidence: external and internal operations. Evidence change can be triggered by a piece of new incoming information or can be the result of an internal process of re-evaluation. We only present here "hard information change" (public announcement) and evidence combination while van Benthem and Pacuit study also "soft information change" (radical and conservative upgrade).

2.7.5. Definition. When new hard evidence φ is received, this induces an *update* $!\varphi$, which changes the agent's prior evidence model $(S, E, \|\cdot\|)$ to $(S', E', \|\cdot\|)$ with: $S' = \|\varphi\|$ and $E'(s) = \{X | \varnothing \neq X = Y \cap \|\varphi\| \text{for some Y} \in E(s)\}$.

2.7.6. Definition. The basic internal operation that van Benthem and Pacuit deal with is *evidence combination*: in this case an agent combines consistent evidence. The evidence model is changed to $(S^*, E^*, \|\cdot\|)$ with: $S^* = S$ and $E^*(s)$ is the smallest set closed under (non-empty) intersection and containing $E(s)$.

[11] Note that every finite model is flat.

Evidence models and partial plausibility models Van Benthem and Pacuit compare their evidence models with partial plausibility models. Every partial plausibility model can be extended to an evidence model defining evidence sets as the downward $\leq$-closed sets of worlds. Let $X_{\downarrow\leq} = \{t \in S \mid \exists x \in X \text{ and } t \leq x\}$ for $X \subseteq S$, then X is $\leq$-closed if $X_{\downarrow\leq} \subseteq X$. Conversely evidence models with a constant function E (such a type of evidence models is said to be *uniform*) can be turned into partial plausibility models such that $s \leq_E t$ iff $\forall X \in E, t \in X$ implies $s \in X$. They note that in partial plausibility models the agent has already combined all of her evidence.

Proof system The proof system for the logic of evidence is given by the minimal modal logic for the separate dynamic modalities (given the usual rules of Necessitation and Replacement of Provable Equivalents) and the following Recursion axioms:

2.7.7. Definition.

$$
\begin{aligned}
[!\varphi]p &\iff (\varphi \to p),\\
[!\varphi](\psi \wedge \chi) &\iff [!\varphi]\psi \wedge [!\varphi]\chi,\\
[!\varphi]\neg\psi &\iff \varphi \to \neg[!\varphi]\psi,\\
[!\varphi]B^{\psi}\theta &\iff \varphi \to B^{\varphi\wedge[!\varphi]\psi}([!\varphi]\theta),\\
[!\varphi]\boxminus^{\psi}\theta &\iff \varphi \to \boxminus^{\varphi\wedge[!\varphi]\psi}([!\varphi]\theta),
\end{aligned}
$$

Till now we have focused on settings with a dynamic dimension, now we move on to a setting with a temporal dimension. Indeed Giacomo Bonanno chooses to use a temporal (modal) logic to model belief change.

2.8 Branching time temporal logics of belief revision

In 2005, Giacomo Bonanno studies a branching time temporal logic of belief revision [17] and develops his setting through a series of papers [19, 18, 21, 20, 22]. In these papers, he models the interaction between information and beliefs over time using a multimodal logic. In his setting, Bonanno only considers agents who receive new factual information, that is information about facts. We consider again a single agent facing new incoming information that triggers a belief change. To model different types of belief change, Bonanno introduces three different logics of increasing strength. For each logic he provides a set of axioms as well as the corresponding property characterizing the axiom in question.

The first logic is called the weakest logic of belief revision L_W. Bonanno introduces L_W in [21] and in this setting he only considers information confirming initial beliefs. Thus this logic captures a very weak notion of belief revision. The second logic is called the logic of Qualitative Bayes Rule (L_{QBR}), which is stronger than the previous logic. L_{QBR} considers new information which does not contradict initial beliefs. Finally the last one is called the logic of AGM (L_{AGM}), which is stronger than the two previous logics since it considers new information contradicting initial beliefs.

In his latest papers [20, 22], Bonanno identifies the condition on his branching-time belief revision frames that is equivalent to the property of AGM-consistency. He adds this property (and the corresponding axiom) to the properties (and axioms) of the logic (L_{AGM}), defining then a new logic: the logic PLS (L_{PLS})[12]. But in his very last paper [22], he changes some of the properties and corresponding axioms making them simpler to handle. So the logic PLS becomes the logic $PLS*$ (L_{PLS*}).

In what follows, we provide an introduction to his basic setting as well as his belief expansion logic L_{QBR} [21] and his last belief revision setting namely the logic called L_{PLS*} [22].

Syntax The language for his branching-time belief logic is an extension of the classical propositional language.

2.8.1. DEFINITION. The formal language $\mathcal{L}_B$ is built up from a countable set of propositional atoms Φ, the usual propositional connectives and five unary modal operators, namely the temporal operators $\bigcirc$ (next instant) and $\bigcirc^{-1}$ (previous instant), the universal modality A, a belief operator B and an information operator I restricted to Boolean formulas that is, $I\varphi$ is a well formed formula iff φ is Boolean. Let p range over Φ.

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \psi \mid \bigcirc\varphi \mid \bigcirc^{-1}\varphi \mid B\varphi \mid I\varphi \mid A\varphi$$

We use the following abbreviations for the dual operators: $\neg\bigcirc^{-1}\neg\varphi := \Diamond^{-1}\varphi$ and $\neg\bigcirc\neg\varphi := \Diamond\varphi$.

The intended interpretation of the operators is as follows:

$\bigcirc\varphi$: at every next instant it will be the case that φ

$\bigcirc^{-1}\varphi$: at the previous instant it was the case that φ

$B\varphi$: the agent believes that φ

[12]Bonanno does not give a name to his new logic, we name it in reference to its new property PLS.

$I\varphi$: the agent is informed that φ

$A\varphi$: φ is true at every world

$\Diamond^{-1}\varphi$: there exists a previous instant at which φ is the case

$\Diamond\varphi$: there exists a next instant at which φ is the case

Semantics Starting with the semantics, Bonanno uses temporal (branching-time) possible world models. As standard in modal logic, such models are based on an underlying Kripke frame.

2.8.2. DEFINITION. We define a *next-time branching frame* $(T, \rightsquigarrow)$, consisting of a countable set of time points or instants T and an "immediate successor" relation $\rightsquigarrow$ on T satisfying the following conditions: $\forall t, u, v \in T$,

1. If $t \rightsquigarrow v$ and $u \rightsquigarrow v$ then $t = u$.

2. If $< t_1, ..., t_n >$ is a sequence with $t_i \rightsquigarrow t_{i+1}$, for every $i = 1, ...n - 1$, then $t_n \neq t_1$.

Condition 1 makes sure that each instant has a unique predecessor and condition 2 excludes cycles in the structure, giving it a tree-form. The intended interpretation of $t \rightsquigarrow u$ is taken to be "u is an immediate successor of t or t is the immediate predecessor of u". Each instant can have several immediate successors. Let $t^{\rightsquigarrow}$ denote the set of all immediate successors of t.

2.8.3. DEFINITION. We introduce a *branching-time belief frame* to be a Kripke frame,

$$(T, \rightsquigarrow, S, \{B_t, I_t\}_{t \in T})$$

where:

– $(T, \rightsquigarrow)$ is a next-time branching frame,

– S is a non-empty set of states,

– B_t is a binary relation on S capturing the beliefs of an agent at t,

– I_t is a binary relation on S modelling the information an agent can receive at t.[13]

Casting the relations in terms of maps, we set:

$$I_t(i) = \{j \in S : iI_tj\}$$
$$B_t(i) = \{j \in S : iB_tj\}$$

[13]It might be customary to think of the belief relation as a *KD45* relation in modal logic and of the information relation as *S4* or *S5* but note that Bonanno leaves these options open.

2.8.4. Example. As an example of a branching-time belief frame we introduce Figure 2.10. The rectangles correspond to the information sets at every instant, whereas the ovals inside them correspond to the belief sets of the agent. In Figure 2.10, we have that $I_t(i) = \{i, j, k, l\}$ and $B_t(i) = \{j, k\}$.

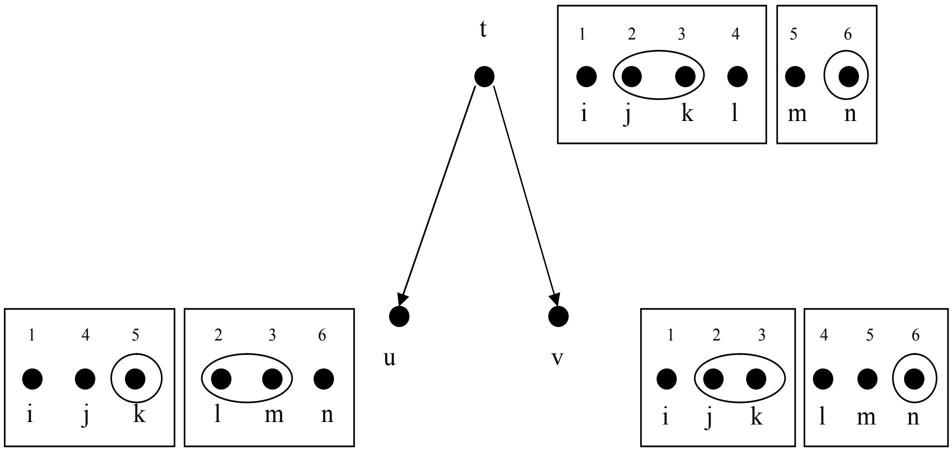

Figure 2.10: Example of a branching-time belief frame

2.8.5. Definition. The *branching-time belief expansion frames* are branching-time belief frames that satisfy the following conditions: $\forall i \in S$ and $\forall t, u \in T$,

1. if $t \rightsquigarrow u$ and $B_t(i) \cap I_u(i) \neq \varnothing$ then $B_u(i) \subseteq B_t(i)$
2. if $t \rightsquigarrow u$ then $B_t(i) \cap I_u(i) \subseteq B_u(i)$
3. if $t \rightsquigarrow u$ and $B_t(i) \cap I_u(i) \neq \varnothing$ then $B_u(i) \subseteq I_u(i)$

2.8.6. Example. As an example of a belief expansion scenario we introduce the following story and we represent the corresponding branching-time belief expansion frame in Figure 2.11. Consider an agent and a dice. Someone throws the dice such that our agent cannot see the upper face. We have 6 possible worlds in our belief expansion frame: i where 1 is the upper face, j where 2 is the upper face and so on. Assume that the agent initially believes that the upper face is 2, 3, 4, 5 or 6 while in reality (unknown to our agent) the upper face is 3. So the agent considers j, k, l, m and n to be the most plausible worlds at the initial world-instant pair (k, t). Then she receives the information that the number on the upper face is odd at the world-instant pair (k, u). So the agent is informed that i, k and m are possible and according to our belief expansion rules she will now come to believe that the upper face is 3 or 5. In our terms, the agent considers k and m to be the most plausible worlds at the world-instant pair (k, u). Consider now the alternative scenario in which the agent receives the information that the number on the upper face is smaller than 4 at the world-instant pair (k, v). In this case, the agent is informed that i, j and k are possible and according to our belief expansion rules she will come to believe that the upper face is 2 or 3. So the agent considers j and k to be the most plausible worlds at the world-instant pair (k, v).

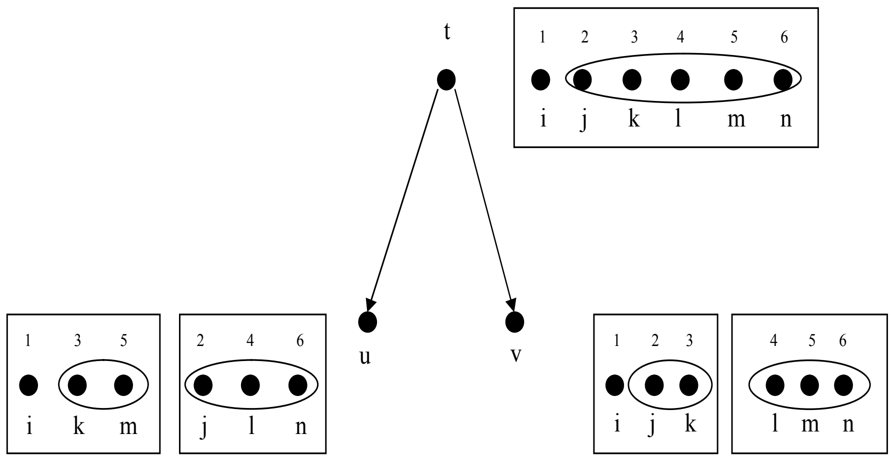

Figure 2.11: Example of a branching-time belief expansion frame

2.8.7. Definition. The *branching-time belief revision frames* are branching-time belief frames in which S is finite and that satisfy the following properties: $\forall i \in S$, $u_0, u_1, \ldots, u_n \in t^{\rightsquigarrow}$ with $u_0 = u_n$ and $\forall k = 1, \ldots, n$

1. if $t \rightsquigarrow u_0$ and $B_t(i) \cap I_{u_0}(i) \neq \varnothing$ then $B_{u_0}(i) = B_t(i) \cap I_{u_0}(i)$
2. $B_t(i) \subseteq I_t(i)$
3. if $t \rightsquigarrow u_0$, $t \rightsquigarrow u_1$ and $I_{u_0}(i) = I_{u_1}(i)$ then $B_{u_0}(i) = B_{u_1}(i)$
4. $B_t(i) \neq \varnothing$
5. if $I_{u_{k-1}}(i) \cap B_{u_k}(i) \neq \varnothing$, then $I_{u_{k-1}}(i) \cap B_{u_k}(i) = B_{u_{k-1}}(i) \cap I_{u_k}(i)$

2.8.8. Example. As an example of a belief revision scenario we introduce the following story and we represent the corresponding branching-time belief revision frame in Figure 2.12. Consider again an agent and a dice. Someone throws the dice such that the agent cannot see the upper face. We have 6 possible worlds in our belief expansion frame: i where 1 is the upper face, j where 2 is the upper face and so on. Assume that the agent initially believes that the upper face is 4, 5 or 6 while in reality (unknown to our agent) the upper face is 3. So the agent considers l, m and n to be the most plausible worlds at the initial world-instant pair (k, t). Then she receives the information that the number on the upper face is a prime number at the world-instant pair (k, u). So the agent is informed that j, k and m are possible and according to our belief revision rules she will come to believe that the upper face is 5. In our terms, the agent considers m to be the most plausible worlds at the world-instant pair (k, u). Consider now the alternative scenario in which the agent receives the information that the number on the upper face is smaller or equal to 4 at the world-instant (k, v). So the agent is informed that i, j, k and l are possible and according to our belief expansion rules she will come to believe that the upper face is 4. So the agent considers l to be the most plausible world at the world-instant pair (k, v).

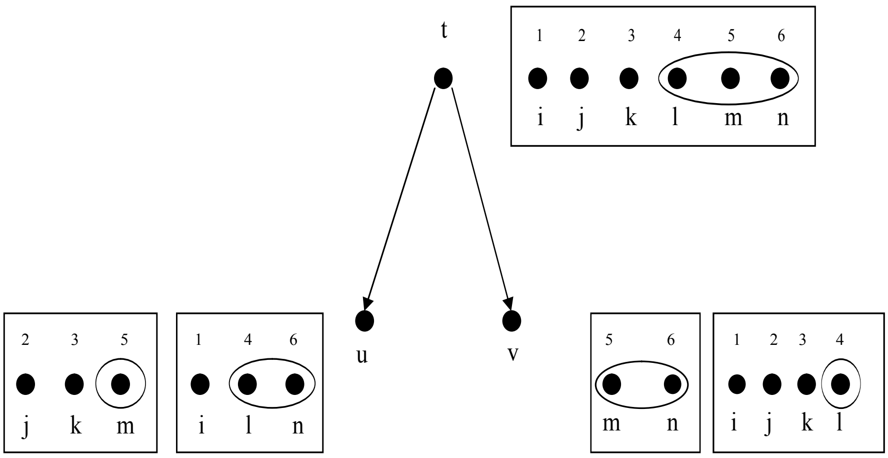

Figure 2.12: Example of a branching-time belief revision frame

2.8.9. FACT. In [22], Bonanno proves that all the branching-time belief revision frames satisfying the above properties, satisfy the AGM postulates.

2.8.10. DEFINITION. A *branching-time belief model* is a Kripke model based on a branching-time belief frame and is obtained by adding a valuation $V : \Phi \to \mathcal{P}(S)$ assigning to each atomic sentence p from a given set of atomic sentence Φ, the set of states in which the sentence is true.

In the same way, a *branching-time belief expansion model* is a Kripke model based on a branching-time belief expansion frame extended with a valuation map. And a *branching-time belief revision model* is a Kripke model based on a branching-time belief revision frame extended with a valuation map.

2.8.11. DEFINITION. In a branching-time belief (respectively revision or expansion) model, the valuation map can be extended to arbitrary well formed sentences

φ in our language, where we use the standard notation in modal logic for satisfaction in a world-time point in a given model[14] $\mathcal{M}$ and denote this as $\mathcal{M},(i,t) \models \varphi$ or equivalently as $(i,t) \in \| \varphi \|_{\mathcal{M}}$.

$\mathcal{M},(i,t) \models p$ iff $i \in V(p)$.

$\mathcal{M},(i,t) \models \neg\varphi$ iff $\mathcal{M},(i,t) \not\models \varphi$.

$\mathcal{M},(i,t) \models \varphi \vee \psi$ iff $\mathcal{M},(i,t) \models \varphi$ or $\mathcal{M},(i,t) \models \psi$.

$\mathcal{M},(i,t) \models \bigcirc\varphi$ iff for all u such that $t \rightsquigarrow u$, $\mathcal{M},(i,u) \models \varphi$.

$\mathcal{M},(i,t) \models \bigcirc^{-1}\varphi$ iff for all u such that $u \rightsquigarrow t$, $\mathcal{M},(i,u) \models \varphi$.

$\mathcal{M},(i,t) \models I\varphi$ iff $\big(j \in I_t(i)$ iff $\mathcal{M},(j,t) \models \varphi\big)$.

$\mathcal{M},(i,t) \models B\varphi$ iff $\big(j \in B_t(i)$ implies $\mathcal{M},(j,t) \models \varphi\big)$.

$\mathcal{M},(i,t) \models A\varphi$ iff for all $j \in S, \mathcal{M},(j,t) \models \varphi$.

It follows that:

$\mathcal{M},(i,t) \models \Diamond\varphi$ iff for at least one u such that $t \rightsquigarrow u$, $\mathcal{M},(i,u) \models \varphi$.

$\mathcal{M},(i,t) \models \Diamond^{-1}\varphi$ iff for at least one u such that $u \rightsquigarrow t$, $\mathcal{M},(i,u) \models \varphi$.

Axiomatics

2.8.12. DEFINITION. A complete axiomatization of the basic logic is given in [83]. The axioms and rules for the basic logic are given as follows.

a) All propositional tautologies.

b) Kripke's axiom K for $\bigcirc, \bigcirc^{-1}, B, A$.

$$\bigcirc(\varphi \to \psi) \to (\bigcirc\varphi \to \bigcirc\psi)$$
$$\bigcirc^{-1}(\varphi \to \psi) \to (\bigcirc^{-1}\varphi \to \bigcirc^{-1}\psi)$$
$$B(\varphi \to \psi) \to (B\varphi \to B\psi)$$
$$A(\varphi \to \psi) \to (A\varphi \to A\psi)$$

c) Temporal axioms.

$$\varphi \to \bigcirc \Diamond^{-1} \varphi$$
$$\varphi \to \bigcirc^{-1} \Diamond \varphi$$

d) Backward uniqueness axiom.

[14]The model $\mathcal{M}$ can be a branching-time belief model or a branching-time belief expansion model or a branching-time belief revision model.

$\Diamond^{-1}\varphi \to \bigcirc^{-1}\varphi$

e) S5 axioms for A.

$$A\varphi \to \varphi$$
$$A\varphi \to AA\varphi$$
$$\neg A\varphi \to A\neg A\varphi$$

f) Inclusion axiom for B.

$$A\varphi \to B\varphi$$

g) Axioms for the information operator I.

$$(I\varphi \wedge I\psi) \to A(\varphi \leftrightarrow \psi)$$
$$A(\varphi \leftrightarrow \psi) \to (I\varphi \leftrightarrow I\psi)$$

h) Inference rules.

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

$$\frac{\varphi}{\bigcirc\varphi} \text{ (}\bigcirc\text{-necessitation)}$$

$$\frac{\varphi}{\bigcirc^{-1}\varphi} \text{ (}\bigcirc^{-1}\text{-necessitation)}$$

$$\frac{\varphi}{A\varphi} \text{ (}A\text{-necessitation)}$$

These axioms and inference rules state that the operators B, A, $\bigcirc$ and $\bigcirc^{-1}$ are normal operators such that the operator A is an $S5$ modal operator, whereas the operator I is a non-normal operator[15]. In this setting an agent can only receive one piece of information at a time such that if two formulas are equivalent then, if this agent is informed about one of the two formulas, it means that she is also informed about the other one. As Bonanno points out in his work, the information operator is different but bears similarities to the "all and only" operator of Humberstone [43], the "only knowing" operator of Levesque [48] and the "assumption operator" of Brandenburger and Keisler [24]. The Inclusion axiom states that if a formula is necessarily true then the agent believes it. Note that the branching-time structure is such that each instant has a unique immediate predecessor but can have several immediate successors. This axiom system forms the basis from which we start adding axioms to deal specifically with the cases of belief expansion and belief revision.

[15]Necessitation for B can be derived from Modus Ponens, Inclusion axiom for B and Necessitation for A.

2.8.13. Definition. A complete axiomatization of the belief expansion logic L_{QBR} is given by starting from the basic axioms and rules and by adding the following axioms [21].

a) No Drop (ND).

$$(\neg B \neg \varphi \wedge B\psi) \rightarrow \bigcirc(I\varphi \rightarrow B\psi)$$

b) No Add (NA).

$$\neg B \neg(\varphi \wedge \neg\psi) \rightarrow \bigcirc(I\varphi \rightarrow \neg B\psi)$$

c) Qualified Acceptance (QA).

$$\neg B \neg \varphi \rightarrow \bigcirc(I\varphi \rightarrow B\varphi)$$

The axiom No Drop states that if the information received does not contradict the initial beliefs, an agent does not drop her initial beliefs. The axiom No Add states that if the information received does not contradict the initial beliefs, an agent does not add a belief about which she is not informed. Note that together, these axioms imply that the information received is also consistent. The axiom Qualified Acceptance states that if the information received does not contradict the initial beliefs, an agent believes the new information.

2.8.14. Fact. These axioms characterize the properties of Definition 2.8.5. Note that Bonanno introduces the property which he calls the Qualitative Bayes Rule in [17, 21]: if there exists an instant t such that $t \rightsquigarrow u$ and if $B_t(i) \cap I_u(i) \neq \emptyset$, then $B_u(i) = I_u(i) \cap B_t(i)$. Bonanno uses the term "Qualitative Bayes Rule" since he relates in [17] this property to the Bayes' rule used in the economics and game theory literature to model belief revision in a probabilistic setting. Then he proves that the conjunction of axioms ND, NA and QA characterizes this property.

2.8.15. Definition. A complete axiomatization of the belief revision logic L_{PLS*} is given by starting from the basic axioms and rules and by adding the following axioms [22].
Let $\bigwedge_{j=1,\ldots,n} \varphi_j$ denote the formula $(\varphi_1 \wedge \ldots \wedge \varphi_n)$ and let $\varphi_0 = \varphi_n$ and $\chi_0 = \chi_n$

a) No Drop (ND).

$$(\neg B \neg \varphi \wedge B\psi) \rightarrow \bigcirc(I\varphi \rightarrow B\psi)$$

b) No Add (NA).

$$\neg B \neg(\varphi \wedge \neg\psi) \rightarrow \bigcirc(I\varphi \rightarrow \neg B\psi)$$

c) Acceptance (A).

$$I\varphi \to B\varphi$$

d) Equivalence (EQ).

$$\Diamond(I\psi \wedge B\varphi) \to \bigcirc(I\psi \to B\varphi)$$

e) PLS.

$$\bigwedge_{j=1,\dots,n} \Diamond(I\varphi_j \wedge \neg B\neg\varphi_{j-1} \wedge B\chi_j) \to \bigwedge_{j=1,\dots,n} \bigcirc((I\varphi_j \to B(\varphi_{j-1} \to \chi_{j-1})) \wedge (I\varphi_{j-1} \to B(\varphi_j \to \chi_j)))$$

f) Consistency (C).

$$B\varphi \to \neg B\neg\varphi$$

The explanation of the axioms No drop and No Add is given above (see Definition 2.8.13). The axiom Acceptance states that the new information is believed. The axiom Equivalence says that differences in beliefs must be due to differences in information. The axiom PLS states that beliefs must be rationalized. The axiom Consistency says that beliefs are consistent.

2.8.16. Fact. As shown in Bonanno [22], these axioms characterize the properties of Definition 2.8.7.

Extension to higher-order information While Bonanno's work is fully developed in the sense that this approach allows one to model the AGM-rules of belief revision in a branching-time setting, the main ideas have till now been mainly pursued in a single-agent context. And in a single agent context, the investigation of higher-order beliefs isn't an issue of primary concern. However this framework can be extended to higher-order information. Previous work on an extension of Bonanno's setting, in an attempt to deal with higher-order beliefs, was provided by Zvesper in [83] where he lifts the Boolean restriction on I-formulas and provides as such an adjustment of the framework. According to Zvesper, this adjustment makes the framework more comparable with the way in which information dynamics is dealt with in Public Announcement Logic.

Zvesper [83] proposed an extension of Bonanno's framework in order to eliminate the Boolean restriction on the I operator. His proposal provides a new interpretation of the branching-time belief models. If an agent receives a piece of new information at an instant t, she now revises her beliefs at an instant u such that u is an immediate successor of t. This change in interpretation captures the idea that the information received describes the state of the world as it was before the receipt of this information.

Zvesper extended the basic setting of Bonanno with additional axioms which, in the restricted case of belief expansion (when the incoming information is consistent with the existing beliefs) can handle the same type of information flow or informational dynamics that we encounter in Public Announcement Logic (abbreviated as *PAL*, see Section 2.4).

2.8.17. DEFINITION. To this end, Zvesper adds some *additional axioms* to the basic set of axioms and rules presented in Definition 2.8.12:

a) Temporal axioms for the atomic propositions.

$$\Diamond p \to p$$
$$\Diamond^{-1} p \to p$$
$$p \to \bigcirc p$$
$$p \to \bigcirc^{-1} p$$

b) Temporal axioms for A.

$$\Diamond A\varphi \to A \Diamond \varphi$$
$$\Diamond^{-1} A\varphi \to A \Diamond^{-1} \varphi$$

c) UA.

$$I\varphi \to AI\varphi$$

d) NM.

$$I\varphi \to (\Diamond B \Diamond^{-1} \psi \to B(\varphi \to \psi))$$

e) PR.

$$I\varphi \to (B(\varphi \to \psi) \to \bigcirc B \bigcirc^{-1} \psi)$$

The axioms (a) and (b) respectively state that facts do not change over time and the states of the model are constant over time which was however implicitly present in the logic of Bonanno. Zvesper's aim in [83] was to show the similarities between the type of dynamics that is encoded in "public announcement actions" (as in *PAL*) and the way we can interpret the I-operator in a branching-time framework. Viewing I as some type of announcement operator, Zvesper uses UA to abbreviate "uniform announcements". In our view this axiom is intended to capture the "public nature" of these announcements. The axiom NM is an abbreviation for "no miracles" indicating that the only way an agent can change her mind is when it is triggered by an announcement. The axiom PR stands for "perfect recall" and indicates intuitively that these agents have a perfect memory.

2.8.18. Definition. Zvesper calls "*public announcement temporal doxastic frames*" the branching-time belief frames that satisfy these two semantic properties:

(1) $I_t(i) = I_t(j)$.

(2) if there exists an instant t such that $t \rightsquigarrow u$, then $B_u(i) = I_t(i) \cap B_t(i)$.

He then further claims that UA, NM and PR are the axioms that characterize exactly these frames.

Note that more general results about the relation between temporal models and the dynamic models of *DEL* (including *PAL*) have recently been studied in [15], while [27, 13] present an extension of these results for belief revision in plausibility models. We will not review these results here in detail because in the chapter where we deal with the temporal framework, we will mainly refer back to the above explained branching-time belief revision setting of Bonanno.

Part II

Developing Soft Dynamic Epistemic Logic

Chapter 3

Belief contraction in total plausibility models

Aim In this chapter we study belief contraction in the framework of Dynamic Epistemic Logic. Our aim is to model different belief contraction operations in the particular setting of total plausibility models.

Summary: In this chapter we consider three different kinds of belief contraction operations. We first present the notions of severe withdrawal, conservative contraction and moderate contraction. Then we axiomatize these *AGM*-friendly versions of contraction in *DEL*. The main points are:

– we provide a brief presentation of the notions of belief contraction in the literature.

– we present the notion of severe withdrawal and discuss the limits of this belief contraction operation and we present two other types of contraction operations namely, conservative and moderate contraction.

– we define these three contracting operations as operations on total plausibility models, we associate to them epistemic actions in *DEL* style and axiomatize them in *DEL* style.

Background In this chapter we use both Grove spheres and (finite) total plausibility models as described in Chapter 2. These types of models are in fact equivalent. Thus we can generate a Grove model from a given total plausibility model and vice versa.

We illustrate this correspondence on the following example. Let us start from a given Grove model in Figure 3.1. This model contains four nested spheres. Since the central sphere represents the actual belief set of the agent and the world s belongs to the central sphere, s is the most plausible world in this model. Since the second sphere contains the world v and the third sphere contains the world t,

v is more plausible than t. Since both u and w belong to the last sphere, u and w are equiplausible and are the less plausible worlds in this model.

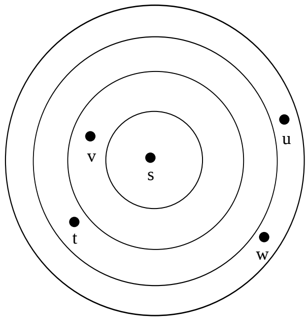

Figure 3.1: Example of a sphere system

From these observations, we can easily draw the corresponding total plausibility model in Figure 3.2.

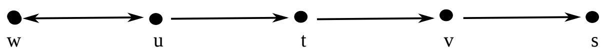

Figure 3.2: Corresponding total plausibility model

In this chapter we will define different dynamic languages on top of two different static languages which we introduced in Chapter 2:

– the language $\mathcal{L}_{CDL}$ is obtained from the language of propositional logic by the addition of a conditional belief operator, and

– the language $\mathcal{L}_{KK_D}$ is obtained from the language of propositional logic by the addition of an irrevocable knowledge operator and a defeasible knowledge operator.

3.1 General background on belief contraction

We focus here on the notion of belief contraction. If an agent may revise her beliefs after receiving a new piece of information, she may also contract her beliefs in the light of new information. For example, consider an agent who believes that Aristotle is lying on my couch. Then this agent receives the information that it is not the case that Aristotle is lying on my couch. After receiving this information she no longer believes that Aristotle is lying on my couch. In other words, the agent removes her belief that Aristotle is lying on my couch. Indeed she considers it now possible that Aristotle is lying on my couch but also that Aristotle is not lying on my couch.

We already mentioned in Chapter 2 the work of Alchourrón, Gärdenfors and Makinson who provide some postulates for belief contraction in [1]. In this chapter, we are only interested in the AGM-friendly versions of belief contraction.

3.1.1 Severe withdrawal

The first notion of contraction we want to study is the notion called "mild contraction" by Levi [50], "severe withdrawal" by Pagnuco and Rott [63] and "Rott contraction" by Ferme and Rodriguez [31].

3.1.1. Definition. *Severe withdrawal* is a belief change operation that removes a belief from the belief set of an agent such that, after contracting this belief set with φ, the most plausible worlds are all the worlds at least as plausible as the best $\neg\varphi$-worlds.

Example We provide one example of severe withdrawal in Figure 3.3. As an example of a severe withdrawal scenario we introduce the following story. Consider an agent and a dice. Someone throws the dice such that the agent cannot see the upper face. We have 6 possible worlds in our sphere system: `i` where 1 is the upper face, `ii` where 2 is the upper face and so on. Assume that the agent initially believes that the upper face is 3 while in reality (unknown to our agent) the upper face is 4. Thus she believes that the upper face is odd that is, she believes φ (the formula φ means "the number on the upper face is odd"). Besides, the agent considers that it is more likely that the upper face is 5 than 1, and that it is more likely that the upper face is 1 than 6 while she considers that it is equally likely that the upper face is 1 or 2, and that it is equally likely that it is 4 or 6. Then according to the agent: $3 < 5 < 1 \equiv 2 < 6 \equiv 4$ ($3 < 5$ is read as it is more likely that the upper face is 3 than 5 and $1 \equiv 2$ is read as it is equally likely that the upper face is 1 or 2). In other words, she considers `iii` more plausible than `v`, `v` more plausible than both `i` and `ii`, and finally `i` and `ii` more plausible than both `iv` and `vi`.

Now we consider the case where the agent receives a piece of information saying that the number on the upper face is even. The agent has to remove her belief that the number on the upper face is odd. So according to our definition for severe withdrawal, the agent will believe that the upper face is 1, 2, 3 or 5. Indeed, the agent considers `i`, `ii`, `iii` and `v` to be more plausible after the severe withdrawal. Then according to the agent: $1 \equiv 2 \equiv 3 \equiv 5 < 6 \equiv 4$.

In Figure 3.3, the numbers represent the spheres of the new Grove system after the revision. Thus all regions labelled with 1 form the first sphere of the new Grove system, the regions labelled with 2 form the second sphere and so on. Finally, the regions labelled with ω contain the states that are outside the union of all the spheres of the Grove system that is, the impossible states.

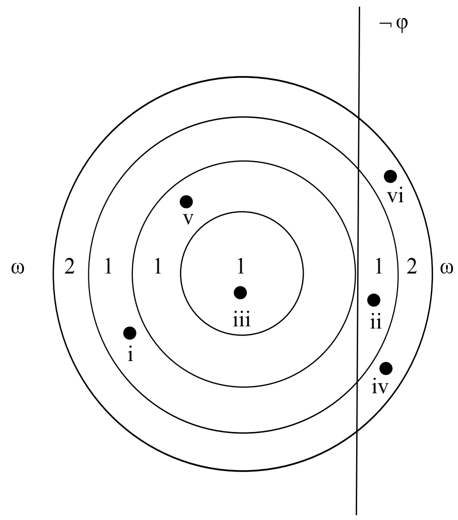

Figure 3.3: Example of severe withdrawal $-\varphi$

3.1.2 Other AGM-type contractions

Severe withdrawal is not the only AGM-friendly semantic contraction operation in the literature. Other options include conservative contraction $-_c P$ and moderate contraction $-_m P$ [72].

3.1.2. Definition. *Conservative contraction* is a belief change operation that removes a belief from the belief set of an agent such that, after contracting this belief set with φ, the most plausible worlds are the best $\neg\varphi$-worlds plus the initial best φ worlds.

3.1.3. DEFINITION. *Moderate contraction* is a belief change operation that removes a belief from the belief set of an agent such that, after contracting this belief set with φ, the most plausible worlds are the best $\neg\varphi$-worlds plus the initial best φ worlds and all the $\neg\varphi$-worlds are promoted. They become better than the rest of the φ-worlds.

Examples We provide two examples in Figures 3.4, 3.5 respectively for conservative contraction and moderate contraction. As an example of a conservative contraction and moderate contraction scenario, we come back to our story with an agent and a dice that we developed above.

After the agent receives the information that the number on the upper face is even, she has to remove her belief that the number on the upper face is odd. According to the definition for conservative contraction, the agent will believe that the upper face is 2 or 3. Indeed, the agent considers `ii` and `iii` to be more plausible after the conservative contraction. The agent still considers that it is more likely that the upper face is 5 rather than 1, and more likely that the upper face is 1 rather than 6 while she still considers that it is equally likely that the upper face is 4 or 6. Then according to the agent: $2 \equiv 3 < 5 < 1 < 6 \equiv 4$.

According to the definition for moderate contraction, the agent will believe that the upper face is 2 or 3. Indeed, the agent considers `ii` and `iii` to be more plausible after the conservative contraction. The agent still considers that it is more likely that the upper face is 5 than 1; she now considers that both 4 and 6 are more likely than 5. Then according to the agent: $2 \equiv 3 < 6 \equiv 4 < 5 < 1$.

3.2 A new approach to severe withdrawal

Now we will model these notions of contraction on (finite) total plausibility models. The language for belief contraction will consist of a base language equipped with a dynamic contraction modality. There are now several options and depending both on the choice of the static base language and the type of contraction operator a different language can be provided.

3.2.1 Language for the logic of severe withdrawal

We define two languages for severe withdrawal. The first language $\mathcal{L}_{Sev1}$ will be defined on top of $\mathcal{L}_{CDL}$ equipped with dynamic modalities for contraction $[-\varphi]$. The second language $\mathcal{L}_{Sev2}$ will be defined on top of $\mathcal{L}_{KK_D}$ equipped with dynamic modalities for contraction $[-\varphi]$.

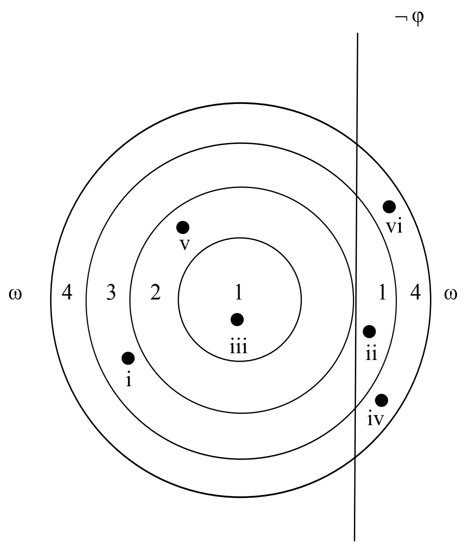

Figure 3.4: Example of a conservative contraction $-_c\varphi$

3.2.2 Semantics for the logic of severe withdrawal

In Dynamic Epistemic Logic, belief contraction is modelled as a model-changing operation. In particular it will take as input a given total plausibility model $\mathcal{M} = (S, \leq, V)$ and produces as output $\mathcal{M}^{-\varphi} = (S, \leq^{-\varphi}, V)$. Note that the current ordering relation $\leq$ of a given total plausibility model will be replaced by the following relation $\leq^{-\varphi}$ in the new model after the severe withdrawal with φ. The intended meaning of this is that all the worlds at least as plausible as the best $\neg\varphi$-worlds have become the best worlds, but apart from that, the old ordering remains the same.

3.2.1. Definition. The initial total plausibility model $\mathcal{M} = (S, \leq, V)$ after a severe withdrawal with φ is transformed into the following model $\mathcal{M}^{-\varphi} = (S, \leq^{-\varphi}, V)$ in which $t \leq^{-\varphi} s$ iff:

- $t \leq s$ or
- $t \leq w$ for some $w \in best \parallel \neg\varphi \parallel_{\mathcal{M}}$

The semantic clause for the dynamic operator $[-\varphi]$ is:

$$\mathcal{M}, s \vDash [-\varphi]\psi \text{ iff } \mathcal{M}^{-\varphi}, s \vDash \psi$$

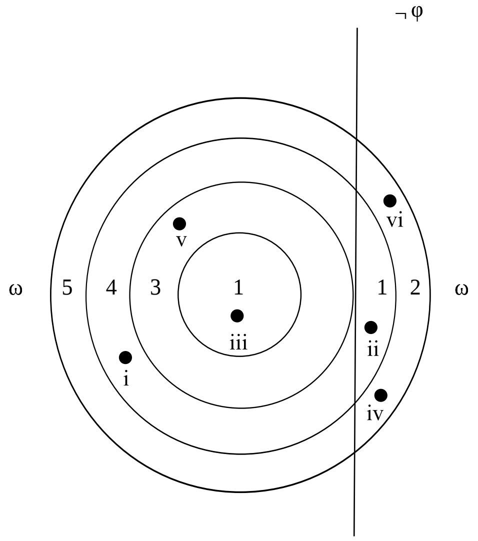

Figure 3.5: Example of a moderate contraction $-_m\varphi$

Example In Figure 3.6 we present a sphere model composed of 6 states s, t, u, v, w, x such that φ is true in s, t, v, x and $\neg\varphi$ is true in u, w.

From this sphere system we can generate the corresponding total plausibility model given in Figure 3.7.

Next, we present the total plausibility model resulting from the severe withdrawal with φ in Figure 3.8. In this model, the states s, t, u and v are equi-plausible, all of them being on top. Thus in the resulting model, the agent does not believe φ anymore and does not believe $\neg\varphi$ either.

3.2.3 Axiom system for the logic of severe withdrawal

3.2.2. Theorem. *A sound and complete proof system for the logic SEV with the language $\mathcal{L}_{Sev2}$ is given by the axioms and rules of $\mathcal{L}_{KK_D}$ plus the following reduction axioms:*

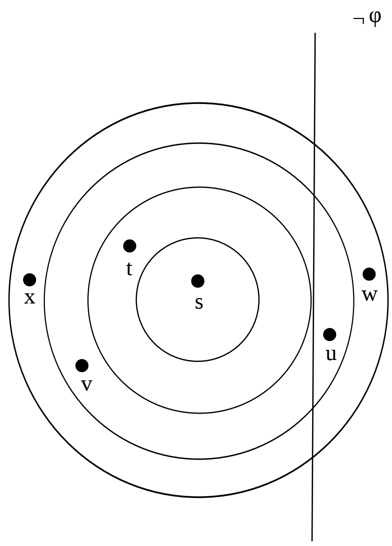

Figure 3.6: Grove system

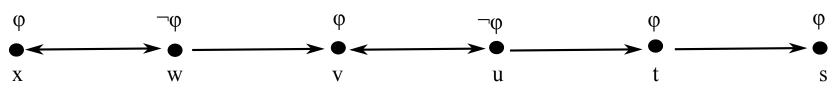

Figure 3.7: Initial total plausibility model

$$[-\varphi]p \iff p$$

$$[-\varphi]\neg\theta \iff \neg[-\varphi]\theta$$

$$[-\varphi](\theta \wedge \psi) \iff [-\varphi]\theta \wedge [-\varphi]\psi$$

$$[-\varphi]K\theta \iff K[-\varphi]\theta$$

$$[-\varphi]K_D\theta \iff (K_D[-\varphi]\theta \wedge (\neg K\varphi \to \neg K\neg(\neg\varphi \wedge K_D[-\varphi]\theta))$$

3.2.3. PROOF. The soundness of the axioms of $\mathcal{L}_{KK_D}$ is proved in [8]. All that remains is to show the soundness of the Reduction axioms. The proof of the first Reduction axioms is straightforward, we focus here only on the proof of the

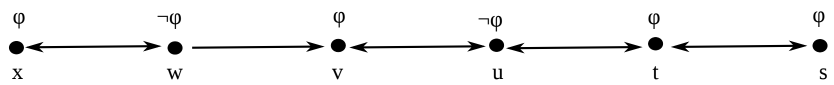

Figure 3.8: Severe withdrawal in the initial total plausibility model of Figure 3.7

reduction axiom for defeasible knowledge.

Suppose $s \vDash_{\mathcal{M}} [-\varphi]K_D\psi$. Then $s \vDash_{\mathcal{M}^{-\varphi}} K_D\psi$. So for all t such that $t \leq^{-\varphi} s$, $t \vDash_{\mathcal{M}^{-\varphi}} \psi$. This means that for all t such that $t \leq^{-\varphi} s$, $t \vDash_{\mathcal{M}} [-\varphi]\psi$.

We know that $t \leq^{-\varphi} s$ iff $t \leq w$ for some $w \in best \parallel \neg\varphi \parallel_{\mathcal{M}}$ or $t \leq s$. Then:

– for all t such that $t \leq w$ for some $w \in best \parallel \neg\varphi \parallel_{\mathcal{M}}$, $t \vDash_{\mathcal{M}} [-\varphi]\psi$ and,

– for all t such that $t \leq s$, $t \vDash_{\mathcal{M}} [-\varphi]\psi$.

This means that $s \vDash_{\mathcal{M}} \neg K\varphi \rightarrow \neg K\neg(\neg\varphi \wedge K_D[-\varphi]\psi)$ and $s \vDash_{\mathcal{M}} K_D[-\varphi]\psi$.

□

Sketch of the proof for completeness From the axiom system for the logic *SEV* with the language $\mathcal{L}_{Sev2}$, we note that what is the case after a severe withdrawal can be expressed by saying what is the case before the severe withdrawal. Using the Reduction Laws, the severe withdrawal operator can be step by step "pushed through" all other operators and at the end, completely eliminated using the Reduction Law for atomic formulas.

3.2.4. LEMMA. *Every formula of $\mathcal{L}_{Sev2}$ is provably equivalent in the above proof system with another formula in $\mathcal{L}_{KK_D}$.*

The completeness of the axioms of $\mathcal{L}_{Sev2}$ follows from the completeness of the axioms of $\mathcal{L}_{KK_D}$ that is proved in [8] and Lemma 3.2.4.

Let φ be a formula in $\mathcal{L}_{KK_D}$ such that it is satisfiable in a total plausibility model and let φ' be a formula in $\mathcal{L}_{Sev2}$ equivalent to φ. Then φ' is also satisfiable in the total plausibility model.

We cannot provide Reduction Laws in the language $\mathcal{L}_{Sev1}$, that is, if the static base language is the language of Conditional Doxastic Logic. Indeed we do not have a reduction axiom for conditional belief.

3.2.4 Objections against severe withdrawal

Many authors consider severe withdrawal to be a bad candidate for modelling contraction. In addition to not satisfying the Recovery principle[1], it does satisfy a highly implausible property, called Expulsiveness. For ontic facts p, q, we have that $\neg Bp \wedge \neg Bq$ implies $[-p]Bq \vee [-q]Bp$. This property does not allow unrelated beliefs to be undisturbed by each other's contraction.

Conservative contraction and moderate contraction are much better behaved than severe withdrawal. Indeed they satisfy the Recovery postulate.

3.3 A new approach to conservative contraction

3.3.1 Language for the logic of conservative contraction

We define two languages for conservative contraction. The first language $\mathcal{L}_{\mathit{Cons1}}$ will be defined on top of the static base language $\mathcal{L}_{CDL}$ equipped with dynamic modalities for contraction $[-_c\varphi]$. The second language $\mathcal{L}_{\mathit{Cons2}}$ will be defined on top of the static base language $\mathcal{L}_{KK_D}$ equipped with dynamic modalities for contraction $[-_c\varphi]$.

3.3.2 Semantics for the logic of conservative contraction

The operation of conservative contraction has as an effect that a given total plausibility model $\mathcal{M} = (S, \leq, V)$ will be transformed into a model $\mathcal{M}^{-_c\varphi} = (S, \leq^{-_c\varphi}, V)$. In this setting the current ordering relation $\leq$ of a given total plausibility model will be replaced by the following relation $\leq^{-_c\varphi}$ in the new model after the conservative contraction with φ. The intended meaning of this is that the best $\neg\varphi$-worlds have become equi-plausible with the best worlds initially on top, but apart from that, the old ordering remains the same.

3.3.1. Definition. The initial total plausibility model $\mathcal{M} = (S, \leq, V)$ after a conservative contraction with φ is transformed into the following model $\mathcal{M}^{-_c\varphi} = (S, \leq^{-_c\varphi}, V)$ in which $t \leq^{-_c\varphi} s$ iff:

- $t \in best \parallel \neg\varphi \parallel_{\mathcal{M}}$ or
- $t \leq s$.

3.3.2. Definition. We define $best^{-_c\varphi} P$:

- $best^{-_c\varphi} P := best\, P$ if $P \cap best \parallel \neg\varphi \parallel_{\mathcal{M}} = \varnothing$ or,

[1] We provide an explanation of the Recovery principle in Chapter 2 when we introduce the *AGM* postulates for belief contraction.

– $best^{-_c\varphi}\, P := (P \cap best \parallel \neg\varphi \parallel_{\mathcal{M}}) \cup (P \cap best\, S)$ if $P \cap best \parallel \neg\varphi \parallel_{\mathcal{M}} \neq \emptyset$.

The semantic clause for the dynamic operator $[-_c\varphi]$ is:

$$\mathcal{M}, s \models [-_c\varphi]\psi \text{ iff } \mathcal{M}^{-_c\varphi}, s \models \psi$$

Example In Figure 3.9 we present a sphere model composed of 6 states s, t, u, v, w, x such that φ is true in s, t, v, x and $\neg\varphi$ is true in u, w.

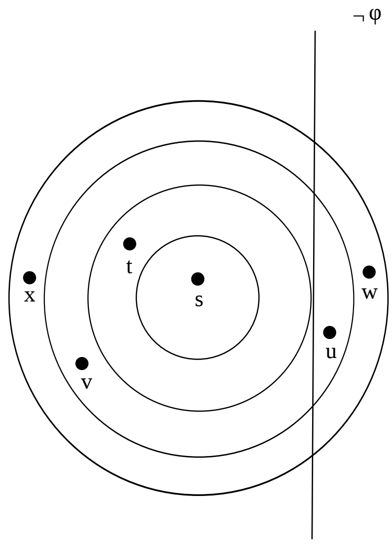

Figure 3.9: Grove system

From this sphere system we can generate the corresponding total plausibility model given in Figure 3.10.

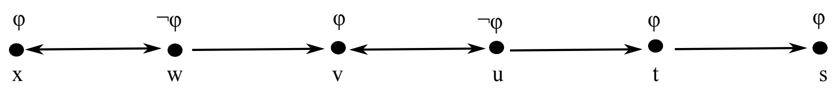

Figure 3.10: Initial total plausibility model

Next we present the total plausibility model resulting from the conservative contraction with φ in Figure 3.11. In this model, the state u is equi-plausible with the state s, both being on top. Thus the agent does not believe φ anymore and does not believe $\neg\varphi$ either.

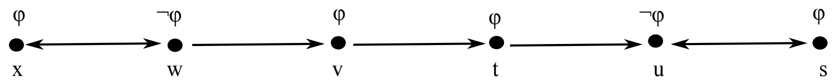

Figure 3.11: Conservative contraction in the initial total plausibility model of Figure 3.10

3.3.3 Axiom system for the logic of conservative contraction

3.3.3. THEOREM. *A sound and complete proof system for the logic CONS₁ with the language $\mathcal{L}_{Cons1}$ is given by the axioms and rules of $\mathcal{L}_{CDL}$ plus the following reduction axioms:*

$$
\begin{array}{lcl}
[-_c\varphi]p & \iff & p \\
[-_c\varphi]\neg\theta & \iff & \neg[-_c\varphi]\theta \\
[-_c\varphi](\theta\wedge\psi) & \iff & [-_c\varphi]\theta \wedge [-_c\varphi]\psi \\
[-_c\varphi]B^{\psi}\theta & \iff & B([-_c\varphi]\psi \to [-_c\varphi]\theta) \wedge B^{\neg\varphi}([-_c\varphi]\psi \to [-_c\varphi]\theta) \\
 & & \wedge \left(B^{\neg\varphi}[-_c\varphi]\neg\psi \to B^{[-_c\varphi]\psi}[-_c\varphi]\theta\right)
\end{array}
$$

3.3.4. PROOF. The soundness of the axioms of $\mathcal{L}_{CDL}$ is proved in [6]. All that remains is to show the soundness of the Reduction axioms. The proof of the first Reduction axioms is straightforward, we focus here only on the proof of the reduction axiom for conditional belief.

Suppose $s \vDash_{\mathcal{M}} [-_c\varphi]B^{\psi}\theta$. Then $s \vDash_{\mathcal{M}^{-_c\varphi}} B^{\psi}\theta$. So $best^{-_c\varphi} \parallel \psi \parallel_{\mathcal{M}^{-_c\varphi}} \subseteq \parallel \theta \parallel_{\mathcal{M}^{-_c\varphi}}$. This means that $best^{-_c\varphi} \parallel [-_c\varphi]\psi \parallel_{\mathcal{M}} \subseteq \parallel [-_c\varphi]\theta \parallel_{\mathcal{M}}$.

We know that:

– $best^{-_c\varphi} \parallel [-_c\varphi]\psi \parallel_{\mathcal{M}} := best \parallel [-_c\varphi]\psi \parallel_{\mathcal{M}}$ if $\parallel [-_c\varphi]\psi \parallel_{\mathcal{M}} \cap best \parallel \neg\varphi \parallel_{\mathcal{M}} = \varnothing$ or,

– $best^{-_c\varphi} \parallel [-_c\varphi]\psi \parallel_{\mathcal{M}} := (\parallel [-_c\varphi]\psi \parallel_{\mathcal{M}} \cap best \parallel \neg\varphi \parallel_{\mathcal{M}}) \cup (\parallel [-_c\varphi]\psi \parallel_{\mathcal{M}} \cap best\, S)$ if $\parallel [-_c\varphi]\psi \parallel_{\mathcal{M}} \cap best \parallel \neg\varphi \parallel_{\mathcal{M}} \neq \varnothing$.

Then:

– if $best \parallel \neg\varphi \parallel_{\mathcal{M}} \subseteq \parallel [-_c\varphi]\neg\psi \parallel_{\mathcal{M}}$ then $best \parallel [-_c\varphi]\psi \parallel_{\mathcal{M}} \subseteq \parallel [-_c\varphi]\theta \parallel_{\mathcal{M}}$ and,

– $best \parallel \neg\varphi \parallel_{\mathcal{M}} \cap \parallel [-_c\varphi]\psi \parallel_{\mathcal{M}} \subseteq \parallel [-_c\varphi]\theta \parallel_{\mathcal{M}}$ and

– $best\, S \cap \parallel [-_c\varphi]\psi \parallel_{\mathcal{M}} \subseteq \parallel [-_c\varphi]\theta \parallel_{\mathcal{M}}$ otherwise.

This means that:

– $s \vDash_{\mathcal{M}} (B^{\neg\varphi}[-_c\varphi]\neg\psi \to B^{[-_c\varphi]\psi}[-_c\varphi]\theta)$ and

– $s \vDash_{\mathcal{M}} B^{\neg\varphi}([-_c\varphi]\psi \to [-_c\varphi]\theta)$ and

– $s \vDash_{\mathcal{M}} B([-_c\varphi]\psi \to [-_c\varphi]\theta)$.

□

Sketch of the proof for completeness From the axiom system for the logic *CONS₁* with the language $\mathcal{L}_{Cons1}$, we note that what is the case after a conservative contraction can be expressed by saying what is the case before the conservative contraction. Using the Reduction Laws, the conservative contraction operator can be step by step "pushed through" all other operators and at the end, completely eliminated using the Reduction Law for atomic formulas.

3.3.5. LEMMA. *Every formula of $\mathcal{L}_{Cons1}$ is provably equivalent in the above axiom system to another formula in $\mathcal{L}_{CDL}$.*

The completeness of the axioms of $\mathcal{L}_{Cons1}$ follows from the completeness of the axioms of $\mathcal{L}_{CDL}$ that is proved in [6] and Lemma 3.3.5.

Let φ be a formula in $\mathcal{L}_{CDL}$ such that it is satisfiable in a total plausibility model and let φ' be a formula in $\mathcal{L}_{Cons1}$ equivalent to φ. Then φ' is also satisfiable in the total plausibility model.

3.3.6. THEOREM. *A sound and complete proof system for the logic $CONS_2$ with the language $\mathcal{L}_{Cons2}$ is given by the axioms and rules of $\mathcal{L}_{KK_D}$ plus the following reduction axioms:*

$$
\begin{array}{lcl}
[-_c\varphi]p & \Longleftrightarrow & p \\
[-_c\varphi]\neg\theta & \Longleftrightarrow & \neg[-_c\varphi]\theta \\
[-_c\varphi](\theta \wedge \psi) & \Longleftrightarrow & [-_c\varphi]\theta \wedge [-_c\varphi]\psi \\
[-_c\varphi]K\psi & \Longleftrightarrow & K[-_c\varphi]\psi \\
[-_c\varphi]K_D\psi & \Longleftrightarrow & K_D[-_c\varphi]\psi \wedge B^{\neg\varphi}[-_c\varphi]\psi
\end{array}
$$

Note that in the last reduction axiom, we use the following conditional operator $B^{\neg\varphi}[-_c\varphi]\psi$ as an abbreviation as defined in Proposition 2.5.12.

3.3.7. PROOF. The soundness of the axioms of $\mathcal{L}_{KK_D}$ is proved in [8]. This leaves us to focus on the Reduction axioms. The proof of the first Reduction axioms is straightforward, we focus here only on the proof of the reduction axiom for defeasible knowledge.

Suppose $s \vDash_{\mathcal{M}} [-_c\varphi]K_D\psi$. Then $s \vDash_{\mathcal{M}^{-c\varphi}} K_D\psi$. So for all t such that $t \leq^{-c\varphi} s$, $t \vDash_{\mathcal{M}^{-c\varphi}} \psi$. This means that for all t such that $t \leq^{-c\varphi} s$, $t \vDash_{\mathcal{M}} [-_c\varphi]\psi$.

We know that $t \leq^{-c\varphi} s$ iff $t \in best \parallel \neg\varphi \parallel_{\mathcal{M}}$ or $t \leq s$. Then:

- for all t such that $t \in best \parallel \neg\varphi \parallel_{\mathcal{M}}$, $t \vDash_{\mathcal{M}} [-_c\varphi]\psi$ and,
- for all t such that $t \leq s$, $t \vDash_{\mathcal{M}} [-_c\varphi]\psi$.

This means that $s \vDash_{\mathcal{M}} B^{\neg\varphi}[-_c\varphi]\psi$ and $s \vDash_{\mathcal{M}} K_D[-_c\varphi]\psi$.

□

Sketch of the proof for completeness From the axiom system for the logic *CONS2* with the language $\mathcal{L}_{Cons2}$, we note that what is the case after a conservative contraction can be expressed by saying what is the case before the conservative contraction. Using the Reduction Laws, the conservative contraction operator can be step by step "pushed through" all other operators and at the end, completely eliminated using the Reduction Law for atomic formulas.

3.3.8. LEMMA. *Every formula of $\mathcal{L}_{Cons2}$ is provably equivalent in the above proof system to another formula in $\mathcal{L}_{KK_D}$.*

The completeness of the axioms of $\mathcal{L}_{Cons2}$ follows from the completeness of the axioms of $\mathcal{L}_{KK_D}$ that is proved in [8] and Lemma 3.3.8.

Let φ be a formula in $\mathcal{L}_{KK_D}$ such that it is satisfiable in a total plausibility model and let φ' be a formula in $\mathcal{L}_{Cons2}$ equivalent to φ. Then φ' is also satisfiable in the total plausibility model.

3.4 A new approach to moderate contraction

3.4.1 Language for the logic of moderate contraction

We define two languages for moderate contraction. The first language $\mathcal{L}_{Mod1}$ will be defined on top of the static base language $\mathcal{L}_{CDL}$ equipped with dynamic modalities for contraction $[-_m\varphi]$. The second language $\mathcal{L}_{Mod2}$ will be defined on top of the static base language $\mathcal{L}_{KK_D}$ equipped with dynamic modalities for contraction $[-_m\varphi]$.

3.4.2 Semantics for the logic of moderate contraction

The operation of moderate contraction has as an effect that a given total plausibility model $\mathcal{M} = (S, \leq, V)$ will be transformed into a model $\mathcal{M}^{-_m\varphi} = (S, \leq^{-_m\varphi}, V)$. In this setting the current ordering relation $\leq$ of a given total plausibility model will be replaced by the following relation $\leq^{-_m\varphi}$ in the new model after the moderate contraction with φ. The intended meaning of this is that the best $\neg\varphi$-worlds become equi-plausible with the best worlds initially on top, then the rest of the $\neg\varphi$-worlds become better than the rest of the φ-worlds, and within these two zones the old ordering remains.

3.4.1. Definition. The initial total plausibility model $\mathcal{M} = (S, \leq, V)$ after a moderate contraction with φ is transformed into the following model $\mathcal{M}^{-_m\varphi} = (S, \leq^{-_m\varphi}, V)$ in which $t \leq^{-_m\varphi} s$ iff:

- $t \in best \parallel \neg\varphi \parallel_{\mathcal{M}}$ or,
- $s \in \parallel \varphi \parallel_{\mathcal{M}}$, $s \notin best \parallel \varphi \parallel_{\mathcal{M}}$ and $t \in \parallel \neg\varphi \parallel_{\mathcal{M}}$ or,
- $s \in \parallel \varphi \parallel_{\mathcal{M}}$, $t \in \parallel \varphi \parallel_{\mathcal{M}}$ and $t \leq s$ or,
- $s \in \parallel \neg\varphi \parallel_{\mathcal{M}}$, $t \in \parallel \neg\varphi \parallel_{\mathcal{M}}$ and $t \leq s$.

3.4.2. Definition. We can define $best^{-_m\varphi}P$:

- $best^{-_m\varphi}\, P := best\, P$ if $P \cap \parallel \neg\varphi \parallel_{\mathcal{M}} = \varnothing$ or,
- $best^{-_m\varphi}\, P := best(\neg\varphi \cap P) \cup (P \cap best\, S)$ if $P \cap \parallel \neg\varphi \parallel_{\mathcal{M}} \neq \varnothing$.

The semantic clause for the dynamic operator $[-_m\varphi]$ is:

$$\mathcal{M}, s \vDash [-_m\varphi]\psi \text{ iff } \mathcal{M}^{-_m\varphi}, s \vDash \psi$$

Example In Figure 3.12 we present a sphere model composed of 6 states s, t, u, v, w, x such that φ is true in s, t, v, x and $\neg\varphi$ is true in u, w.

From this sphere system we can generate the corresponding total plausibility model given in Figure 3.13.

Next we present the total plausibility model resulting from the moderate contraction with φ in Figure 3.14. In this model, the state u is equi-plausible with the state s, both being on top while w is more plausible than t, v, x. Thus the agent does not believe φ anymore and does not believe $\neg\varphi$ either. However she has a propensity to consider $\neg\varphi$ more plausible than φ. If she further revises/contracts her beliefs, she will finally come to believe $\neg\varphi$ more easily than φ.

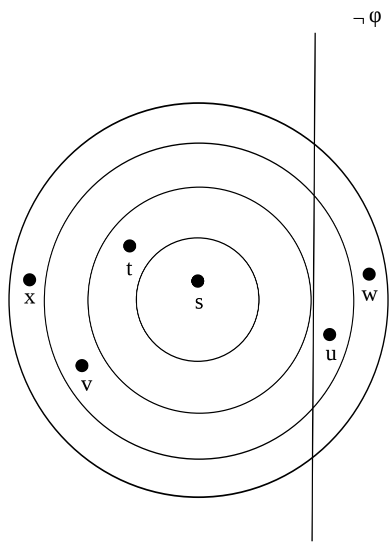

Figure 3.12: Grove system

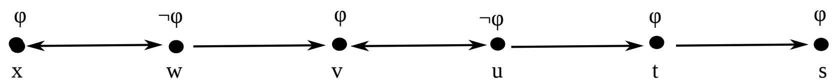

Figure 3.13: Initial total plausibility model

3.4.3 Axiom system for the logic of moderate contraction

3.4.3. THEOREM. *A sound and complete proof system for the logic MOD_1 with the language $\mathcal{L}_{Mod1}$ is given by the axioms and rules of $\mathcal{L}_{CDL}$ plus the following reduction axioms:*

$$
\begin{array}{lcl}
[-_m\varphi]p & \Longleftrightarrow & p \\
[-_m\varphi]\neg\theta & \Longleftrightarrow & \neg[-_m\varphi]\theta \\
[-_m\varphi](\theta \wedge \psi) & \Longleftrightarrow & [-_m\varphi]\theta \wedge [-_m\varphi]\psi \\
[-_m\varphi]B^{\psi}\theta & \Longleftrightarrow & B([-_m\varphi]\psi \to [-_m\varphi]\theta) \wedge B^{\neg\varphi\wedge[-_m\varphi]\psi}[-_m\varphi]\theta \\
 & & \wedge \big(K^{\neg\varphi}[-_m\varphi]\neg\psi \to B^{[-_m\varphi]\psi}[-_m\varphi]\theta\big)
\end{array}
$$

3.4.4. PROOF. The soundness of the axioms of $\mathcal{L}_{CDL}$ is proved in [6]. All that remains is to show the soundness of the Reduction axioms. The proof of the

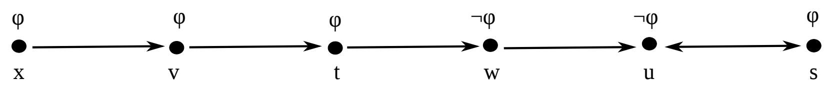

Figure 3.14: Moderate contraction in the initial total plausibility model of Figure 3.13

first Reduction axioms is straightforward, we focus here only on the proof of the reduction axiom for conditional belief.

Suppose $s \vDash_{\mathcal{M}} [-_m\varphi]B^{\psi}\theta$. Then $s \vDash_{\mathcal{M}^{-m\varphi}} B^{\psi}\theta$. So $best^{-m\varphi} \parallel \psi \parallel_{\mathcal{M}^{-m\varphi}} \subseteq \parallel \theta \parallel_{\mathcal{M}^{-m\varphi}}$. This means that $best^{-m\varphi} \parallel [-_m\varphi]\psi \parallel_{\mathcal{M}} \subseteq \parallel [-_m\varphi]\theta \parallel_{\mathcal{M}}$.
We know that:

- $best^{-m\varphi} \parallel [-_m\varphi]\psi \parallel_{\mathcal{M}} := best \parallel [-_m\varphi]\psi \parallel_{\mathcal{M}}$ if $\parallel [-_m\varphi]\psi \parallel_{\mathcal{M}} \cap \parallel \neg\varphi \parallel_{\mathcal{M}} = \varnothing$ or,
- $best^{-m\varphi} \parallel [-_m\varphi]\psi \parallel_{\mathcal{M}} := best(\neg\varphi \cap \parallel [-_m\varphi]\psi \parallel_{\mathcal{M}}) \cup (\parallel [-_m\varphi]\psi \parallel_{\mathcal{M}} \cap best\, S)$ if $\parallel [-_m\varphi]\psi \parallel_{\mathcal{M}} \cap \parallel \neg\varphi \parallel_{\mathcal{M}} \neq \varnothing$.

Then:

- if $\parallel \neg\varphi \parallel_{\mathcal{M}} \subseteq \parallel [-_m\varphi]\neg\psi \parallel_{\mathcal{M}}$ then $best \parallel [-_m\varphi]\psi \parallel_{\mathcal{M}} \subseteq \parallel [-_m\varphi]\theta \parallel_{\mathcal{M}}$ and,
- $best(\neg\varphi \cap \parallel [-_m\varphi]\psi \parallel_{\mathcal{M}}) \subseteq \parallel [-_m\varphi]\theta \parallel_{\mathcal{M}}$ and
- $best\, S \cap \parallel [-_m\varphi]\psi \parallel_{\mathcal{M}} \subseteq \parallel [-_m\varphi]\theta \parallel_{\mathcal{M}}$ otherwise.

This means that:

- $s \vDash_{\mathcal{M}} K^{\neg\varphi}[-_m\varphi]\neg\psi \to B^{[-_m\varphi]\psi}[-_m\varphi]\theta$ and
- $s \vDash_{\mathcal{M}} B^{\neg\varphi \wedge [-_m\varphi]\psi}[-_m\varphi]\theta$ and
- $s \vDash_{\mathcal{M}} B([-_m\varphi]\psi \to [-_m\varphi]\theta)$.

□

Sketch of the proof for completeness From the axiom system for the logic *MOD1* with the language $\mathcal{L}_{Mod1}$, we note that what is the case after a conservative contraction can be expressed by saying what is the case before the conservative contraction. Using the Reduction Laws, the moderate contraction operator can be step by step "pushed through" all other operators and at the end, completely eliminated using the Reduction Law for atomic formulas.

3.4.5. Lemma. *Every formula of $\mathcal{L}_{Mod1}$ is provably equivalent in the above proof system to another formula in $\mathcal{L}_{CDL}$.*

The completeness of the axioms of $\mathcal{L}_{Mod1}$ follows from the completeness of the axioms of $\mathcal{L}_{CDL}$ that is proved in [6] and Lemma 3.4.5.

Let φ be a formula in $\mathcal{L}_{CDL}$ such that it is satisfiable in a total plausibility model and let φ' be a formula in $\mathcal{L}_{Mod1}$ equivalent to φ. Then φ' is also satisfiable in the total plausibility model.

We cannot provide an axiom system for the logic MOD_2 with the language $\mathcal{L}_{Mod2}$. Indeed we do not have a reduction axiom for defeasible knowledge. This reduction axiom would require a new operator or equivalently a new language. Let us consider what would be such an axiom:

$$[-_m\varphi]K_D\psi \iff B^{\neg\varphi}[-_m\varphi]\psi \wedge (\varphi \rightarrow K_D(\varphi \rightarrow [-_m\varphi]\psi))$$

$$\wedge(\neg\varphi \rightarrow K_D(\neg\varphi \rightarrow [-_m\varphi]\psi)) \wedge (\varphi \wedge \neg best\varphi \rightarrow K^{\neg\varphi}[-_m\varphi]\psi))$$

where the semantics of $best\varphi$ is given by $s \vDash best\varphi$ iff $s \in best \parallel \varphi \parallel_{\mathcal{M}}$.

The problem here is the operator "best" which cannot be expressed in our language.

However, we can provide the following reduction axioms:

$$\begin{aligned}
[-_m\varphi]p &\iff p \\
[-_m\varphi]\neg\theta &\iff \neg[-_m\varphi]\theta \\
[-_m\varphi](\theta \wedge \psi) &\iff [-_m\varphi]\theta \wedge [-_m\varphi]\psi \\
[-_m\varphi]K\psi &\iff K[-_m\varphi]\psi
\end{aligned}$$

Conclusion

Belief revision has been widely explored in *DEL* contrary to belief contraction. However, belief contraction is also a very interesting notion and is worth being studied in the setting of *DEL*. A belief contraction operation really comes with its own reduction axioms. We explored in this chapter three different notions of contraction: severe withdrawal, conservative contraction and moderate contraction. We clarified the mechanism of each of these operations. We also explained the limits of the mechanism of severe withdrawal while stressing the advantages of conservative and moderate contraction.

In the next chapter, we continue to develop Soft *DEL* by designing a formal setting allowing to make explicit the connections between plausibility models, evidence models and uniting some existing different settings.

Chapter 4

Justification models

Aim In this chapter we introduce justification models as a generalization of some of the models we have seen before. So we are after a general setting that can encompass both total and partial plausibility models, sphere models as well as the evidence models of van Benthem and Pacuit.

Summary The main points are:

– we introduce justification models defining the notions of evidence and justification. We provide an example to illustrate this new type of model.

– we study some special classes of justification models. We start with plausibility models and show how they are related to justification models. We show how a justification model can be mapped into a plausibility model which allows us to define epistemic and doxastic attitudes.

– we then introduce counting models and weighting models, proving that they can be considered as a special kind of justification models and we show that (introspective) evidence models[1] are exactly a special kind of justification models.

– then we define the notion of update on justification models. We study this update operation in each class of justification models.

– we provide a language and an axiomatization for the logic of justification.

– finally, we focus on total justification models providing some important properties about defeasible knowledge.

[1]We define an introspective evidence model in Definition 4.2.9.

Background We will need to use the evidence models as well as the (partial and total) plausibility models we introduced in Chapter 2.

Van Benthem and Pacuit provide a semantic approach to evidence in [16]. They deal with possibly false and possibly mutually inconsistent evidence. The belief revision in evidence models does not satisfy the *AGM* postulates since the pre-order that can be induced on states in these models is not total. That means that non every two states are comparable. Van Benthem and Pacuit show in [16] how (uniform) evidence models can be turned into partial plausibility models and how a partial plausibility model can be extended to an evidence model.

4.1 Justification models as a new formalization of belief, evidence and justification

In this section we introduce a new type of model that we call justification model. Justification models will be used to capture the agents' evidence or justification. We claim that they provide a good formalization of belief, evidence and justification. Below we first introduce the formal details of justification models.

4.1.1 General presentation of justification models

We introduce a (finite) pointed *justification model* via the following definition.

4.1.1. Definition. As for our formalisation, we introduce a pointed justification model $\mathcal{M}$ to be a tuple $(S, E, \leq, \|\cdot\|, s_0)$ consisting of:

- a finite set S of possible worlds,
- a family $E \subseteq \mathcal{P}(S)$ of non-empty subsets $e \subseteq S$ ($\varnothing \notin E$), called evidence (sets) such that S is itself an evidence set ($S \in E$). A *body of evidence* (or an *argument*) is any consistent family of evidence sets, i.e. any $F \subseteq E$ such that $\bigcap F \neq \varnothing$. We denote by $\mathcal{E} \subseteq \mathcal{P}(E)$ the family of all bodies of evidence.
- a partial preorder $\leq$ on $\mathcal{E}$, satisfying the following constraints:

$$F \subseteq F' \Rightarrow F \leq F'$$

$$F \leq F', G \leq G' \text{ and } F' \cap G' = \varnothing \Rightarrow F \cup G \leq F' \cup G'$$

$$F < F', G \leq G' \text{ and } F' \cap G' = \varnothing \Rightarrow F \cup G < F' \cup G'$$

  whenever $F, F', G, G', F \cup G, F' \cup G'$ are bodies of evidence.
- a standard $\|\cdot\|$ valuation map,
- an actual state of affairs s_0.

Note that the empty family of evidence sets $\varnothing$ is a body of evidence. Since formally speaking $\bigcap \varnothing = \{s \in S \mid \forall e \in E(e \in \varnothing \Rightarrow s \in e)\}$, $\varnothing$ is a consistent family of evidence sets.

Note also that $\leq$ is a partial preorder, connecting only the consistent families of evidence sets and we read $F \leq G$ as the body of evidence G is (considered as) at least as convincing or easier to accept (by some implicit agent) as the body of evidence F. Similarly the strict version $F < G$ denotes that the body of evidence G is (considered as) more convincing, easier to accept (by some implicit agent) than the body of evidence F.

We can impose further conditions on $\leq$ to obtain a total pre-order. Indeed we can require that either $F \leq F'$ or $F' \leq F$. We call a justification model with a total pre-order a *total justification model*. In total justification models, all evidence sets are comparable.

Assumption We assume here that the agent is introspective regarding to evidence. Informally it means that we assume that the agent knows what evidence she has[2].

Explanation of the conditions in Definition 4.1.1 The first condition expresses that if an argument F entails argument F' then F' is at least as convincing as F. Note that this implies that $\varnothing$ is the least convincing argument, that is, $\varnothing \leq F$ for all $F \in \mathcal{E}$. The second condition states that if F' is at least as convincing as F and G' is at least as convincing as G such that the argument F' and G' are not consistent, then the union of F' and G' is at least as convincing as the union of F and G. The last condition says that if F' is more convincing than F and G' is at least as convincing as G such that the argument F' and G' are not consistent, then the union of F' and G' is more convincing than the union of F and G.

Example We illustrate a justification model in Figure 4.1.

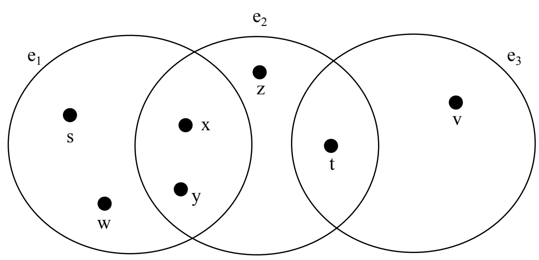

Figure 4.1: Justification model

In this figure, s, t, v, w, x, y, z are possible worlds. There are three evidence sets: $e_1 = \{s, x, w, y\}$, $e_2 = \{x, y, z, t\}$ and $e_3 = \{t, v\}$. Thus $E = \{e_1, e_2, e_3\}$. We can then define five bodies of evidence in addition to the empty family of evidence sets $\varnothing$: $D = \{e_1\}$, $F = \{e_2\}$, $G = \{e_3\}$, $H = \{e_1, e_2\}$ and $I = \{e_2, e_3\}$[3]. Thus $\mathcal{E} = \{\varnothing, \{e_1\}, \{e_2\}, \{e_3\}, \{e_1, e_2\}, \{e_2, e_3\}\}$.

Now we can state the following relations between consistent families of evidence sets:

[2] If we drop this assumption, then E is no longer a family of evidence sets but a relation $E \subseteq S \times \wp(S)$. In that case it coincides with the definition of the evidence relation of van Benthem and Pacuit [16].

[3] Note that $\{e_1, e_3\}$ is not a body of evidence since the evidence sets $\{e_1\}$ and $\{e_3\}$ are not consistent sets.

- $\{e_1\} \subseteq \{e_1, e_2\}$, $D \preceq H$,
- $\{e_2\} \subseteq \{e_1, e_2\}$, $F \preceq H$,
- $\{e_2\} \subseteq \{e_2, e_3\}$, $F \preceq I$,
- $\{e_3\} \subseteq \{e_2, e_3\}$, $G \preceq I$.

4.1.2 Plausibility in justification models

On every justification model, we can define a plausibility order on states in a canonical way.

4.1.2. DEFINITION. We define the notion of *largest body of evidence consistent with a given state* $s \in S$ and write it as $E_s := \{e \in E \mid s \in e\}$.

We can induce a plausibility relation on states directly from the partial pre-order on $\mathcal{E}$: for two states $s, t \in S$, we put

$$s \leq_E t \text{ iff } E_t \preceq E_s$$

Example Let us go back to Figure 4.1 above. The largest body of evidence consistent with x is $E_x := \{e_1, e_2\}$. The largest body of evidence consistent with s is $E_s := \{e_1\}$. Since $E_s \preceq E_x$, then $x \leq_E s$.

Epistemic and doxastic notions Given a justification model, we can define all epistemic and doxastic notions usually defined on plausibility models as irrevocable knowledge (K), belief (B), conditional belief (B^-), strong belief (Sb) and defeasible knowledge (K_D) using this plausibility order $\leq_E$.

$$KP := \{s \in S : P = S\}$$

$$BP := \{s \in S : best_E \subseteq P\}$$

$$B^Q P := \{s \in S : best_E Q \subseteq P\}$$

$$SbP := \{s \in S : P \neq \varnothing \text{ and } w <_E t \text{ for all } t \in P \text{ and all } w \notin P\}$$

$$K_D P = \{s \in S : t \not>_E s \text{ implies } t \in P\}$$

Note that in case the plausibility order $\leq_E$ is total, Proposition 2.5.6 holds, this is:

$$s \vDash K_D P \text{ iff } s \vDash B^Q P \text{ for all } Q \text{ such that } s \vDash Q$$

4.2 Special classes of justification models

Justification models are a very general framework, subsuming a lot of different existing settings. We will show that partial and total plausibility models and evidence models are a special class of justification models and we will introduce counting models and weighting models as special classes of justification models. The relations between justification models, counting models, weighting models, plausibility models and evidence models is given by the following Figure 4.2.

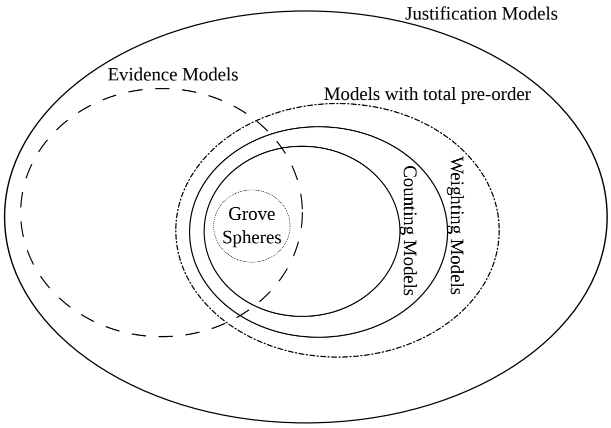

Figure 4.2: Relations between settings

4.2.1 Plausibility models

How can we interpret a plausibility model as a justification model? As we will see there are at least two ways to interpret a plausibility model $(S, \leq, \|\cdot\|, s_0)$ as a justification model $(S, E, \leq, \|\cdot\|, s_0)$. Plausibility models are a special kind of justification models in which:

- S is the set of possible worlds,
- the set of evidence sets $E = \{\downarrow w : w \in S\}$ where $\downarrow w = \{s \in S : s \leq w\}$

Now we can take either of the following two possible options:

1. $F \leq_1 F'$ iff $F \subseteq F'$ or
2. $F \leq_2 F'$ iff $| F | \leq | F' |$.[4]

In the second case, plausibility models are a special case of counting models which we introduce in the next section. In the first case, plausibility models $(S, \leq, \|\cdot\|, s_0)$ are a special kind of justification models $(S, E, \leq_1, \|\cdot\|, s_0)$ in which:

[4]Note that while the inclusion order is not necessarily total, the cardinality order is total.

1. the pre-order on bodies of evidence is given by inclusion ($\preceq_1 = \subseteq$) and
2. the evidence sets are nested that is, $\forall e, e' \in E$ either $e \subseteq e'$ or $e' \subseteq e$ (it means that the pre-order is a total pre-order).
3. a body of evidence F corresponds to any family of spheres of a plausibility model[5],
4. $\mathcal{E}$ corresponds to all the families of spheres.

4.2.1. Definition. Let us define two *plausibility maps* $Just_1$ and $Just_2$, mapping plausibility models to justification models:

- $(S, \leq, \|.\|, s_0) \overset{\text{Just}_1}{\longmapsto} (S, E, \preceq_1, \|.\|, s_0)$
- $(S, \leq, \|.\|, s_0) \overset{\text{Just}_2}{\longmapsto} (S, E, \preceq_2, \|.\|, s_0)$.

The plausibility map $Just_1$ corresponds to the case where the pre-order on bodies of evidence is given by inclusion while the plausibility map $Just_2$ corresponds to the case where the pre-order on bodies of evidence is given by cardinality. We will call the justification models that can be obtained in one of these two ways (by applying $Just_1$ or $Just_2$ to a plausibility model), *sphere-based justification models.*

Mapping justification models to plausibility models Any justification model $(S, E, \preceq, \|.\|, s_0)$ can be mapped into a type of plausibility model $(S, \leq, \|.\|, s_0)$.

4.2.2. Definition. We define the *plausibility map Plau*, mapping justification models to plausibility models: $(S, E, \preceq, \|.\|, s_0) \overset{\text{Plau}}{\longmapsto} (S, \leq, \|.\|, s_0)$.

A justification model with a partial pre-order gives a partial plausibility model while a justification model with a total pre-order gives a total plausibility model.

4.2.3. Proposition. *Total justification models induce total plausibility models:*

$$\forall F, F' (F \preceq F' \vee F' \preceq F) \iff \forall s, s' (s \leq_E s' \vee s' \leq_E s)$$

Note that the map $Plau$ mapping justification models to plausibility models is not injective. So, two different justification models can give the same plausibility model.

If we interpret a plausibility model $\mathcal{M}$ as a justification model $\mathcal{M}'$ and then apply the map $Plau$, we obtain the initial plausibility model $\mathcal{M}$. The converse

[5] We saw in Chapter 3 that plausibility models and Grove models are equivalent. We use this equivalence and discuss the relation between justification models and plausibility models in terms of spheres.

is false since if we apply the map *Plau* on a justification model $\mathcal{M}'$ to obtain a plausibility model $\mathcal{M}$ and then interpret this plausibility model $\mathcal{M}$ as a justification model, we do not obtain the initial justification model $\mathcal{M}'$. Thus, we have both:

– $Plau(Just_1(\mathcal{M})) = \mathcal{M}$ for any plausibility model $\mathcal{M}$ and $Just_1(Plau(\mathcal{M}')) \neq \mathcal{M}'$ for any justification model $\mathcal{M}'$,

– $Plau(Just_2(\mathcal{M})) = \mathcal{M}$ for any plausibility model $\mathcal{M}$ and $Just_2(Plau(\mathcal{M}')) \neq \mathcal{M}'$ for any justification model $\mathcal{M}'$.

4.2.2 Counting models

4.2.4. Definition. A *counting model* is a justification model $(S, E, \preceq, \|\cdot\|, s_0)$ in which the pre-order is given by the cardinality order, i.e. $F \preceq F'$ iff $| F | \leq | F' |$.[6]

Thus a body of evidence G is more convincing than a body of evidence F iff the number of evidence sets $e \in G$ is bigger than the number of evidence sets $e \in F$: $F \preceq G$ iff $| F | \leq | G |$. The intuition is that the more evidence the agent has, the stronger it is.

We represent an example of counting model in Figure 4.3. In this example, there are 4 states namely s, t, v and w. There are 5 evidence sets e_1, e_2, e_3, e_4 and e_5. Since the order on evidence is induced from cardinality, we have $F \preceq G$ iff $| F | \leq | G |$. Thus, $| \{e_1\} | = 1$, $| \{e_3, e_4\} | = 2$, $| \{e_2, e_4\} | = 2$ and $| \{e_2, e_3, e_5\} | = 3$. Then, $\{e_2, e_4\} \prec \{e_2, e_3, e_5\}$, $\{e_3, e_4\} \preceq \{e_2, e_4\}$, $\{e_1\} \prec \{e_3, e_4\}$.

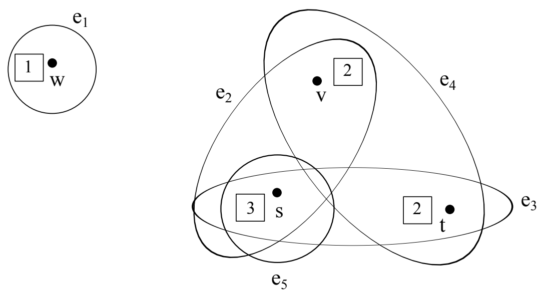

Figure 4.3: Counting model

4.2.3 Weighting models

First we define weighting models.

[6]Note that cardinality generates a total pre-order.

4.2.5. DEFINITION. A pointed *weighting model* is a structure $(S, E, f, \|\cdot\|, s_0)$ where $f : E \to \mathbb{N}$.

Weighting models can be considered as a special kind of justification models. Let $(S, E, f, \|\cdot\|, s_0)$ be a weighting model. The function f can be extended to $\mathcal{E}$ such that $f(E) = \sum_{e \in E} f(e)$ and $E \preceq_f E'$ iff $f(E) \leq f(E')$. Any weighting model endowed with $\preceq_f$ is a justification model.

We represent an example of a weighting model in Figure 4.4. In this example, there are 4 states namely s, t, v and w. There are 5 evidence sets e_1, e_2, e_3, e_4 and e_5. There is a function $f : E \to \mathbb{N}$ defined such that $f(e_1) = 1, f(e_2) = 2, f(e_3) = 1, f(e_4) = 3$ and $f(e_5) = 3$. Thus, $f(e_1) = 1$, $f(\{e_3, e_4\}) = 4$, $f(\{e_2, e_4\}) = 5$ and $f(\{e_2, e_3, e_5\}) = 6$. Then $\{e_2, e_4\} \prec_f \{e_2, e_3, e_5\}$, $\{e_3, e_4\} \prec_f \{e_2, e_4\}$, $\{e_1\} \prec_f \{e_3, e_4\}$.

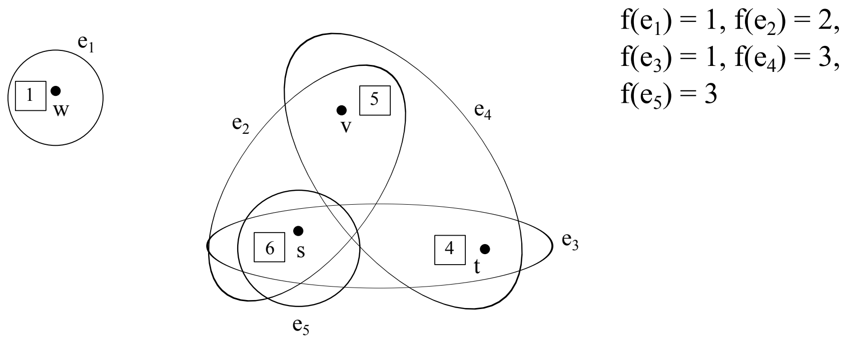

Figure 4.4: Weighting model

4.2.6. PROPOSITION. *A counting model is a special case of a weighting model in which* $f(e) = 1$ *for all* $e \in E$.

4.2.7. PROPOSITION. *Every weighting model* $(S, E, f, \|\cdot\|, s_0)$ *is a justification model* $(S, E, \preceq, \|.\|, s_0)$.

4.2.8. PROOF. We prove that weighting models (and so counting models) satisfy the three constraints which the pre-order must satisfy for it to be a justification model. First note that since we use the order on natural numbers, transitivity follows.

– if $F \subseteq F'$ then $f(F) = \sum_{e \in F} f(e) \leq \sum_{e \in F'} f(e) = f(F')$ that is, $F \preceq F'$.

– Let $F \preceq F', G \preceq G'$ and $F' \cap G' = \emptyset$.
$F \cup G = \sum_{e \in F \cup G} f(e) = \sum_{e \in F} f(e) + \sum_{e \in G} f(e) - \sum_{e \in F \cap G} f(e)$.
Moreover $F' \cup G' = \sum_{e \in F' \cup G'} f(e) = \sum_{e \in F'} f(e) + \sum_{e \in G'} f(e) - \sum_{e \in F \cap G} f(e)$.
Since $F' \cap G' = \emptyset$ then $F' \cup G' = \sum_{e \in F' \cup G'} f(e) = \sum_{e \in F'} f(e) + \sum_{e \in G'} f(e)$.
Since $F \preceq F', G \preceq G'$ then $f(F) + f(G) - \sum_{e \in F \cap G} f(e) \leq f(F') + f(G')$.
It means that $F \cup G \preceq F' \cup G'$.

– Let $F \prec F', G \preceq G'$ and $F' \cap G' = \emptyset$.
$F \cup G = \sum_{e \in F \cup G} f(e) = \sum_{e \in F} f(e) + \sum_{e \in G} f(e) - \sum_{e \in F \cap G} f(e)$.
Moreover $F' \cup G' = \sum_{e \in F' \cup G'} f(e) = \sum_{e \in F'} f(e) + \sum_{e \in G'} f(e) - \sum_{e \in F \cap G} f(e)$.
Since $F' \cap G' = \emptyset$ then $F' \cup G' = \sum_{e \in F' \cup G'} f(e) = \sum_{e \in F'} f(e) + \sum_{e \in G'} f(e)$.
Then $f(F) + f(G) - \sum_{e \in F \cap G} f(e) < f(F') + f(G')$.
It means that $F \cup G \prec F' \cup G'$.

□

4.2.4 Evidence models

Recall that we introduced the setting of van Benthem and Pacuit in Chapter 2. Before we continue, we need to introduce one more notion, that's the notion of an introspective evidence model:

4.2.9. DEFINITION. An evidence model $\mathcal{M}$ is *introspective* iff we have sEX iff tEX for all $s, t \in S$ and for all $X \subseteq S$.

Put in another way, we have $\boxminus\varphi \Rightarrow K \boxminus \varphi$ and $\neg \boxminus \varphi \Rightarrow K \neg \boxminus \varphi$.[7]

4.2.10. PROPOSITION. *Introspective evidence models are a special kind of justification models namely, they correspond exactly to those justification models in which the pre-order on bodies of evidence is given by inclusion.*

In an introspective evidence model, the evidence relation E boils down to our notion of E that is, it becomes a family of evidence sets $E \subseteq \mathcal{P}(S)$ such that $E_s = \{e \mid s \in E\}$ for any $s \in S$. Moreover, the notions of irrevocable knowledge (K), belief (B) and conditional belief (B^-) defined in evidence models (see Chapter 2) do exactly correspond to the notions we defined in Section 4.1.

[7]We recall that irrevocable knowledge has been defined on evidence models in Chapter 2.

4.2.5 Refined justification models

The evidence sets of a refined justification model may be of two types: genuine evidence and biases (or defaults). In refined justification models, we consider that an agent may have some preferences that are not genuinely based on evidence but based on the trustworthiness of his senses and reason. For example, I prefer to believe that I am actually writing my thesis, sitting at my desk instead of being in my bed, dreaming that I am writing my thesis, even if I have no real evidence to justify this belief. Bias comes from inside the agent. The biases are all the pieces of evidence the agent has because she trusts her senses or her reason according to her own experience. We do note that the difference between genuine evidence and a bias is only a generic difference. Both types of evidence sets behave in the same way in our models.

In refined justification models, the definition of E is the following:

4.2.11. Definition. $E = E_0 \cup B$ such that E_0 is the family of evidence sets representing the genuine evidence the agent has while B is the family of evidence sets representing the biases of the agent.

We put some conditions on B: all biases $b \in B$ should strictly increase the strengh of a body of evidence. Formally, $F < F \cup \{b\}$ such that $F \cup \{b\} \in \mathcal{E}$.

Let us take the famous example of Tweety. Consider an agent who learns that Tweety is a bird. According to the experience of the agent, birds typically fly. Indeed she saw many birds flying and she remembers seeing many birds flying. Even if the agent has no genuine evidence about Tweety himself, she has some prior experiences about birds. Since our agent trusts her senses and her memory, she prefers to believe that Tweety flies instead of believing that Tweety does not fly: "Tweety flies" is a bias of the agent.

4.2.6 Important notions in justification models

We remind the reader of our previous definitions, indicating the fact that every argument $F \in \mathcal{E}$, that every evidence set $e \in E$ and that every state $s \in S$.

4.2.12. Definition. An argument F is *sound* at s iff $s \in \bigcap F$.

Note that the empty argument $\varnothing$ is always sound at every state s since $s \in \bigcap \varnothing = S$.

4.2.13. Definition. An argument F *supports* Q (or F is an *argument for* Q) iff $\bigcap F \subseteq Q$.

In a refined justification model, we can consider a "softer" kind of support.

4.2.14. Definition. In a refined justification model, an argument F *weakly supports* Q (or F is a *"soft" argument for* Q) conditional on some set $B' \subseteq B$ of biases iff $F \cup B'$ supports Q.

4.2.15. Definition. A *justification for* Q is an argument F such that all arguments at least as strong as F support Q, i.e. $\forall F'(F \leq F' \Rightarrow \bigcap F' \subseteq Q)$.

4.2.16. Definition. An argument F *supports* Q *conditional on* P (or F is an *argument for* Q *conditional on* P) iff $\bigcap F \cap P \subseteq Q$.

4.2.17. Definition. A *justification for* Q *given* P is an argument F that is consistent with P such that all arguments at least as strong as F support Q conditional on P, i.e. $\bigcap F \cap P \neq \varnothing$ and $\forall F'(F \leq F' \Rightarrow \bigcap F' \cap P \subseteq Q)$.

4.3 Dynamics of Justification Models

We consider the case in which an agent is confronted with new incoming information and accommodates this new information. To model this, we adopt the standard view in Dynamic Epistemic Logic (see Chapter 2). We consider here *updates* of justification models.

4.3.1. Definition. Given a justification model $\mathcal{M} = (S, E, \leq, \|\cdot\|, s_0)$ and some subset $P \subseteq S$, we define the relativization of the justification model $\mathcal{M}$ to P as $\mathcal{M}|P = (S', E', \leq', \|\cdot\|', s_0')$ with:

$$S' = P$$

$$E' = \{e \cap S' \mid e \in E,\ e \cap S' \neq \varnothing\}$$

$$F' \leq' G' \text{ iff } \{e \in E \mid e \cap S' \in F'\} \leq \{e \in E \mid e \cap S' \in G'\}$$

$$\|\cdot\|' = \|\cdot\| \cap S'$$

$$s_0' = s_0$$

where the new set of states S' is now reduced to the set of states satisfying P. The new evidence set E' is taken to be the old evidence E that is consistent with the states surviving the update and the order $\leq'$ on new bodies of evidence $F' \leq' G'$ reflects the fact that the new evidence within G' is at least as strong as the new evidence in F'.

We define the restriction $F|\mathcal{M}'$ of the argument F to the justification model $\mathcal{M}'$ as follows.

4.3.2. Definition. If F is an argument for a justification model $\mathcal{M}$, and if $\mathcal{M}' = \mathcal{M}|P$ is the relativization of the justification model $\mathcal{M}$ to a subset P, then $F|\mathcal{M}'$ is a body of evidence for the model $\mathcal{M}'$ (called the restriction $F|\mathcal{M}'$ of the argument F to model $\mathcal{M}'$), defined by

$$F|\mathcal{M}' = \{e \cap S' \mid e \in F\}$$

Suppose we are given a language $\mathcal{L}$ for a class of justification models $\mathcal{M}$.

4.3.3. Definition. Let φ be any formula in $\mathcal{L}$. When new hard evidence φ is received, this induces an update $!\varphi$, which changes the agent's prior justification model $\mathcal{M} = (S, E, \leq, \|\cdot\|, s_0)$ to the justification model $\mathcal{M}|\varphi = (S', E', \leq', \|\cdot\|', s_0')$ with:

$$S' = \|\varphi\|_{\mathcal{M}}$$

$$E' = \{e \cap S' \mid \ e \in E, \ e \cap S' \neq \varnothing\}$$

$$F' \leq' G' \text{ iff } \{e \in E \mid e \cap S' \in F'\} \leq \{e \in E \mid e \cap S' \in G'\}$$

$$\|\cdot\|' = \|\cdot\| \cap S'$$

$$s_0' = s_0$$

where the new set of states S' is now reduced to the set of states satisfying the new information φ. The new evidence set E' is taken to be the old evidence E that is consistent with the states surviving the update and the order $\leq'$ on new bodies of evidence $F' \leq' G'$ reflects the fact that the new evidence within G' is at least as strong as the new evidence in F'.

It is easy to see that if $\mathcal{M}$ is a justification model $(S, E, \leq, \|\cdot\|, s_0)$ and $\mathcal{M}|\varphi$ is the justification model $(S', E', \leq', \|\cdot\|, s_0)$ that is the result of updating $\mathcal{M}$ with $!\varphi$ then we have $\leq_E' = \leq_E \cap (S' \times S')$. The new plausibility relation on states $\leq_E'$ after the update with φ is exactly the result of updating the old plausibility relation $\leq_E$ with φ.

There are other ways to transform a given justification model into a new one, for instance by just adding a new body of evidence and giving it a degree of plausibility in relation to the other bodies of evidence. This would correspond to obtaining soft evidence of which the agent is not fully certain. The theory of soft evidence upgrades is interesting to be worked out in itself, this we'll do in future work.

4.3.1 Plausibility models

When restricting to sphere-based justification models, our update operation coincides with the usual update on plausibility models (see Chapter 2):

$$\leq'_E = \leq_E \cap (S' \times S')$$

For $s \in S'$, we have $E_s = \{e \in E \mid s \in e\}$ and $E'_s = \{e \cap S' \mid e \in E_s\}$. Note that from $e \in E_s$ and $s \in S'$, we have $s \in e \cap S' \neq \emptyset$.

For $s, t \in S'$, we have:

$$\begin{aligned} s \leq'_E t \iff E'_t \preceq' E'_s &\iff \{e \cap S' \mid e \in E_t\} \preceq' \{e \cap S' \mid e \in E_s\} \\ &\iff \{e \cap S' \mid e \in E, t \in e\} \preceq' \{e \cap S' \mid e \in E, s \in e\} \\ &\iff \{e \mid e \in E, t \in e\} \preceq \{e \mid e \in E, s \in e\} \\ &\iff E_t \preceq E_s \\ &\iff s \leq_E t \end{aligned}$$

4.3.2 Counting models

4.3.4. Proposition. *The class of counting models is not closed under update. When a given counting model is updated, the result of updating is a justification model but not necessarily a counting model.*

We present an example of such a problem below.

Consider the counting model depicted in Figure 4.5 where there are three pieces of evidence e_1, e_2 and e_3 and 4 states s, t, u and v. The most plausible states are the states s and u since $E_s := \{e_1, e_2\}$, $E_t := \{e_1\}$, $E_u := \{e_1, e_3\}$, $E_v := \{e_3\}$ and so $\mid E_s \mid < \mid E_t \mid$, $\mid E_s \mid < \mid E_v \mid$, $\mid E_u \mid < \mid E_t \mid$ and $\mid E_u \mid < \mid E_v \mid$.

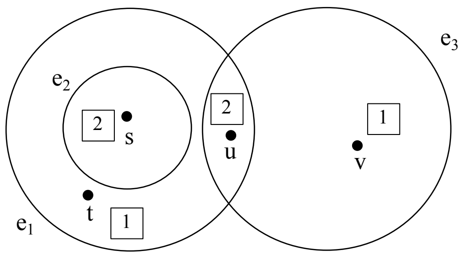

Figure 4.5: Initial counting model

What happens now if we want to update this counting model? It is easily to see that a problem arises when dealing with (some) updates on this model. Suppose that the agent receives the hard information that P such that P is only true in s and v. Next, the model is updated with $!P$ and the states u and t

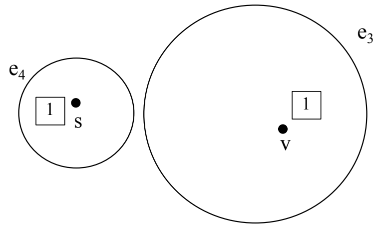

Figure 4.6: Counting updated model

are deleted as illustrated in Figure 4.6. Then s and v are equiplausible since $E_v := \{e_3\}$, $E_s := \{e_4\}$ and so $\mid E_s \mid = \mid E_w \mid$.

But this should not be the case, we should have $\mid E_s \mid < \mid E_v \mid$ that is, we should have the justification model as depicted in Figure 4.7.

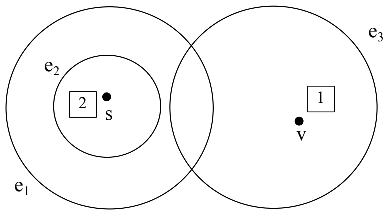

Figure 4.7: Labelling solution

The problem is clearly visible if we use plausibility models. We first provide the corresponding initial (total) plausibility model in Figure 4.8, and then represent the updated plausibility model after the update with P in Figure 4.9. Here it is obvious that the state s is still more plausible than the state v.

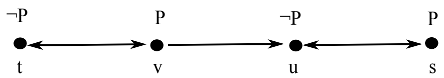

Figure 4.8: Initial plausibility model

4.3.3 Weighting models

One solution to the problem in Proposition 4.3.4 is to deal with weighting models instead of counting models.

4.3.5. Definition. We define the *map* Wei, mapping weighting models to justification models: $(S, E, f, \|\cdot\|, s_0) \overset{\text{Wei}}{\longmapsto} (S, \leq, \|.\|, s_0)$.

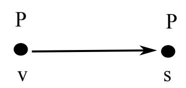

Figure 4.9: Plausibility model updated

4.3.6. PROPOSITION. *The class of weighting models is closed under update:*

$$Wei(\mathcal{M})|\varphi = Wei(\mathcal{M}|\varphi)$$

.

4.3.7. PROOF. Let $(S, E, f, \|\cdot\|, s_0)$ be a weighting model $\mathcal{M}$ where $f : E \to \mathbb{N}$. We define the result of updating this model with φ. The weighting model $\mathcal{M} = (S, E, f, \|\cdot\|, s_0)$ is changed to the weighting model $\mathcal{M}|\varphi = (S', E', f', \|\cdot\|', s'_0)$ with:

$$S' = \|\varphi\|_S$$

$$E' = \{e \cap S' \mid \ e \in E,\ e \cap S' \neq \varnothing\}$$

$$f'(e') = \sum\{f(e) \mid e \in E \text{ such that } e \cap S' \neq \varnothing\}$$

$$\|\cdot\|' = \|\cdot\| \cap S'$$

$$s'_0 = s_0$$

□

Coming back to Figure 4.5, we put $f(e_1) = f(e_2) = f(e_3) = 1$. Then after the update with P, $v < w$ since as depicted in Figure 4.7, $f(e_1) + f(e_2) > f(e_3)$.

4.3.4 Evidence models

On evidence models, the update operation coincides with the update defined by van Benthem and Pacuit in [16].

4.4 Language and axiomatization

4.4.1 The language of evidence logic

We can re-use the language introduced by van Benthem and Pacuit in [16] for our justification models.

Syntax

Assume given any static object language $\mathcal{L}_{EV}$ containing propositional letters coming from a set Φ, Boolean connectives, a conditional belief operator B^{-} and an evidence operator $\boxminus$. This evidence operator comes from the logic of van Benthem and Pacuit but can be used in justification models in general.

Dynamic modalities $[!\varphi]$ are added to $\mathcal{L}_{EV}$ to obtain a dynamic language.

Axiomatization

The axioms of van Benthem and Pacuit [16](see Chapter 2) hold for general justification models.

4.4.1. Theorem. *The Reduction axioms of van Benthem and Pacuit are sound in the class of all justification models.*

4.4.2. Proof. The proof of the first axioms is straightforward, we focus here only on the proof of the reduction axiom for conditional belief. Assume a justification frame satisfying our three conditions. We now show that $[!\varphi]B^{\psi}\theta \iff \varphi \to B^{\varphi\wedge[!\varphi]\psi}([!\varphi]\theta)$ holds in such a frame.

1. In the direction from left to right we assume as given a model $\mathcal{M}$ and a state s and assume that $[!\varphi]B^{\psi}\theta$ is true at s. To show that $\varphi \to B^{\varphi\wedge[!\varphi]\psi}([!\varphi]\theta)$ is also true at s, we need to assume that if φ is true at s then $B^{\varphi\wedge[!\varphi]\psi}([!\varphi]\theta)$ is true at s. So assume φ is true at s. To show that $B^{\varphi\wedge[!\varphi]\psi}([!\varphi]\theta)$ is true at s we need to show that $[!\varphi]\theta$ is true at t for all most plausible states t where $\varphi \wedge [!\varphi]\psi$ is true. Thus, we need to show that for all most plausible states t such that φ is true at t and $[!\varphi]\psi$ is true at t (that is, ψ is true at t), θ is true at t.

   We have assumed that $[!\varphi]B^{\psi}\theta$ and φ are true at s, hence $B^{\psi}\theta$ is true at s. And $B^{\psi}\theta$ is true at s iff θ is true at t for all most plausible states t such that ψ is true at t.

   So we have shown that θ is true at t for all most plausible states t such that ψ is true at t and we are done.

2. In the direction from right to left we assume as given a model $\mathcal{M}$ and a state s and assume that $\varphi \to B^{\varphi\wedge[!\varphi]\psi}([!\varphi]\theta)$ is true at s. We need to show that $[!\varphi]B^{\psi}\theta$ is also true at s. We need to show that either $\neg\varphi$ is true at s or if φ is true at s, $B^{\psi}\theta$ is true at s. Thus, if φ is true at s, we have to show that θ is true at t for all most plausible states t such that ψ is true at t.

   We know that $\varphi \to B^{\varphi\wedge[!\varphi]\psi}([!\varphi]\theta)$ is true at s. So either $\neg\varphi$ is true at s or $B^{\varphi\wedge[!\varphi]\psi}([!\varphi]\theta)$ is true at s.

– assume $\neg\varphi$ is true at s, then $[!\varphi]B^{\psi}\theta$ is also true at s by definition of the update modality.

– assume $B^{\varphi\wedge[!\varphi]\psi}([!\varphi]\theta)$ is true at s, then for all most plausible states t such that $\varphi\wedge[!\varphi]\psi$ is true at t, $[!\varphi]\theta$ is true at t. Then, for all most plausible states t such that φ is true at t and $[!\varphi]\psi$ is true at t (that is, ψ is true at t), θ is true at t and we are done.

□

4.4.2 The language of justification logic

The language $\mathcal{L}_{EV}$ is not expressive enough to capture the main interesting features of justification models. We need to introduce a more expressive language.

Syntax

The language of justification logic $\mathcal{L}_{JL}$ is defined as follows.

4.4.3. DEFINITION. Let Φ be a set of propositional atoms such that p ranges over Φ.

$$\varphi ::= p \mid \neg\varphi \mid \varphi\wedge\psi \mid K\varphi \mid K_D\varphi \mid \mathit{sound} \mid \forall^{ev}\varphi \mid [\leq]\varphi$$

We can define a belief modality as an abbreviation:

$$B\varphi := K\neg K_D\neg K_D\varphi$$

In total justification models, this boils down to:

$$B\varphi := \neg K_D\neg K_D\varphi$$

We can as well define the following modalities:

$$\mathit{Supp}\,\varphi := K(\mathit{sound}\to\varphi)$$
$$\mathit{Just}\,\varphi := [\leq]\mathit{supp}\,\varphi$$
$$\boxplus\varphi := \exists^{ev}\mathit{supp}\,\varphi$$

The intended interpretation of the operators is as follows:

$K\varphi$: the agent knows that φ

$K_D\varphi$: the agent defeasibly knows that φ

sound: the current argument F is sound (true) at the actual state s, i.e. the current pieces of evidence $e\in F$ are true

$\forall^{ev}\varphi$: for every argument F, φ is the case

$[\leq]\varphi$: for every argument F' at least as convincing as the current argument F, φ is the case

$Supp\,\varphi$: the current argument F supports φ

$Just\,\varphi$: the current argument is a justification for φ

$\boxplus\varphi$: there exists an argument supporting φ

Semantics

The formulas are interpreted at a state s and a body of evidence F such that F is the current argument.

4.4.4. DEFINITION. Given a justification model $\mathcal{M}$, a semantics for $\mathcal{L}_{JL}$ is required to satisfy the following constraints:

$$\begin{array}{lll}
s, F \vDash p & \text{iff} & s \in V(p) \\
s, F \vDash \neg\varphi & \text{iff} & s, F \nvDash \varphi \\
s, F \vDash \varphi \wedge \psi & \text{iff} & (s, F \vDash \varphi) \wedge (s, F \vDash \psi) \\
s, F \vDash K\varphi & \text{iff} & t, F \vDash \varphi \text{ for every } t \in S \\
s, F \vDash K_D\varphi & \text{iff} & t, F \vDash \varphi \text{ for every } t \in S \text{ such that } t \leq s \\
s, F \vDash sound & \text{iff} & s \in \bigcap F \\
s, F \vDash \forall^{ev}\varphi & \text{iff} & s, F' \vDash \varphi \text{ for every } F' \in \mathcal{E} \\
s, F \vDash [\leq]\varphi & \text{iff} & \forall F'(F \leq F' \Rightarrow s, F' \vDash \varphi)
\end{array}$$

Axiomatization

All the axioms of the logic KK_D (see Chapter 2) hold for total justification models.

4.4.5. THEOREM. *A sound (but not complete) proof system for the logic JL with the language $\mathcal{L}_{JL}$ over the class of total justification models is given by the axioms and rules of $\mathcal{L}_{KK_D}$ plus the following reduction axioms:*

Necessitation Rules for both $\forall^{ev}$ and $[\leq]$

$S5$-axioms for $\forall^{ev}$

$S4$-axioms for $[\leq]$

$[\leq]K\varphi \to K[\leq]\varphi$

$\forall^{ev}K\varphi \to K\forall^{ev}\varphi$

$\forall^{ev}K_D\varphi \to K_D\forall^{ev}\varphi$

$\forall^{ev}\varphi \to [\leq]\varphi$

$\forall^{ev}(\varphi \vee [\leq]\psi) \wedge \forall^{ev}(\psi \vee [\leq]\varphi) \to \forall^{ev}\varphi \vee \forall^{ev}\psi$

The last axiom is the analogue of the axiom *Totality* for bodies of evidence (arguments).

We do not have a complete system for the logic *JL* over the class of total justification models. We would need more axioms to connect doxastic and epistemic modalities with evidence modalities. Obtaining a complete system is still ongoing work.

4.5 Justifiable beliefs

All the next statements can be encoded in the formal syntax of the language $\mathcal{L}_{JL}$.

4.5.1 Belief in justification models

4.5.1. PROPOSITION. *An agent believes Q iff every argument can be strengthened to a justification for Q, i.e.*

$$\forall F \exists F' \geq F (\forall F'' \geq F' (\bigcap F'' \subseteq Q))$$

This fact can be captured by the following validity:

$$Bp \iff \forall^{ev} \langle \leq \rangle just\, p$$

Or writing it more explicitly:

$$Bp \iff \forall^{ev} \langle \leq \rangle [\leq] supp\, p$$

To prove Proposition 4.5.1, we first state and prove Lemma 4.5.2.

4.5.2. LEMMA. *An agent believes Q iff all maximal (in the sense of strength order) arguments supports Q, i.e.*

$$\forall F \in Max_{\leq}\mathcal{E}(\bigcap F \subseteq Q)$$

where $Max_{\leq}\mathcal{E} = \{F \in \mathcal{E} \mid F \nless F' \text{ for any } F' \in \mathcal{E}\}$.

4.5.3. PROOF. – In the direction from left to right, we start from a given justification model $\mathcal{M}$ in which BQ is true at s. So we know that $bestS \subseteq Q$. Let $F \in Max_{\leq}\mathcal{E}$ and $t \in \bigcap F$. Then $F \subseteq E_t$, so $F \leq E_t$. Suppose $t \notin bestS$. Then $\exists w <_E t$, so $E_t < E_w$, so $F < E_w$. This contradicts $F \in Max_{\leq}\mathcal{E}$. Then $t \in bestS$, so $\bigcap F \subseteq bestS$. Hence $\bigcap F \subseteq Q$.

– In the direction from right to left we assume as given a justification model $\mathcal{M}$ and a state s such that $\forall F \in Max_{\leq}\mathcal{E}(\bigcap F \subseteq Q)$. Let $t \in bestS$. Suppose $E_t \notin Max_{\leq}\mathcal{E}$. Then $\exists F' \in \mathcal{E}$ such that $E_t < F'$. Let $w \in \bigcap F'$, so $F' \subseteq E_w$, so $F' \leq E_w$. Then $E_t < E_w$, so $w <_E t$. This contradicts $t \in bestS$. Hence, $E_t \in Max_{\leq}\mathcal{E}$. Then, $t \in \bigcap E_t \subseteq Q$. Hence $t \in Q$, so $bestS \subseteq Q$. Hence, BQ is true at s.

□

Now we can prove Proposition 4.5.1.

4.5.4. PROOF. – In the direction from left to right, we start from a given justification model $\mathcal{M}$ in which BQ is true at s. Let $F \in \mathcal{E}$. Then F can be strengthened to a maximal argument F', i.e. $\exists F' \geq F(F' \in Max_{\leq}\mathcal{E})$. Indeed since S is finite, so is $\mathcal{E}$. By Lemma 4.5.2, since BQ is true at s and $F' \in Max_{\leq}\mathcal{E}$, $\bigcap F' \subseteq Q$. So F' supports Q. Let $F'' \geq F'$. Then $F'' \in Max_{\leq}\mathcal{E}$ and by Lemma 4.5.2, $\bigcap F'' \subseteq Q$. So F'' supports Q. Then, F' is a justification for Q. Hence, F can be strengthened to a justification for Q.

– In the direction from right to left we assume as given a justification model $\mathcal{M}$ and a state s such that $\forall F \exists F' \geq F(\forall F'' \geq F'(\bigcap F'' \subseteq Q))$. Let $F \in Max_{\leq}\mathcal{E}$. Then $F \geq F'$. Take $F'' := F$. Hence, $\bigcap F \subseteq Q$. By Lemma 4.5.2, BQ is true at s.

□

4.5.5. PROPOSITION. *An agent believes Q conditional on P iff every argument consistent with P can be strengthened to a justification for Q given P, i.e.*

$$\forall F(\bigcap F \cap P \neq \varnothing \Rightarrow \exists F' \geq F(\bigcap F' \cap P \notin \varnothing \wedge \forall F'' \geq F'(\bigcap F'' \cap P \subseteq Q)))$$

To prove Proposition 4.5.5, we first state and prove Lemma 4.5.6.

4.5.6. LEMMA. *An agent believes Q conditional on P iff all maximal (in the sense of strength order) arguments consistent with P supports Q conditional on P, i.e.*

$$\forall F \in \mathcal{E}(F \in Max^P_{\leq}\mathcal{E} \Rightarrow \bigcap F \cap P \subseteq Q)$$

where $Max^P_{\leq}\mathcal{E} = \{F \in \mathcal{E} \mid \bigcap F \cap P \neq \varnothing$ *and* $F \nless F'$ *for any* $F' \in \mathcal{E}(\bigcap F' \cap P \neq \varnothing)\}$.

4.5.7. PROOF. – In the direction from left to right, we start from a given justification model $\mathcal{M}$ in which $B^P Q$ is true at s. So we know that $bestP \subseteq Q$. Let $F \in Max^P_{\leq}\mathcal{E}$ and $t \in \bigcap F \cap P$. Then $F \subseteq E_t$, so $F \leq E_t$. Suppose $t \notin bestP$. Then $\exists w <_E t$, so $E_t < E_w$, so $F < E_w$. This contradicts $F \in Max^P_{\leq}$. Then $t \in bestP$. So we proved that $\forall t(t \in \bigcap F \cap P \Rightarrow t \in bestP)$. Hence, $\bigcap F \cap P \subseteq bestP$, i.e. $\bigcap F \cap P \subseteq Q$.

– In the direction from right to left we assume as given a justification model $\mathcal{M}$ and a state s such that $\forall F \in \mathcal{E}(F \in\in Max^P_{\leq}\mathcal{E} \Rightarrow \bigcap F \cap P \subseteq Q)$. Let $t \in bestP$. Then $\bigcap E_t \cap P \neq \varnothing$. Suppose $E_t \notin Max^P_{\leq}\mathcal{E}$. Then $\exists F' \in \mathcal{E}$ such that $(\bigcap F' \cap P \neq \varnothing)$ and $E_t < F'$. Let $w \in \bigcap F' \cap P$, so $F' \subseteq E_w$, so $F' \leq E_w$. Then $E_t < E_w$, so $w <_E t$. This contradicts $t \in bestP$. Hence, $E_t \in Max^P_{\leq}\mathcal{E}$. Then, $t \in \bigcap E_t \cap P \subseteq Q$. Hence $t \in Q$. So we proved that $\forall t(t \in bestP \Rightarrow t \in Q)$. Hence, $bestP \subseteq Q$, i.e. $B^P Q$ is true at s.

□

Now we can prove Proposition 4.5.5.

4.5.8. Proof. – In the direction from left to right, we start from a given justification model $\mathcal{M}$ in which $B^P Q$ is true at s. Let $F \in \mathcal{E}$ such that F is consistent with P, i.e. $\bigcap F \cap P \neq \varnothing$. Then F can be strengthened to a maximal argument F' consistent with P, i.e. $\exists F' \geq F(F' \in Max^P_{\leq}\mathcal{E})$. Indeed since S is finite, so is $\mathcal{E}$. By Lemma 4.5.6, since $B^P Q$ is true at s and $F' \in Max^P_{\leq}\mathcal{E}$, $\bigcap F' \cap P \subseteq Q$. So F' supports Q conditional on P. Let $F'' \geq F'$. Then $F'' \in Max^P_{\leq}\mathcal{E}$ and by Lemma 4.5.6, $\bigcap F'' \cap P \subseteq Q$. So F'' supports Q conditional on P. Then, F' is a justification for Q given P. Hence, F can be strengthened to a justification for Q given P.

– In the direction from right to left we assume as given a justification model $\mathcal{M}$ and a state s such that $\forall F(\bigcap F \cap P \neq \varnothing \Rightarrow \exists F' \geq F(\bigcap F' \cap P \notin \varnothing \wedge \forall F'' \geq F'(\bigcap F'' \cap P \subseteq Q)))$. Let $F \in Max^P_{\leq}\mathcal{E}$. Then $F \geq F'$. Take $F'' := F$. Hence, $\bigcap F \cap P \subseteq Q$. By Lemma 4.5.6, $B^P Q$ is true at s.

□

4.5.2 Knowledge and belief in total justification models

From now on, we restrict ourselves to justification models with a total pre-order. In total justification models, a belief is a justified belief.

4.5.9. Proposition. *In total justification models, an agent believes Q iff there exists a justification F for Q, i.e.*

$$\exists F \forall F' \geq F(\bigcap F' \subseteq Q)$$

This fact can be captured by the following validity:

$$Bp \iff \exists^{ev} just\, p$$

Or writing it more explicitly:

$$Bp \iff \exists^{ev}[\leq] supp\, p$$

4.5.10. PROOF. – In the direction from left to right, we start from a given total justification model $\mathcal{M}$ in which BQ is true at s. By Proposition 4.5.1, every argument can be strengthened to a justification for Q. Take any argument and strengthen it, then we have a justification for Q.

– In the direction from right to left we assume as given a total justification model $\mathcal{M}$ and a state s such that there exists a justification F for Q. We have to show that $s \vDash BQ$. Since F is a justification for Q, by Definition 4.2.15, $\forall F' \in \mathcal{E}(F \preceq F' \Rightarrow \bigcap F' \subseteq Q)$. Take any argument G such that $F \preceq G$ or $G \preceq F$. If $F \preceq G$, $\bigcap G \subseteq Q$. If $G \preceq F$, G can be strengthened to a justification for Q since $G \preceq F$ and $\forall F' \in \mathcal{E}$ such that $F \preceq F', \bigcap F' \subseteq Q$. By Proposition 4.5.1, BQ is true at s.

□

4.5.11. PROPOSITION. *In total justification models, an agent defeasibly knows Q at s iff there exists a sound (true) justification F for Q at s, i.e.*

$$\exists F(s \in \bigcap F \wedge \forall F' \succeq F(\bigcap F' \subseteq Q))$$

.

This fact can be captured by the following validity:

$$K_D p \iff \exists^{ev}(sound \wedge just\, p)$$

4.5.12. PROOF. – In the direction from left to right, we start from a given total justification model $\mathcal{M}$ in which $K_D Q$ is true at s. Then $\forall t(t \leq_E s \Rightarrow t \in Q)$. So, $\forall t(E_s \preceq E_t \Rightarrow t \in Q)$. Take $F := E_s$. Since $s \in \bigcap E_s$, $s \in \bigcap F$. Let $F' \succeq F$ and $t \in \bigcap F'$. Then $F' \subseteq E_t$, so $F' \preceq E_t$. Since $E_s \preceq F'$, $E_s \preceq E_t$. Then $t \leq_E s$, so $t \in Q$. Hence, $\exists F(s \in \bigcap F \wedge \forall F' \succeq F(\bigcap F' \subseteq Q))$.

– In the direction from right to left we assume as given a total justification model $\mathcal{M}$ and a state s such that $\exists F(s \in \bigcap F \wedge \forall F' \succeq F(\bigcap F' \subseteq Q))$. Let $t \leq_E s$. We want to show that $t \in Q$. As we know, $s \in \bigcap F$. Then $F \subseteq E_s$, so $F \preceq E_s$. Since $t \leq_E s$, $E_s \preceq E_t$. Then $F \preceq E_t$. By assumption, $\bigcap E_t \subseteq Q$. So, $t \in Q$. Hence, $K_D Q$ is true at s.

□

4.5.13. PROPOSITION. *An agent believes Q conditional on P iff there exists a justification F for Q given P, i.e.*

$$\exists F(\bigcap F \cap P \neq \varnothing \wedge \forall F' \succeq F(\bigcap F' \cap P \subseteq Q))$$

.

4.5.14. PROOF. – In the direction from left to right, we start from a given total justification model $\mathcal{M}$ in which $B^P Q$ is true at s. By Proposition 4.5.5, every argument consistent with P can be strengthened to a justification for Q given P. Take any argument consistent with P and strengthen it, then we have a justification for Q given P.

– In the direction from right to left we assume as given a total justification model $\mathcal{M}$ and a state s such that there exists a justification F for Q given P. We have to show that $s \vDash B^P Q$. Since F is a justification for Q given P, by Definition 4.2.17, $\bigcap F \cap P \neq \emptyset$ and $\forall F' \in \mathcal{E}(F \leq F' \Rightarrow \bigcap F' \cap P \subseteq Q)$. Take any argument G consistent with P ($\bigcap G \cap P \neq \emptyset$) such that $F \leq G$ or $G \leq F$. If $F \leq G$, $\bigcap G \cap P \subseteq Q$. If $G \leq F$, G can be strengthened to a justification for Q given P since $G \leq F$ and $\forall F' \in \mathcal{E}$ such that $F \leq F', \bigcap F' \cap P \subseteq Q$. By Proposition 4.5.5, $B^P Q$ is true at s.

□

Conclusion

We have provided a very general setting which can encompass many other settings. Indeed justification models subsume plausibility models, counting and weighting models as well as evidence models.

In the next chapter we want to use these justification models to fix the agent's justifications, irrevocable knowledge, beliefs, conditional beliefs, strong beliefs and defeasible knowledge. This will be a key ingredient of the game semantics (for defeasible knowledge) we introduce in this next chapter.

Chapter 5

Playing for knowledge

Aim In this chapter our aim is to provide a game semantics for the "justification games" used to define Keith Lehrer's notion of "defeasible knowledge". Indeed while the specifics of Lehrer's system were not formalized in logical terms and that has left philosophers to settle the misunderstandings via argumentation and on-going philosophical debates, we believe that a formalisation of this type of game can lead to useful insights into the different accounts of knowledge within formal epistemology.

Summary In this chapter we focus on the many definitions of the concept of "knowledge". We first present the traditional understanding of knowledge as "justified true belief" as well as Edmund Gettier's counterexamples to this conception [37]. Then we introduce the notion of defeasible knowledge theorized by Keith Lehrer in [46, 47]. Finally we formalise Lehrer's conception providing a game semantics for "defeasible knowledge". The main points are:

– we discuss the various definitions of knowledge, in particular the definition of knowledge as "justified true belief". We introduce the famous Gettier's counterexamples shattering this conception of knowledge.

– we develop the notion of "defeasible knowledge" in the form that was theorized by Lehrer. We introduce the "ultra-justification game" as an essential ingredient of Lehrer's informal account of knowledge as "undefeated justified acceptance".

– we use total justification models to offer a qualitative representation of an agent's information and justification within the framework of Dynamic Epistemic Logic.

– Finally we propose a "game semantics for defeasible knowledge", as a formalization of Lehrer's conception. We apply our formal model to some examples, discussing the limits of this formalisation and indicating some possible ways to overcome them.

Background In the traditional literature, there is a common understanding that the famous definition of knowledge as "justified true belief" can/should be credited to Plato in *Meno* and *Theaetetus*. Edmund Gettier himself notes in [37] that Plato seems to consider such a "definition [of knowledge] at *Theaetetus* 201, and perhaps accepting one at *Meno* 98". However Rohit Parikh rightly points out that Socrates presents an objection to this conception of knowledge in the *Theaetetus*. Thus he shows that contrary to common belief, Plato does not endorse this definition of knowledge.

In [37] Gettier exposes two counter-examples to that definition claiming that having a true justified belief about a given proposition is not sufficient for someone's knowing this proposition. Indeed these counter-examples show that even true justified beliefs, instead of being real knowledge, can just be lucky guesses. Gettier's counter-examples are widely accepted by epistemologists as proving that the analysis of knowledge must be modified.

One of the strategies followed by epistemologists to solve the so-called Gettier problem is to find a suitable condition to the definition of knowledge such that knowledge is a justified true belief plus "something". Keith Lehrer provides such a condition in [46, 47]. Indeed he defines knowledge as undefeated justified acceptance (true belief).

In this chapter we use the setting of justification models as described in Chapter 4.

5.1 Is knowledge justified true belief?

It is generally accepted in contemporary Epistemology that the partitional model of knowledge (first proposed by Hintikka [41] in terms of equivalence relations, and later rediscovered by Aumann) does not provide an adequate picture for "knowledge", as the term is used in day-to-day life or even in empirical science. From the 1960s when Gettier's counterexamples [37] shattered the traditional understanding of knowledge as "justified true belief", many distinguished philosophers proposed various concepts of "knowledge" deemed to be closer to the target. In this section, we focus on the notion of "defeasible knowledge" in the form that was theorized by Lehrer [46, 47].

5.1.1 Discussion about the definition of knowledge

Knowledge as justified true belief The most common interpretation of knowledge is that knowledge is a true belief that can be justified. In other words, knowledge is defined as "justified true belief". Thus an agent S knows a proposition p iff:

(1.) p is true

(2.) S believes p

(3.) S is justified in believing that p.

The first condition is known as the truth condition and is not controversial. The content of knowledge must be true that is, false propositions cannot be known. The second condition is called the belief condition. Knowledge encodes a propositional attitude towards a proposition. Finally the last condition is the justification condition. Knowledge is not only true belief because an agent could know a proposition being lucky. This agent knows this proposition only if he is able to justify his belief. In other words only if he can provide reasons for his belief.

Gettier problem In [37] Gettier provides some counter-examples to this tripartite analysis of knowledge. We present here a so-called Gettier counter-example. Consider an agent Smith who is justified to believe (1.) that is, he has some reasons to believes (1.).

(1.) Jones owns a Ford.

Indeed he has evidence that Jones owns a Ford since Jones has offered Smith a ride while driving a Ford, showed Smith some papers stating he owns a Ford and told Smith he owns a Ford.

Now consider an agent Brown who is Smith's friend. Brown took vacations and went abroad. Smith cannot remember where Brown has gone but it could be Boston, Barcelona or Brest-Litovsk.

And consider the following propositions:

(2.) Either Jones owns a Ford, or Brown is in Boston.

(3.) Either Jones owns a Ford, or Brown is in Barcelona.

(4.) Either Jones owns a Ford, or Brown is in Brest-Litovsk.

Each of these propositions is entailed by (1.). Suppose Smith is a perfect logician. Then Smith is completely justified in believing (2.), (3.) and (4.).

Consider now that Jones does not own a Ford, but an Honda. Besides, by a strange coincidence (and remember entirely unknown to Smith), Brown is indeed in Barcelona. Then (2.) is true, Smith believes that (2.) is true (since he believes that Jones owns a Ford) and he is justified in believing that (2.) is true (since he is justified in believing that Jones owns a Ford). However one cannot claim that Smith knows that (2.) is true.

The Gettier problem is posed in terms of a problem in first order logic. The problem is mainly due to the claim that justification is preserved by entailment: if an agent S is justified to believe P and if P entails Q, then S would be justified to believe Q. Thus Gettier claims that the three conditions are not sufficient to define knowledge.

A fourth condition Several philosophers provide answers to the Gettier problem. Most of them add a fourth condition to the conditions of truth, belief and justification[1]. We focus here on the solution provided by Lehrer in [46, 47].

5.1.2 Lehrer's solution to the Gettier problem

Lehrer defines knowledge as undefeated justified acceptance: an agent knows p in case she is justified to accept p and her justification cannot be defeated. He considers several types of justifications varying from a subjective to more objective ones. Both make use of the notions of coherence and reasonableness, but the notion of truth only plays a role in the last one.

[1]Some philosophers provide another condition (4.), thus Alvin Goldman who states that a belief is justified only if this belief has been caused by the truth of another belief, considers that a justified true belief is knowledge if the agent is able to correctly reconstruct the causal chain. However, some philosophers prefer to use another notion of justification or to use a primitive notion of knowledge to solve Gettier problem instead of adding a further condition.

Personal justification The subjective type of justification is called personal justification. An agent is personally justified to accept p at time t iff p is coherent with the agent's "evaluation system" at t that is, it is more reasonable for the agent to accept p than any objection against it on the basis of this evaluation system at t.

Evaluation system The evaluation system of an agent is a collection of three things: an acceptance system, a preference system and a reasoning system. Together this captures both the relevant background information that the agent has acquired about the world in her quest for truth as well as the limited reasoning capacity of the agent. As Lehrer puts it, "the evaluation system of a person consists of what the person accepts, what the person prefers concerning acceptance, and how the person reasons concerning acceptance" [47, p.127]. Then the evaluation system tells what it is more reasonable to accept that is, what sources of information can be trusted (senses, memory...). Note that in the first edition of Lehrer's book, the evaluation system was equated with the acceptance system [46]. But as real agents are not considered to be logically omniscient, it is important in Lehrer's account that one should be able to state the agent's reasoning system explicitly, and similarly the agent's preferences (or conditional acceptances) are an important ingredient when analyzing the agent's knowledge.

Example of personal justification Let's consider an example of a personal justification, that is, an example of coherence with an evaluation system. Imagine I see what looks like a vase on the table in my house, but my reason tells me it could also be a jug of water because they sometimes look like the same. So my senses (my eyes) and my reason disagree on the nature of the object standing on my table. I do not own a jug of water and *I accept that I do not own any.* But someone could have placed one when I left the house. However *I prefer to accept that nobody came into my house than somebody did.* Thus my evaluation system (composed of my acceptance that I do not own any jug of water and my preference concerning my acceptance that nobody came into my house) tells me that it is more reasonable to trust my eyes and accept that it is a vase than to accept it is a jug of water. Of course I can be wrong (somebody can actually came to put this jug of water on my table) but I am personally justified to accept that I see a vase.

Hence, an agent who is personally justified to accept a proposition p at t might be wrong about the truth-value of p but will be right in claiming that p fits well, in the sense of coherence, with the other propositions she accepts. The evaluation system is fallible but according to Lehrer, it has to be used to decide what to accept because relevant information is contained in it. The evaluation system providing personal justification for an acceptance can be in error even though the acceptance itself is true. So we cannot say that a person knows every acceptance to be true that she is personally justified in accepting provided only that it is also true.

Truth-compatible subsystem In order to make the transition to a more objective type of justification, Lehrer introduces in [46] the notion of "complete" justification but replaces this in [47] with the idea of justification on the basis of a truth-compatible subsystem of an evaluation system. Such a truth-compatible subsystem of an original evaluation system contains only the accepted items that are actually true, it deletes those states of preference in which something false is preferred over something that is true and its reasoning system is restricted to sound reasonings.

Undefeated justification This notion of a truth-compatible subsystem plays an essential role in Lehrer's definition of irrefutable or undefeated justification as follows: an agent is justified to accept p in a way that is undefeated at t iff she is justified in accepting p at t on the basis of what Lehrer calls the ultra-system at t (consisting of a truth-compatible subsystem of an original evaluation system at t and the remaining so-called "unmarked" states). Note the difference in the notions of undefeated justification and personal justification by stressing the role played in the former but not the latter by ingredients that are objectively true.

Justification game To clarify these notions of justified acceptance, Lehrer defines for each notion of justification a corresponding justification game. In a justification game, an agent (called the Claimant) claims that she is justified to accept p at time t while an opponent (called a Skeptic or Critic) tries to show that this is in fact not the case.

Personal justification game In the personal justification game, the Skeptic can object to the claim of the Claimant by using an objection (or so-called "competitor" as it was called in the first edition [46]) o to p iff it is more reasonable for the agent to accept that p on the assumption that o is false than on the assumption that o is true on the basis of her evaluation system at t. Then the Claimant has to answer or neutralize the Skeptic to win a round in the game. If she can answer (show that o does not cohere with her evaluation system, that is, it is more reasonable for her to accept p than o on the basis of her evaluation system) or neutralize (show that there is a neutralizing statement n which together with o is not an objection against p and it is as reasonable to accept n together with o as it is to accept o alone on the basis of her evaluation system) all the objections raised by the Skeptic, sshe wins the game. If he wins the game, she is personally justified to accept p.

Ultra-justification game In the ultra-justification game, the agent's payoff is "defeasible knowledge". In this game the opponent (called Ultra-critic) is supposed to be aware of the truth-value of what the Claimant accepts. The objections can be raised in a similar fashion as in the personal justification game, but now such objections can only be met (answered or neutralized) if they happen to refer to truthful pieces of information (only the content of the truth-compatible subsystem and the existence – but not the content – of the unmarked states of the ultra-system can be used). If the Claimant wins the game, she is justified to accept p in a way that is undefeated.

Defeasible knowledge If we adopt Lehrer's definition in [47, p.169] of defeasible knowledge and we use the setting of his justification game to give an explication to condition (4.) below, knowledge will in his setting be reduceable to undefeated justified acceptance.

5.1.1. DEFINITION. *S knows* that p if and only if

(1.) S accepts that p,

(2.) it is true that p,

(3.) S is justified in accepting that p, and

(4.) S is justified in accepting that p in a way that is not defeated by any false statement (that does not depend on any false statement).

Hence if knowledge of p is reduced to undefeated justified acceptance of p, we can say that if the Claimant wins every round of the ultra-justification game then the Claimant knows p (in the (in)defeasible sense of knowledge).

No false lemma We want to clarify the meaning of condition (4.) above. What does it mean "not defeated by any false statement" and "does not depend on any false statement"? How a justification can be defeated by a false statement?

Lehrer insists on the fact that condition (4.) does not imply the simple denial of false statements. Whenever the false statement is the result of a perceptual error or is the premise of some reasoning, condition (4.) does not imply that the justification the agent has to accept p must not contain any false statements (or beliefs). Moreover Lehrer argues against Peter Klein and Risto Hilpinen's proposal according to which a justification depends on a false statement iff the person holding the justification would not be justified anymore if she knew the false statement to be false. Lehrer provides examples where knowing some statement to be false is misleading[2].

[2]We develop such an example in Section 5.2.3.

In fact, Lehrer only requires that the agent has *some* justification that does not depend on any false statement or is not defeated by any false statement in his definition of knowledge. Here the need and the interest of the ultra-justification game is fully revealed. Remember that this type of game involves using the ultra-system of the agent that is, a system retaining only what is true in the agent's evaluation system, an evaluation system free of error. Indeed the Ultra-critic can ask the Claimant to eliminate acceptances, preferences and reasonings that do not belong to the ultra-system of the Claimant. The ultra-justification game allows to delete all false statements that could be part of the justification of an agent who can then only use true statements. Then the Claimant wins the game if she can answer or neutralize all the objections of the Ultra-critic proving she is justified to accept p in a way that is undefeated by any false statement.

If an agent has at least one justification that does not depend on false statement, she will win the ultra-justification game because once she will delete all the false statements belonging to her evaluation system, the Ultra-critic will not be able to defeat the remaining statements and so the corresponding justification[3].

5.2 An original game semantics for defeasible knowledge

In this section we formalize Lehrer's concept of "(in)defeasible knowledge" in terms of a game semantics that we design for this purpose. First, we analyse the notions of belief and knowledge we define in Section 4.5.2 and 4.2.6 from the point of view of Lehrer's theory of knowledge. Next, we prove Lehrer's notion of (in)defeasible knowledge (see Definition 5.1.1) to be equivalent to the formal concept of defeasible knowledge.

5.2.1 Knowledge, belief and justification

In Definition 4.2.15, we define the notion of justification: a justification for Q is an argument F such that all arguments at least as strong as F support Q $(\forall F'(F \preceq F' \Rightarrow \bigcap F' \subseteq Q))$. In Lehrer's personal justification game, this means that the argument F cannot be defeated by a stronger argument since all stronger arguments support Q. This notion of justification introduced in Chapter 4 corresponds exactly to the notion of personal justification of Lehrer.

In Definition 4.5.2, we define the notion of belief in total justification models: an agent believes Q iff there exists a justification F for Q $(\exists F \forall F' \succeq F(\bigcap F' \subseteq Q))$. An agent believes Q iff all the arguments stronger than F support Q. In Lehrer's terminology, this means that the agent is personally justified in accepting Q.

[3] We provide an example to illustrate this in Section 5.2.3.

The notion of justified true belief is captured in our formal system as follows: $p \wedge \exists^{ev} just\, p$. We would like to emphasize here the difference between this definition of justified true belief with our definition of (defeasible) knowledge. We define (defeasible) knowledge in Definition 4.5.2: in total justification models, an agent defeasibly knows Q at s iff there exists a sound justification F for Q at s $(\exists F(s \in \bigcap F \wedge \forall F' \geq F(\bigcap F' \subseteq Q)))$. In our definition of defeasible knowledge, it is not just the belief in Q that has to be sound but above all the evidence the agent has for Q. An agent defeasibly knows Q iff the evidence she has for Q is sound and her evidence supports Q. In Lehrer's ultra-justification game, this means that the argument F cannot be defeated by a stronger argument since all stronger arguments support Q nor by soundness. This notion of justification corresponds exactly to the notion of undefeated justification of Lehrer.

5.2.2 Ultra-justification game

Our setting will start from a given justification model $\mathcal{M}$ which fixes the agent's justifications, irrevocable knowledge, beliefs, conditional beliefs, strong beliefs and defeasible knowledge. We take this to be the basis of the agent's evaluation system. From now we only focus on the class of justification models with a total pre-order on bodies of evidence. We distinguish between two kinds of justification models inside this class: the general kind of justification models where the evidence sets are not nested but can be mutually inconsistent or only partially overlapping and the AGM kind where all the evidence sets are nested.

Assumptions We equate Lehrer's notion of "acceptance" with our notion of "belief" and assume our agent to be logically omniscient. Another assumption we make is that our agent holds only consistent beliefs. These simplifying assumptions render the formalization less complicated and prove sufficient to give a first formal analysis of Lehrer's justification games. However we are aware of the fact that these restrictions will have to be lifted in future work if we want to have a fully accurate formalization of Lehrer's account including a formal analysis going beyond the setting provided here.

5.2.1. Definition. Given a total justification model[4] $\mathcal{M}_0 = (S_0, E_0, \leq_0, \|\cdot\|_0, s_0)$ and a claim $Q \subseteq S_0$, we define the *ultra-justification game* $G(\mathcal{M}_0, Q)$.

The ultra-justification game is a two players game where the players are called Claimant and Ultra-critic. The justification model of the Claimant is supposed to be known by the Ultra-critic. That means the Ultra-critic knows the epistemic and doxastic attitudes of the Claimant as well as his justification such that the

[4] In the rest of the chapter we assume that all justification models and plausibility models are total (i.e. connected) models and we will only explicitly mention the word "total" further on in case confusion is possible.

Ultra-critic knows which belief (and conditional belief, strong belief) is false and which justification is unsound.

Every move for the Claimant (or Believer) is bound by the precondition that the information he conveys to his opponent about himself holding certain justifications, beliefs, conditional beliefs or about the strength of his beliefs, has to be truthful. In other words, the Believer cannot make claims that go against the information he accepts in his own evaluation system, even if what he believes might actually be false in reality.

The pre-condition for any move of the Ultra-critic is that all the information he conveys has to be true in the actual world, i.e. the Ultra-critic cannot lie. This is why we use public announcement operators ! in the formalization of the information that the Ultra-critic conveys.

A play (or run) is a sequence of moves of the players where the set of legal *moves* of each player is defined below.

The game is played on *positions* that is, on pairs $P = (\mathcal{M}, F)$ such that $\mathcal{M} = (S, E, \leq, \|\cdot\|, s)$ is a justification model such that $s = s_0 (\in S)$ and $F \in \mathcal{E}_{\mathcal{M}}$ is an argument for that model $\mathcal{M}$ (i.e. F is a body of evidence in that model $\mathcal{M}$).

The initial position is $P_0 = (\mathcal{M}_0, F_0)$ where $\mathcal{M}_0$ is the initial total justification model and $F_0 = \varnothing$.

The game is played in rounds composed of a move made by the Ultra-critic, followed by a move made by the Claimant. We call the moves of the Ultra-critic "challenges" and the moves of the Claimant "defences".

Consider a position $P_{n-1} = (\mathcal{M}_{n-1}, F_{n-1})$ where $n \geq 1$.

1. First, the Ultra-critic makes a move by:

- **a.** either challenging the current argument F_{n-1} as *unsound*, i.e. announcing that $s_0 \notin \bigcap F_{n-1}$ which induces an update $!(\neg \bigcap F_{n-1})$ of the current justification model $\mathcal{M}_{n-1}$ or
- **b.** challenging the current argument F_{n-1} as *unconvincing* (it is not a justification for Q), by finding an *objection*, i.e. an argument $F' \in \mathcal{E}_{\mathcal{M}_{n-1}}$ such that $F_{n-1} \leq F'$ and $\mathcal{M}_{n-1}, s_0 \vDash \neg K \neg (\bigcap F' \wedge \neg Q)$, i.e. $\bigcap F' \nsubseteq Q \cap S_{n-1}$.

After this, the current justification model is updated to a new justification model $\mathcal{M}_n$ given by:

- $\mathcal{M}_n := \mathcal{M}_{n-1} | (\neg \bigcap F_{n-1})$ if the Ultra-critic made a move of type $(a.)$;
- $\mathcal{M}_n := \mathcal{M}_{n-1}$ if the Ultra-critic made a move of type $(b.)$.

2. Next, the Claimant correspondingly defends himself by:

- **a'.** answering a challenge of type ($a.$) with a new argument $F_n \in \mathcal{E}_{\mathcal{M}_n}$, ending at a new position $P_n = (\mathcal{M}_n, F_n)$; or
- **b'.** answering a challenge of type ($b.$) with a new argument $F_n \in \mathcal{E}_{\mathcal{M}_n}$ such that $F' < F_n$, ending at a new position $P_n = (\mathcal{M}_n, F_n)$.

If at any round, a player cannot make a move, then he *loses* and the other player *wins*. If the Ultra-critic cannot make a move, i.e. he cannot challenge a given argument, this argument is said to be undefeated. If the Claimant cannot make a move, i.e. he cannot defend a given argument, this argument is said to be defeated. So the Claimant wins the ultra-justification game $G(\mathcal{M}_0, Q)$ iff he offered at least one justification for Q that is left undefeated (his original belief in Q is undefeated).

5.2.2. Proposition. *Every play ends in finitely many steps with one of the two players winning.*

5.2.3. Proof. Every move of the Claimant either shrinks the total justification model $\mathcal{M}$ or goes to a strictly more convincing argument F_n. We know that every total justification model is finite. Since S_0 is finite, there are only finitely many updates that shrink the justification model. So there exists some number n, such that starting from the n-th round, the justification model stays the same forever, i.e. $\forall m > n, \mathcal{M}_n := \mathcal{M}_m$.

From round n onwards, the Claimant can only make moves of type ($b'.$). So he can only defend himself providing a new argument such that this argument is strictly more convincing than the argument of the Ultra-critic, which means that at each round, the arguments go stronger: $\forall m > n, F_n < \cdots < F_m$ for every argument $F_n \in \mathcal{E}_{\mathcal{M}_n}$. Since $\mathcal{M}_n$ is finite, $\mathcal{E}_{\mathcal{M}_n}$ is also finite and there are only finitely many available arguments. Hence, there is no infinite ascending chain of arguments, i.e. at some round $m > n$, the Claimant last argument F_m is either defeated (the Ultra-critic wins) or is undeafeated (i.e. the Ultra-critic cannot challenge it, that is, he cannot make a move) and the Claimant wins. □

5.2.4. Corollary. *The game is determined: there exits a winning strategy for one of the players.*

5.2.5. Theorem. *The Claimant defeasibly knows Q iff he has a winning strategy in the ultra-justification game $G(\mathcal{M}_0, Q)$. Else, the Ultra-critic has a winning strategy in the ultra-justification game $G(\mathcal{M}_0, Q)$.*

5.2.6. Proof. – Assume the Claimant defeasibly knows Q. We are given a total justification model $\mathcal{M}_0$ in which $K_D Q$ is true at s_0. We have to show that there exists a winning strategy for the Claimant in $G(\mathcal{M}_0, Q)$.

We denote by L_n the (finite) set of available moves for the Claimant at round n at step (2.), i.e. after the Ultra-critic made his n-th move and the justification model has been updated to $\mathcal{M}_n$. Since $K_D Q$ is true at s_0 and by Proposition 4.5.11, we know that there exists a sound justification $F \in \mathcal{E}_0$ for Q at s_0. All we need to show is the following:

Claim: for every n, if the step (2.) is reached, i.e the Ultra-critic made his n-th move and the justification model has been updated to $\mathcal{M}_n$, then $L_n \neq \varnothing$. More precisely, we will show that we will always have

$$F|\mathcal{M}_n \in L_n$$

The desired conclusion will follow from this claim. Since for every round n, either the Ultra-critic cannot make his move (hence, he loses) or his challenge can be answered by the Claimant (by choosing any $F|\mathcal{M}_n \in L_n$). Hence, the Claimant can never lose. So by Corollary 5.2.4, he will win.

Proof of the claim: by induction on n. At round $n = 0$, we just have $L_0 := \mathcal{E}_0 = \{G \in \mathcal{E}_0 \mid G_0 = \varnothing \leq G\}$. So we have $F \in L_0$ since $\varnothing \leq F$.

At any later round n, if the Ultra-critic cannot move, the Claimant wins and we are done.

Otherwise if the Ultra-critic made a move of type (a.), announcing $!(\neg \bigcap F_{n-1})$ then we know that we have $\mathcal{M}_n := \mathcal{M}_{n-1}|(\neg \bigcap F_{n-1})$. We want to show that $F|\mathcal{M}_n \in L_n$, i.e. $F|\mathcal{M}_n \in \mathcal{E}_{\mathcal{M}_n}$. By induction hypothesis, we know that $F|\mathcal{M}_{n-1} \in L_{n-1} \subseteq \mathcal{E}_{\mathcal{M}_{n-1}}$. We know that $F|\mathcal{M}_n := (F|\mathcal{M}_{n-1})|\mathcal{M}_n$. By definition of $\mathcal{E}_{\mathcal{M}_n}$, $\mathcal{E}_{\mathcal{M}_n} = \{G|\mathcal{M}_n \mid G \in \mathcal{E}_{\mathcal{M}_{n-1}}\}$. Then $F|\mathcal{M}_n := (F|\mathcal{M}_{n-1})|\mathcal{M}_n \in \mathcal{E}_{\mathcal{M}_n}$.

If the Ultra-critic made a move of type (b.), he found an argument $F' \in \mathcal{E}_{n-1}$ such that $F_{n-1} \leq F'$ and $\bigcap F' \nsubseteq Q \cap S_{n-1}$, i.e. $\bigcap F' \nsubseteq Q$. In this case, we know that $\mathcal{M}_n := \mathcal{M}_{n-1}$. We want to show that $F|\mathcal{M}_n \in L_n$. By induction hypothesis, we know that $F|\mathcal{M}_{n-1} \in L_{n-1}$. We have to prove that $F' <_{\mathcal{M}_n} F|\mathcal{M}_n$. Suppose $F|\mathcal{M}_n \leq_{\mathcal{M}_n} F'$. By definition of $\leq_{\mathcal{M}_n}$, we have $\{e \in E_0 \mid e \cap S_n \in F|\mathcal{M}_n\} \leq_{\mathcal{M}_0} \{e \in E_0 \mid e \cap S_n \in F'\}$. But F is sound, i.e. $s \in e$ for all $e \in F$, and so $s \in e \cap S_n \neq \varnothing$ for all $e \in F$. Hence, $\forall e \in E_0 (e \in F \Rightarrow e \cap S_n \in F|\mathcal{M}_n)$. So $F \subseteq \{e \in E_0 \mid e \cap S_n \in F|\mathcal{M}_n\}$, and hence, $F \leq_{\mathcal{M}_0} \{e \in E_0 \mid e \cap S_n \in F|\mathcal{M}_n\} \leq_{\mathcal{M}_0} \{e \in E_0 \mid e \cap S_n \in F'\}$. Since F is a justification for Q in $\mathcal{M}_0$, we must have $\bigcap\{e \in E_0 \mid e \cap S_n \in F'\} \subseteq Q$. But F' is an argument in $\mathcal{M}_{n-1} := \mathcal{M}_n$. Hence, we have $e \cap S_n = e$ for all $e \in F'$, and so $\bigcap F' = \bigcap\{e \in E_0 \mid e \in F'\} = \bigcap\{e \in E_0 \mid e \cap S_n \in F'\} \subseteq Q$, which contradicts the fact that F' was chosen as an objection by the Ultra-critic (with $\bigcap F' \nsubseteq Q$).

– For the other direction, assume the Claimant does not defeasibly know Q. We are given total justification model $\mathcal{M}_0$ in which $K_D Q$ is false at s_0. We want to show that the Ultra-critic has a winning strategy (and hence, the Claimant has not a winning strategy). By Proposition 4.5.11, there does not exist a sound justification $F \in \mathcal{E}_0$ for Q at s_0.

  At every round n, the previous argument F_{n-1} of the Claimant will be defeated by the Ultra-critic. The Ultra-critic will challenge either by showing that F_{n-1} is unsound at s_0 or that F_{n-1} is unconvincing (it is not a justification for Q), i.e. by providing an objection F' with $F_{n-1} \leq F'$ such that $\bigcap F' \nsubseteq Q$. The Claimant can never win. Hence, by Corollary 5.2.4, he will lose.

□

We apply our setting to the following examples. In all the applications of our game semantics, we will use refined justification models as defined in Section 4.2.5, which allow us to distinguish between soft arguments (that weakly support a conclusion, given some implicit biases) and "stronger" arguments (that make the biases explicit).

5.2.3 Applications

The first example is inspired by [47].

Example 1: Zebra

Imagine an agent who is dreaming that she is at the Amsterdam Zoo (Artis) looking at a Zebra. The atomic propositions in this example are *zebra* (there is a zebra), *dream* (the agent is dreaming), *see* (the agent sees a zebra) and *Zoo* (the agent is at the Amsterdam Zoo). We represent the agent's evidences via the refined justification model $\mathcal{M}_0 = (S_0, E_0, \leq_0, \|\cdot\|_0, s_0)$ described in Figure 5.1.

Note that in this example, we do not consider "seeing" as a factual attitude: what the agent sees is not necessarily true. However, we do consider "seeing" as being fully introspective, that is, if an agent sees something, she irrevocably knows that she sees it and if she does not see something, she irrevocably knows she does not see it. So the agent irrevocably knows that she sees a zebra by introspection. Moreover, the agent irrevocably knows that if she sees a zebra, then either she is at the Zoo or she is dreaming (we assume an agent living in Amsterdam, far away from any savannah). In the same way, she irrevocably knows that if she is at the Amsterdam Zoo, then there is a zebra (we assume she already went to the Zoo where there is indeed a zebra).

In accordance with the knowledge of the agent, her epistemic state consists of four worlds $S_0 = \{s, t, u, v\}$. The valuation of the atomic propositions is given as follows: *zebra* is true at s, u, v, *dream* is true at s, t, u, *see* is true at s, t, u, v and

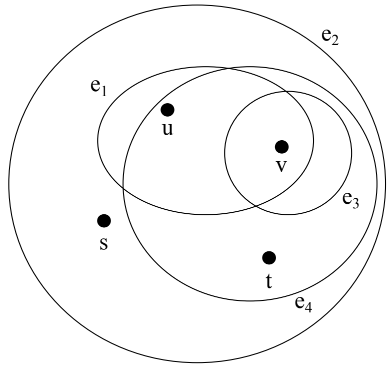

Figure 5.1: Initial refined justification model

Zoo is true at u, v. So at the state $s = s_0$, there is a zebra, the agent sees a zebra, the agent is not at the Zoo and the agent is dreaming. At the state t, there is not a zebra, the agent sees a zebra, the agent is not at the Zoo and the agent is dreaming. At the state u, there is a zebra, the agent sees a zebra, the agent is at the Zoo and the agent is dreaming. At the state v, there is a zebra, the agent sees a zebra, the agent is at the Zoo and the agent is not dreaming. Formally we have:

- $\| zebra \|_0 = \{s, u, v\}$,
- $\| dream \|_0 = \{s, t, u\}$,
- $\| see \|_0 = \{s, t, u, v\}$,
- $\| Zoo \|_0 = \{u, v\}$.

We remind the reader that, in a refined justification model, not all evidence sets represent genuine evidence. Some evidences sets are biases giving the agent's default beliefs. In the refined justification model $\mathcal{M}_0$, there are four evidence sets:

$$E_0 = \{Zoo, see, \neg dream, zebra \to Zoo\}$$

with:

- $Zoo = \{u, v\} = e_1$,
- $see = \{s, t, u, v\} = e_2$,
- $\neg dream = \{v\} = e_3$,
- $zebra \to Zoo = \{t, u, v\} = e_4$.

The evidence sets e_1 and e_2 represent genuine evidence (through not necessarily truthful). The evidence set e_2 represents the piece of evidence the agent has, based on her perception: her eyes. The evidence set e_1 represents the piece of evidence the agent has, based on her memory: she remembers coming to the Zoo. The evidence sets e_3 and e_4 represent the biases of the agent. By default, the agent assumes that she is not dreaming, she has no evidence to the contrary so she prefers to believe she is awake (as usually people do). She also assumes that if there is a zebra, then she is at the Amsterdam Zoo (since there is no savannah near Amsterdam).

Note that the evidence set e_2 has the property that $e_2 = \mathit{see} = S_0$.

We have $\mathcal{E}_0 = \{\varnothing, \{e_1\}, \{e_2\}, \{e_3\}, \{e_4\}, \{e_1, e_2\}, \{e_1, e_3\}, \{e_1, e_4\}, \{e_2, e_3\}, \{e_2, e_4\}, \{e_3, e_4\}, \{e_1, e_2, e_3\}, \{e_1, e_2, e_4\}, \{e_1, e_3, e_4\}, \{e_2, e_3, e_4\}, \{e_1, e_2, e_3, e_4\}\}$.

The pre-order $\leq_0$ on $\mathcal{E}_0$ is given by inclusion $\subseteq$, i.e. $F \leq_0 G$ iff $F \subseteq G$ (so in fact we get an evidence model).

From refined justification model to plausibility model We can easily turn this refined justification model into a plausibility model:

1. $s \leq_{E_0} t$ iff $E_t \leq_0 E_s$ iff $E_t \subseteq E_s$
2. $E_s := \{e_2\}$, $E_t := \{e_2, e_4\}$, $E_u := \{e_1, e_2, e_4\}$ and $E_v := \{e_1, e_2, e_3, e_4\}$
3. $\{e_1, e_2, e_4\} \leq_0 \{e_1, e_2, e_3, e_4\}$ so $E_u \leq_0 E_v$
4. $\{e_2, e_4\} \leq_0 \{e_1, e_2, e_4\}$ so $E_t \leq_0 E_u$
5. $\{e_2\} \leq_0 \{e_2, e_4\}$ so $E_s \leq_0 E_t$

So we have $E_s \leq_0 E_t \leq_0 E_u \leq_0 E_v$ that is, $v \leq_{E_0} u \leq_{E_0} t \leq_{E_0} s$.

Plausibility model We represent the agent's beliefs and knowledge via the plausibility model described in Figure 5.2 consisting of four possible states (s, t, u, v) where the double circled state indicates the real world and the arrows represent the plausibility relation on states (we skip the reflexive and transitive arrows).

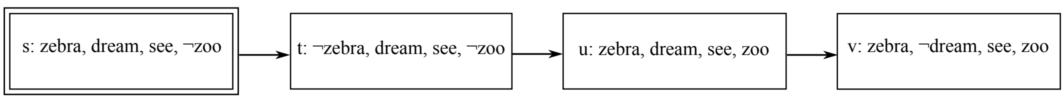

Figure 5.2: Initial plausibility model

One can easily see that our justification model is sphere-based.

The informal dialogue The dialogue starts with our agent claiming to know that she sees a zebra:

Claimant: There is a zebra here.

$$B(zebra)$$

Ultra-critic: Why do you think so? (Justify!)

$$\neg K\neg(\neg zebra)$$

Claimant: I believe there is a zebra because I see a zebra.

$$\{see\} \in \mathcal{E}_{\mathcal{M}_1} \text{ and } B(see) \text{ and } B^{see}(zebra)$$

Ultra-critic: Maybe you are sleeping and dreaming that you see a zebra.
(Your evidence is consistent with the negation of "zebra"! You need to provide further justification!)

$$\neg K\neg(see \wedge dream \wedge \neg zebra)$$

Claimant: It is more reasonable for me to accept that there is a zebra because I see the zebra than to accept that I am dreaming a zebra!

$$\{see, \neg dream\} \in \mathcal{E}_{\mathcal{M}_2} \text{ and } B(see \wedge \neg dream) \text{ and } B^{see \wedge \neg dream}(zebra)$$

Ultra-critic: You are dreaming! You are asleep! You are only seeing the zebra in your dreams!

$$!dream$$

Update The announcement of "dream" is taken as a public announcement, which formally will change the refined justification model and the plausibility model as described respectively in Figures 5.3 and 5.4:

So $\mathcal{M}_3 := \mathcal{M}_2|(dream)$ where $E_{\mathcal{M}_3} = \{zoo, see, dream, zebra \to zoo\}$ with:

- $Zoo = \{u\} = e_1$,
- $see = \{s, t, u\} = e_2$,
- $zebra \to Zoo = \{t, u\} = e_4$,
- $dream = \{s, t, u\} = e_5$.

Note that after the update, *dream* is true in all the remaining states ($S_2|(dream) = dream$), so *dream* is an evidence set.

We have $\mathcal{E}_{\mathcal{M}_3} = \{\varnothing, \{e_1\}, \{e_2\}, \{e_4\}, \{e_5\}, \{e_1, e_2\}, \{e_1, e_4\}, \{e_1, e_5\}, \{e_2, e_4\}, \{e_2, e_5\}, \{e_4, e_5\}, \{e_1, e_2, e_4\}, \{e_1, e_2, e_5\}, \{e_1, e_4, e_5\}, \{e_2, e_4, e_5\}, \{e_1, e_2, e_4, e_5\}\}$.

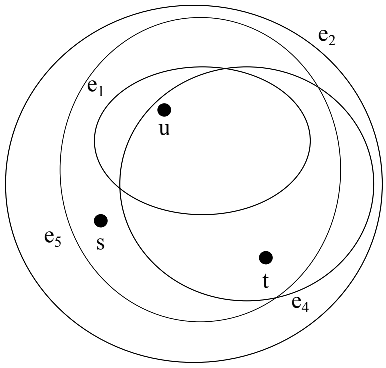

Figure 5.3: Refined justification model updated with $!dream$

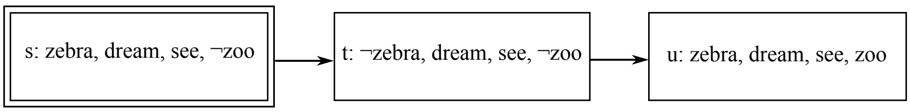

Figure 5.4: Plausibility model updated with $!dream$

The dialogue continued Then the dialogue continues and our agent claims:
Claimant: I still believe there is a zebra here, coincidental with my dreaming of it. I distinctly remember coming to the Zoo. Maybe I just fell asleep at the Zoo? This would also explain why I am seeing a zebra.

$$Zoo \in \mathcal{E}_{\mathcal{M}_3} \text{ and } B(zoo) \wedge B^{Zoo\wedge dream\wedge see}(zebra)$$

Ultra-critic: You are not at the zoo. You are asleep in your bed, dreaming of zebras.

$$!\neg Zoo$$

Update The announcement of "$\neg Zoo$" is taken as a public announcement, which formally will change the refined justification model and the plausibility model as described respectively in Figures 5.5 and 5.6:

So $\mathcal{M}_4 := \mathcal{M}_3|(\neg Zoo)$ where $E_{\mathcal{M}_4} = \{\neg zoo, see, dream, zebra \to zoo\}$ with:

- $\neg Zoo = \{s, t\} = e_6$,
- $see = \{s, t\} = e_2$,
- $dream = \{s, t\} = e_5$,
- $zebra \to Zoo = \{t\} = e_4$

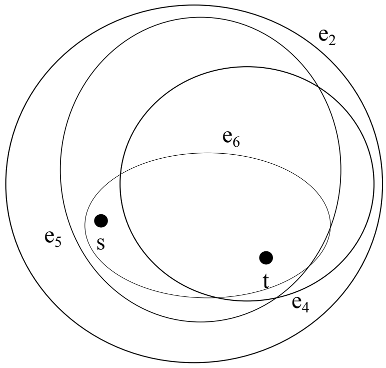

Figure 5.5: Refined justification model updated with $!\neg Zoo$

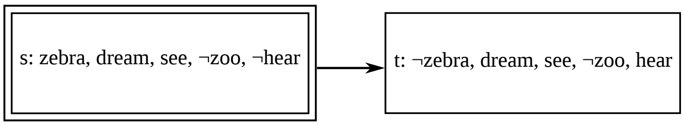

Figure 5.6: Plausibility model updated with $!\neg Zoo$

Note again that after the update, $\neg Zoo$ is true in all the remaining states ($S_3|(\neg Zoo) = \neg Zoo$), so $\neg Zoo$ is an evidence set.

We have $\mathcal{E}_{\mathcal{M}_4} = \{\varnothing, \{e_2\}, \{e_5\}, \{e_4\}, \{e_6\}, \{e_2, e_5\}, \{e_2, e_4\}, \{e_2, e_6\}, \{e_5, e_4\}, \{e_5, e_6\}, \{e_4, e_6\} \{e_2, e_5, e_4\}, \{e_2, e_5, e_6\}, \{e_2, e_4, e_6\}, \{e_5, e_4, e_6\}, \{e_2, e_5, e_4, e_6\}\}$.

The dialogue ended Then the dialogue ends with our agent claiming:

Claimant: Given that I am asleep in my bed, I have no justification left to believe there is a zebra here in the bedroom. So I give up: I no longer believe it!

Conclusion of the dialogue The agent loses because she cannot provide further argument for *zebra*. She does not even believe *zebra* anymore. She didn't defeasibly "know" that there was a zebra. In this case our agent does not defeasibly know *zebra* since she based her belief on false evidence. Sadly enough, her initial justified belief was in fact true: however implausible this might seem to her, there is a zebra in her bedroom. The agent did not know: she only had a true justified belief.

The formal game Formally, we model the informal dialogue as a play in our ultra-justification game $G(\mathcal{M}_0, zebra)$ as follows.

Claimant: There is a zebra here.

$$B(zebra)$$

This means that at round 0, $F_0 = \varnothing$.

Ultra-critic: Why do you think so? (Justify!)

$$\neg K \neg(\neg zebra)$$

This means that at round 1, the Ultra-critic chooses $F'_0 = F_0 = \varnothing$ such that F'_0 does not support $zebra$: $\bigcap F'_0 = \bigcap \varnothing = S_0 \nsubseteq zebra$ because of the state t ($t \vDash \neg zebra$).

Claimant: I believe there is a zebra because I see a zebra.

$$\{see\} \in \mathcal{E}_{\mathcal{M}_1} \text{ and } B(see) \text{ and } B^{see}(zebra)$$

This means that the Claimant chooses a new argument $F_1 = \{see\}$ in $\mathcal{M}_1 := \mathcal{M}_0$, which is a soft argument in the sense of Definition 4.2.14 that weakly supports $zebra$ conditional on the default $\neg dream$.

Ultra-critic: Maybe you are sleeping and dreaming that you see a zebra.
(Your evidence is consistent with the negation of "zebra"! You need to provide further justification!)

$$\neg K \neg(see \wedge dream \wedge \neg zebra)$$

This means that at round 2, the Ultra-critic chooses $F'_1 = F_1 = \{see\}$ because F'_1 does not support $zebra$: $\bigcap F'_1 = \bigcap \{see\} \nsubseteq zebra$ because of the state t ($t \vDash see \wedge dream \wedge \neg zebra$).

Claimant: It is more reasonable for me to accept that there is a zebra because I see the zebra than to accept that I am dreaming a zebra!

$$\{see, \neg dream\} \in \mathcal{E}_{\mathcal{M}_2} \text{ and } B(see \wedge \neg dream) \text{ and } B^{see \wedge \neg dream}(zebra)$$

This means that the Claimant makes explicit his bias $\neg dream$ by adding it to his argument, obtaining a new argument $F_2 = \{see, \neg dream\}$ in $\mathcal{M}_2 := \mathcal{M}_1$, which is an argument that supports $zebra$.

Ultra-critic: You are dreaming! You are asleep! You are only seeing the zebra in your dreams!

$$!dream$$

This means that at round 3, the Ultra-critic challenges F_2 as unsound and announces that $s \notin \bigcap F_2$ which induces an update $!(\neg \bigcap F_2)$ of the refined justification model and the plausibility model.

Update The announcement of "dream" is taken as a public announcement, which formally will change the refined justification model and the plausibility model as described respectively in Figures 5.3 and 5.4:

So $\mathcal{M}_3 := \mathcal{M}_2|(\neg \bigcap F_2)$ where $E_{\mathcal{M}_3} = \{zoo, see, dream, zebra \to zoo\}$ with:

- $Zoo = \{u\} = e_1$,
- $see = \{s, t, u\} = e_2$,
- $zebra \to Zoo = \{t, u\} = e_4$,
- $dream = \{s, t, u\} = e_5$.

Note that after the update, *dream* is true in all the remaining states ($S_2|(\neg \bigcap F_2) = dream$), so *dream* is an evidence set.

We have $\mathcal{E}_{\mathcal{M}_3} = \{\varnothing, \{e_1\}, \{e_2\}, \{e_4\}, \{e_5\}, \{e_1, e_2\}, \{e_1, e_4\}, \{e_1, e_5\}, \{e_2, e_4\}, \{e_2, e_5\}, \{e_4, e_5\}, \{e_1, e_2, e_4\}, \{e_1, e_2, e_5\}, \{e_1, e_4, e_5\}, \{e_2, e_4, e_5\}, \{e_1, e_2, e_4, e_5\}\}$.

The game continued Then the game continues and our agent claims:

Claimant: I still believe there is a zebra here, coincidental with my dreaming of it. I distinctly remember coming to the Zoo. Maybe I just fell asleep at the Zoo? This would also explain why I am seeing a zebra.

$$Zoo \in \mathcal{E}_{\mathcal{M}_3} \text{ and } B(zoo) \wedge B^{Zoo \wedge dream \wedge see}(zebra)$$

This means that the Claimant chooses a new argument $F_3 = \{see, dream, Zoo\}$ in $\mathcal{M}_3$, which is an argument that supports *zebra*.

Ultra-critic: You are not at the zoo. You are asleep in your bed, dreaming of zebras.

$$!\neg Zoo$$

This means that at round 4, the Ultra-critic challenges F_3 as unsound and announces that $s \notin \bigcap F_3$ which induces an update $!(\neg \bigcap F_3)$ of the refined justification model and the plausibility model.

Update The announcement of "$\neg Zoo$" is taken as a public announcement, which formally will change the refined justification model and the plausibility model as described respectively in Figures 5.5 and 5.6:

So $\mathcal{M}_4 := \mathcal{M}_3|(\neg \bigcap F_3)$ where $E_{\mathcal{M}_4} = \{\neg zoo, see, dream, zebra \to zoo\}$ with:

- $\neg Zoo = \{s, t\} = e_6$,
- $see = \{s, t\} = e_2$,

– $dream = \{s, t\} = e_5$,

– $zebra \to Zoo = \{t\} = e_4$

Note again that after the update, $\neg Zoo$ is true in all the remaining states ($S_3|(\neg \bigcap F_3) = \neg Zoo$), so $\neg Zoo$ is an evidence set.

We have $\mathcal{E}_{\mathcal{M}_4} = \{\varnothing, \{e_2\}, \{e_5\}, \{e_4\}, \{e_6\}, \{e_2, e_5\}, \{e_2, e_4\}, \{e_2, e_6\}, \{e_5, e_4\}, \{e_5, e_6\}, \{e_4, e_6\} \{e_2, e_5, e_4\}, \{e_2, e_5, e_6\}, \{e_2, e_4, e_6\}, \{e_5, e_4, e_6\}, \{e_2, e_5, e_4, e_6\}\}$.

The game ended Then the game ends with our agent claiming:

Claimant: Given that I am asleep in my bed, I have no justification left to believe there is a zebra here in the bedroom. So I give up: I no longer believe it!

Conclusion of the game The agent loses this round of the ultra-justification game because she cannot provide further argument for *zebra*. She does not even believe *zebra* anymore. Then the Claimant loses the game: she didn't defeasibly "know" that there was a zebra. In this case our agent does not defeasibly know *zebra* since she based her belief on false evidence. Sadly enough, her initial justified belief was in fact true: however implausible this might seem to her, there is a zebra in her bedroom. The agent did not know: she only had a true justified belief.

Example 2: Ferrari

The second example illustrates the meaning of Lehrer's condition (4.) in his definition of knowledge and underlines the interest of the ultra-justification game. In particular it shows that Lehrer only requires that the agent has at least one justification that does not depend on any false statement to win the ultra-justification game.

Suppose an agent S is in a room with Mr. Nogot and Mr. Knewit. Mr. Nogot does not own a Ferrari contrary to Mr. Knewit. However the agent S is justified in accepting that Mr. Nogot owns a Ferrari because S saw Mr. Nogot drove a Ferrari and Mr. Nogot showed S the papers stating he owns a Ferrari. Then suppose someone asks S if she knows whether anyone in the room owns a Ferrari, S replies claiming she knows that *at least one person in the room owns a Ferrari* (P). It seems that though S has a justified true belief that P, she does not know it. However, suppose S also is justified in accepting that Mr. Knewit owns a Ferrari because S sold Mr. Knewit her Ferrari. Though part of the justification of S (Mr. Nogot owns a Ferrari) is a false statement/belief, she also has justification that does not depend on this false statement/belief.

The atomic propositions in this example are P (at least one person in the room owns a Ferrari, i.e. $Nogot \vee Knewit$)[5], *reliable* (Mr. Nogot is reliable), *buy* (Mr. Knewit bought the Ferrari of the agent), *Mr. Nogot* (Mr. Nogot owns a Ferrari), *Mr. Knewit* (Mr. Knewit owns a Ferrari). We represent the agent's evidence via the refined justification model $\mathcal{M}_0 = (S_0, E_0, \preceq_0, \|\cdot\|_0, s_0)$ described in Figure 5.7.

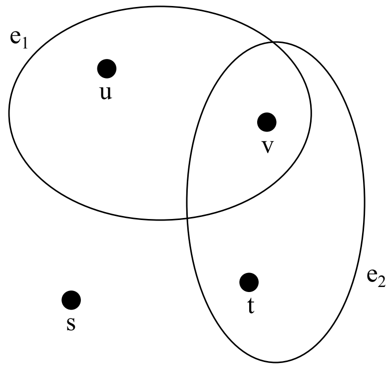

Figure 5.7: Initial refined justification model

For simplicity, we assume the agent irrevocably knows that if Mr. Nogot is reliable then Mr. Nogot does own a Ferrari while if Mr. Nogot is not reliable, (i.e. Mr. Nogot was lying about the papers stating he owns a Ferrari), he does not actually own a Ferrari[6]. We also assume that the agent irrevocably knows that if Mr. Knewit bought her Ferrari then Mr. Knewit does own a Ferrari (since the agent knows Mr. Knewit really wants to own a Ferrari, not to sold one) while if Mr. Knewit did not buy the Ferrari of the agent, then Mr. Knewit does not own a Ferrari (the agent knows that nobody else could have sold one Ferrari to Mr. Knewit).

In accordance with the knowledge of the agent, her epistemic state consists of four worlds $S_0 = s, t, u, v$. The valuation of the atomic propositions is given as follows: P is true at t, u, v, *Nogot* is true at u, v, *Knewit* is true at v, t, *reliable* is true at u, v and *buy* is true at t, v. So at the state s, Mr. Nogot is not reliable, Mr. Nogot does not own a Ferrari, Mr. Knewit did not buy the Ferrari, Mr. Knewit does not own a Ferrari, nobody in the room owns a Ferrari. At the state $t = s_0$, Mr. Nogot is not reliable, Mr. Nogot does not own a Ferrari, Mr. Knewit bought the Ferrari, Mr. Knewit owns a Ferrari, at least one person in the room owns a Ferrari. At the state u, Mr. Nogot is reliable, Mr. Nogot owns a Ferrari, Mr. Knewit did not buy the Ferrari, Mr. Knewit does not own a Ferrari, at least one person in the room owns a Ferrari. At the state v, Mr. Nogot is reliable,

[5] We assume that our agent knows that she does not own a Ferrari at that moment.

[6] It is not unusual to stop trusting and believing in people when one realize they are liars while one continue to trust them as long as one has evidence that they are telling the truth.

Mr. Nogot owns a Ferrari, Mr. Knewit bought the Ferrari, Mr. Knewit owns a Ferrari, at least one person in the room owns a Ferrari. Formally we have:

- $\| P \|_0 = \{t, u, v\}$,
- $\| Nogot \|_0 = \{u, v\}$,
- $\| Knewit \|_0 = \{t, v\}$,
- $\| reliable \|_0 = \{u, v\}$,
- $\| buy \|_0 = \{t, v\}$,

In the refined justification model $\mathcal{M}_0$, there are two evidence sets:

$$E_0 = \{reliable, buy\}$$

with:

- $reliable = \{u, v\} = e_1$,
- $buy = \{t, v\} = e_2$.

The evidence set e_2 represents genuine evidence, i.e. the piece of evidence the agent has, based on her memory: she remembers selling a Ferrari to Mr. Knewit. The evidence set e_1 represents a bias. By default, she assumes Mr. Nogot is reliable since she has no evidence to the contrary (and we assume the agent prefers to trust people are not liars).

We have the family of bodies of evidence $\mathcal{E}_0 = \{\varnothing, \{e_1\}, \{e_2\}, \{e_1, e_2\}\}$.

As pre-order $\preceq_0$, we take the cardinality order, i.e. $F \preceq_0 G$ iff $\| F \| \leq \| G \|$. So the refined justification model $\mathcal{M}_0$ is a counting model.

From refined justification model to plausibility model We can easily turn this refined justification model into a plausibility model:

1. $s \leq_{E_0} t$ iff $E_t \preceq_0 E_s$ iff $\| E_t \| \leq \| E_s \|$.
2. $E_t := \{e_2\}$, $E_u := \{e_1\}$ and $E_v := \{e_1, e_2\}$
3. $\{e_1\} \preceq_0 \{e_1, e_2\}$ so $E_u \preceq_0 E_v$
4. $\{e_2\} \preceq_0 \{e_1, e_2\}$ so $E_t \preceq_0 E_v$
5. $\{e_1\} \equiv_0 \{e_2\}$ so $E_u \equiv_0 E_t$

So we have $E_s \preceq_0 E_u \equiv_0 E_t \preceq_0 E_v$ that is, $v \leq_{E_0} t \equiv_{E_0} u \leq_{E_0} s$.

Plausibility model We represent the agent's beliefs and knowledge via the plausibility model described in Figure 5.8 consisting of four possible states (s, t, u, v) where the double circled state indicates the real world and the arrows represent the plausibility relation on states (we skip the reflexive and transitive arrows).

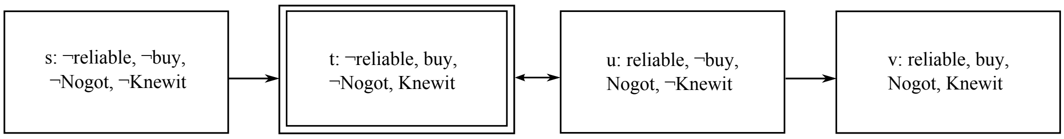

Figure 5.8: Initial plausibility model

The informal dialogue The dialogue starts with our agent claiming to know that someone in the room owns a Ferrari:

Claimant: At least one person in the room owns a Ferrari.

$$B(P)$$

Ultra-critic: Why do you think so? (Justify!)

$$\neg K \neg(\neg P)$$

Claimant: I believe at least one person in the room owns a Ferrari because Mr. Nogot is reliable (and he showed me the papers stating he owns a Ferrari).

$$\{reliable\} \in \mathcal{E}_{\mathcal{M}_1} \text{ and } B(reliable) \text{ and } B^{reliable}(P)$$

Ultra-critic: Mr. Nogot is not reliable, he lied, he does not own a Ferrari!

$$!\neg reliable$$

Update The announcement of "$\neg$*reliable*" is taken as a public announcement, which formally will change the refined justification model and the plausibility model as described respectively in Figures 5.9 and 5.10:

So $\mathcal{M}_2 := \mathcal{M}_1|(\neg reliable)$ where $E_{\mathcal{M}_2} = \{\{e_2\}, \{e_3\}\}$ with:

– $buy = \{t\} = e_2$

– $\neg reliable = \{s, t\} = e_3$

Note that after the update, $\neg reliable$ is true in all the remaining states $(S_1|(\neg reliable) = \neg reliable)$, so $\neg reliable$ is an evidence set.

We have $\mathcal{E}_{\mathcal{M}_2} = \{\varnothing, \{e_2\}, \{e_3\}, \{e_2, e_3\}\}$.

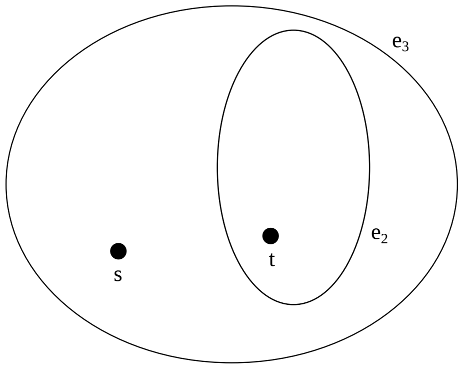

Figure 5.9: Refined justification model updated with $!\neg reliable$

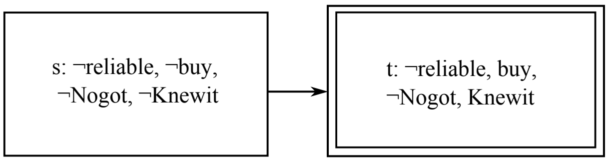

Figure 5.10: Plausibility model updated with $!\neg reliable$

The dialogue ended Then the dialogue ends with our agent claiming:
Claimant: I still believe at least one person in the room owns a Ferrari because I remember Mr. Knewit bought mine.

$$\{buy\} \in \mathcal{E}_{\mathcal{M}_2} \text{ and } B(buy) \text{ and } B^{buy}(P)$$

Conclusion of the dialogue The Ultra-critic cannot object against this last argument because it is actually true that Mr. Knewit bought the Ferrari and so owns one. The Claimant provides a sound justification for P. The Claimant wins, she defeasibly knows at least one person in the room owns a Ferrari because she is justified to accept it in a way that is undefeated by the falsity of any statement (even if her justification for P contains a false statement/belief).

The formal game Formally, we model the informal dialogue as a play in our ultra-justification game $G(\mathcal{M}_0, P)$ as follows.
Claimant: At least one person in the room owns a Ferrari.

$$B(P)$$

This means that at round 0, $F_0 = \varnothing$.
Ultra-critic: Why do you think so? (Justify!)

$$\neg K \neg(\neg P)$$

This means that at round 1, the Ultra-critic chooses $F'_0 = F_0 = \varnothing$ such that F'_0 does not support P: $\bigcap F'_0 = \bigcap \varnothing = S_0 \notin P$ because of the state s ($s \vDash \neg P$).

Claimant: I believe at least one person in the room owns a Ferrari because Mr. Nogot is reliable (and he showed me the papers stating he owns a Ferrari).

$$\{reliable\} \in \mathcal{E}_{\mathcal{M}_1} \text{ and } B(reliable) \text{ and } B^{reliable}(P)$$

This means that the Claimant chooses a new argument $F_1 = \{reliable\}$ in $\mathcal{M}_1 := \mathcal{M}_0$, which is an argument that supports P.

Ultra-critic: Mr. Nogot is not reliable, he lied, he does not own a Ferrari!

$$!\neg reliable$$

This means that at round 2, the Ultra-critic challenges F_1 as unsound and announces that $t \notin \bigcap F_1$ which induces an update $!(\neg \bigcap F_1)$ of the refined justification model and the plausibility model.

Update The announcement of "$\neg$*reliable*" is taken as a public announcement, which formally will change the refined justification model and the plausibility model as described respectively in Figures 5.9 and 5.10:

So $\mathcal{M}_2 := \mathcal{M}_1|(\neg \bigcap F_1)$ where $E_{\mathcal{M}_2} = \{\{e_2\}, \{e_3\}\}$ with:

- $buy = \{t\} = e_2$
- $\neg reliable = \{s, t\} = e_3$

Note that after the update, $\neg reliable$ is true in all the remaining states ($S_1|(\neg \bigcap F_1) = \neg reliable$), so $\neg reliable$ is an evidence set.

We have $\mathcal{E}_{\mathcal{M}_2} = \{\varnothing, \{e_2\}, \{e_3\}, \{e_2, e_3\}\}$.

The game ended Then the game ends with our agent claiming:

Claimant: I still believe at least one person in the room owns a Ferrari because I remember Mr. Knewit bought mine.

$$\{buy\} \in \mathcal{E}_{\mathcal{M}_2} \text{ and } B(buy) \text{ and } B^{buy}(P)$$

This means that the Claimant chooses a new argument $F_2 = \{buy\}$ in $\mathcal{M}_2$, which is a sound justification for P.

Conclusion of the game The Ultra-critic cannot object against this last argument because it is actually true that Mr. Knewit bought the Ferrari and so owns one. The Claimant provides a sound justification for P. The Claimant wins this run of the game, she defeasibly knows at least one person in the room owns a Ferrari because she is justified to accept it in a way that is undefeated by the falsity of any statement (even if her justification for P contains a false statement/belief).

Example 3: Grabit

The last example is divided into two parts.

First scenario First, we suppose an agent called Harry who sees a man he knows very well, Tom Grabit, in the library. Harry can see him taking a book and leaving the library without paying.

The atomic propositions in this example are *see* (Harry saw Tom stealing a book), Tom (Tom stole the book) and $Twin$ (there exists somebody different from Tom who looks just like Tom, that we call "a twin of Tom"). We represent the agent's evidences via the refined justification model $\mathcal{M}_0 = (S_0, E_0, \preceq_0, \|\cdot\|_0, s_0)$ described in Figure 5.11.

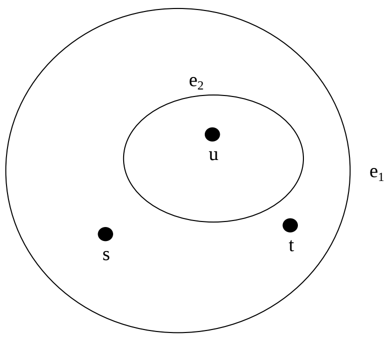

Figure 5.11: Initial refined justification model

Note that again, we do not consider "seeing" as a factual attitude but as being fully introspective. So Harry irrevocably knows that he saw "Tom" stealing a book by introspection.

In accordance with the knowledge of the agent, his epistemic state consists of three worlds $S_0 = s, t, u$. The valuation of the atomic propositions is given as follows: *see* is true at s, t, u, Tom is true at s, u, $Twin$ is true at s, t. So at the state $u = s_0$, Harry saw "Tom" stealing a book and in fact, Tom did steal the book and moreover there does not exist "a twin of Tom". At the state t, Harry saw "Tom" stealing a book, but there exists "a twin of Tom" who actually stole the book (Tom did not steal the book). At the state s, Harry saw "Tom" stealing a book and in fact Tom stole the book, although there exists "a twin of Tom". Formally we have:

- $\| see \|_0 = \{s, t, u\}$,
- $\| Tom \|_0 = \{s, u\}$,
- $\| Twin \|_0 = \{s, t\}$,

In the refined justification model $\mathcal{M}_0$, there are two evidence sets:

$$E_0 = \{see, \neg Twin\}$$

with:

- $see = \{s, t, u\} = e_1$,
- $\neg Twin = \{u\} = e_2$,

The evidence set e_1 represents genuine evidence, i.e. the piece of evidence the agent has, based on his perception: his eyes. The evidence set e_2 represents a bias of the agent. By default, the agent assumes that there does not exist somebody different from Tom who looks just like Tom since he has no evidence to the contrary (he never met such a person).

Note that the evidence set e_1 has the property that $e_1 = see = S_0$.

We have $\mathcal{E}_0 = \{\emptyset, \{e_1\}, \{e_2\}, \{e_1, e_2\}\}$.

The pre-order $\preceq_0$ on $\mathcal{E}_0$ is given by inclusion $\subseteq$, i.e. $F \preceq_0 G$ iff $F \subseteq G$ (so in fact we get an evidence model).[7]

From refined justification model to plausibility model We can easily turn this refined justification model into a plausibility model:

1. $s \leq_{E_0} t$ iff $E_t \preceq_0 E_s$ iff $E_t \subseteq E_s$
2. $E_s := \{e_1\}$, $E_t := \{e_1\}$ and $E_u := \{e_1, e_2\}$
3. $\{e_1\} \preceq_0 \{e_1, e_2\}$ so $E_s \preceq_0 E_u$ and $E_t \preceq_0 E_u$.

So we have $E_s \equiv_0 E_t \preceq_0 E_u$ that is, $u \equiv_{E_0} t \leq_{E_0} s$.

Plausibility model We represent the agent's beliefs and knowledge via the plausibility model described in Figure 5.12 consisting of three possible states (s, t, u) where the double circled state indicates the real world and the arrows represent the plausibility relation on states (we skip the reflexive and transitive arrows).

[7] Note that the cardinality order would lead to the same conclusions about the defeasible knowledge of the agent.

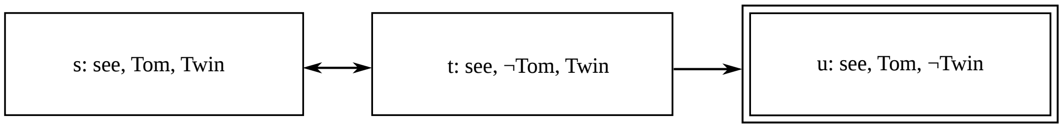

Figure 5.12: Initial plausibility model

The informal dialogue The dialogue starts with our agent claiming to know that Tom stole the book.

Claimant: Tom stole the book.

$$B(Tom)$$

Ultra-critic: Why do you think so? (Justify!)

$$\neg K \neg(\neg Tom)$$

Claimant: I believe Tom stole the book because I saw somebody looking just like Tom stealing a book.

$$\{see\} \in \mathcal{E}_{\mathcal{M}_1} \text{ and } B(see) \text{ and } B^{see}(Tom)$$

Ultra-critic: Maybe there exists "a twin of Tom"!
(Your evidence is consistent with the negation of "Tom"! You need to provide further justification!)

$$\neg K \neg(see \wedge Twin \wedge \neg Tom)$$

Claimant: It is more reasonable for me to accept that Tom stole the book because I saw somebody looking just like Tom stealing a book than to accept that there exists "a twin of Tom"!

$$\{see, \neg Twin\} \in \mathcal{E}_{\mathcal{M}_2} \text{ and } B(see \wedge \neg Twin) \text{ and } B^{see \wedge \neg Twin}(Tom)$$

Conclusion of the dialogue The Ultra-critic cannot object against this last argument because it is actually true that there does not exist "a twin of Tom". The Claimant provides a sound justification for Tom. The Claimant wins, she defeasibly knows Tom stole a book from the library (even despite the remark of the Ultra-critic about the possible existence of "a twin of Tom").

The formal game Formally, we model the informal dialogue as a play in our ultra-justification game $G(\mathcal{M}_0, Tom)$ as follows.

Claimant: Tom stole the book.

$$B(Tom)$$

This means that at round 0, $F_0 = \emptyset$.

Ultra-critic: Why do you think so? (Justify!)

$$\neg K \neg(\neg Tom)$$

This means that at round 1, the Ultra-critic chooses $F'_0 = F_0 = \varnothing$ such that F'_0 does not support Tom: $\bigcap F'_0 = \bigcap \varnothing = S_0 \notin Tom$ because of the state t ($t \vDash \neg Tom$).

Claimant: I believe Tom stole the book because I saw somebody looking just like Tom stealing a book.

$$\{see\} \in \mathcal{E}_{\mathcal{M}_1} \text{ and } B(see) \text{ and } B^{see}(Tom)$$

This means that the Claimant chooses a new argument $F_1 = \{see\}$ in $\mathcal{M}_1 := \mathcal{M}_0$, which is a soft argument in the sense of Definition 4.2.14 that weakly supports Tom conditional on the default $\neg Twin$.

Ultra-critic: Maybe there exists "a twin of Tom"!
(Your evidence is consistent with the negation of "Tom"! You need to provide further justification!)

$$\neg K \neg (see \wedge Twin \wedge \neg Tom)$$

This means that at round 2, the Ultra-critic chooses $F'_1 = F_1 = \{see\}$ because F'_1 does not support Tom: $\bigcap F'_1 = \bigcap \{see\} \notin Tom$ because of the state t ($t \vDash see \wedge Twin \wedge \neg Tom$).

Claimant: It is more reasonable for me to accept that Tom stole the book because I saw somebody looking just like Tom stealing a book than to accept that there exists "a twin of Tom"!

$$\{see, \neg Twin\} \in \mathcal{E}_{\mathcal{M}_2} \text{ and } B(see \wedge \neg Twin) \text{ and } B^{see \wedge \neg Twin}(Tom)$$

This means that the Claimant chooses a new argument $F_2 = \{see, \neg Twin\}$ in $\mathcal{M}_2 := \mathcal{M}_1$, which is a sound argument that supports Tom.

Conclusion of the game The Ultra-critic cannot object against this last argument because it is actually true that there does not exist "a twin of Tom". The Claimant provides a sound justification for Tom. The Claimant wins this run of the game, she defeasibly knows Tom stole a book from the library (even despite the remark of the Ultra-critic about the possible existence of "a twin of Tom").

Second scenario Suppose again that the agent Harry sees a man he knows very well, Tom Grabit, in the library. Harry can see him taking a book and leaving the library without paying. Suppose now that there is really a Twin of Tom, even if it is actually Tom who stole the book.

The atomic propositions in this scenario are exactly the same as those in the first scenario. The refined justification model $\mathcal{M}_0 = (S_0, E_0, \preceq_0, \|\cdot\|_0, s_0)$ representing the agent's evidences in this second scenario is identical to the refined justification model in the first scenario with one exception: the actual state $s_0 = s$.

We represent the agent's beliefs and knowledge via the plausibility model induced from the refined justification model $\mathcal{M}_0 = (S_0, E_0, \preceq_0, \|\cdot\|_0, s_0)$ described in Figure 5.13.

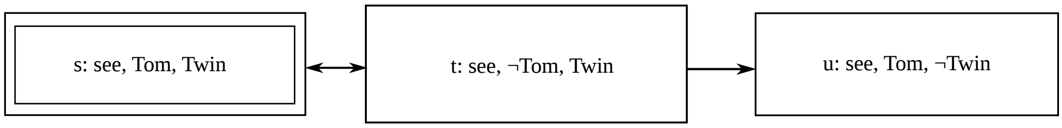

Figure 5.13: Initial plausibility model

The informal dialogue Now the dialogue starts and the Claimant claims to know that Tom stole the book.

Claimant: Tom stole the book.

$$B(Tom)$$

Ultra-critic: Why do you think so? (Justify!)

$$\neg K \neg(\neg Tom)$$

Claimant: I believe Tom stole the book because I saw somebody looking just like Tom stealing a book.

$$\{see\} \in \mathcal{E}_{\mathcal{M}_1} \text{ and } B(see) \text{ and } B^{see}(Tom)$$

Ultra-critic: Maybe there exists a Twin of Tom!
(Your evidence is consistent with the negation of "Tom"! You need to provide further justification!)

$$\neg K \neg(see \wedge Twin \wedge \neg Tom)$$

Claimant: It is more reasonable for me to accept that Tom stole the book because I saw somebody looking just like Tom stealing a book than to accept that there exists "a twin of Tom"!

$$\{see, \neg Twin\} \in \mathcal{E}_{\mathcal{M}_2} \text{ and } B(see \wedge \neg Twin) \text{ and } B^{see \wedge \neg Twin}(Tom)$$

Ultra-critic: There exists a Twin of Tom!

$$!Twin$$

Updates The announcement of "$Twin$" is taken as a public announcement, which formally will change the refined justification model and the plausibility model as described respectively in Figures 5.14 and 5.15:

So $\mathcal{M}_3 := \mathcal{M}_2|(Twin)$ where $E_{\mathcal{M}_3} = \{see, Twin\}$ with $see = \{s,t\} = e_1$ and $Twin = \{s,t\} = e_3$.

Note that after the update, $Twin$ is true in all the remaining states ($S_2|(Twin) = Twin$), so $Twin$ is an evidence set.

We have $\mathcal{E}_{\mathcal{M}_3} = \{\varnothing, \{e_1\}, \{e_3\}, \{e_1, e_3\}\}$.

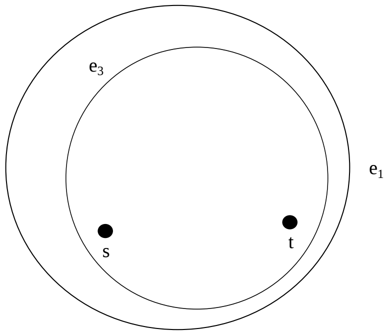

Figure 5.14: Refined justification model updated with $!Twin$

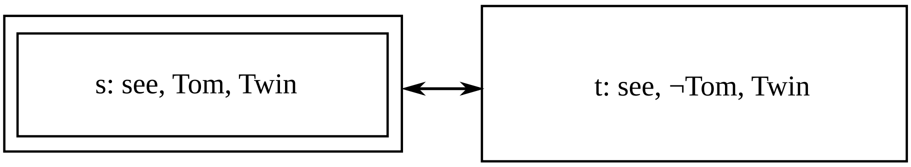

Figure 5.15: Plausibility model updated with $!Twin$

The dialogue ended Then the dialogue ends with our agent claiming:

Claimant: Given that there exists a Twin of Tom, I have no justification left to believe Tom stole the book. So I give up: I no longer believe it!

Conclusion of the dialogue The agent loses because she cannot provide further argument for Tom. She does not even believe Tom anymore. Then the Claimant didn't defeasibly "know" that Tom stole the book. In this case our agent does not defeasibly know Tom since she based her belief on false evidence. Sadly enough, her initial justified belief was in fact true: even if Tom has actually a Twin, he did steal the book. The agent did not know: she only had a true justified belief.

The formal game Formally, we model the informal dialogue as a play in our ultra-justification game $G(\mathcal{M}_0, Tom)$ as follows.

Claimant: Tom stole the book.

$$B(Tom)$$

This means that at round 0, $F_0 = \varnothing$.

Ultra-critic: Why do you think so? (Justify!)

$$\neg K \neg(\neg Tom)$$

This means that at round 1, the Ultra-critic chooses $F'_0 = F_0 = \varnothing$ such that F'_0 does not support Tom: $\bigcap F'_0 = \bigcap \varnothing = S_0 \nsubseteq Tom$ because of the state t ($t \vDash \neg Tom$).

Claimant: I believe Tom stole the book because I saw somebody looking just like Tom stealing a book.

$$\{see\} \in \mathcal{E}_{\mathcal{M}_1} \text{ and } B(see) \text{ and } B^{see}(Tom)$$

This means that the Claimant chooses a new argument $F_1 = \{see\}$ in $\mathcal{M}_1 := \mathcal{M}_0$, which is a soft argument in the sense of Definition 4.2.14 that weakly supports Tom conditional on the default $\neg Twin$.

Ultra-critic: Maybe there exists a Twin of Tom!
(Your evidence is consistent with the negation of "Tom"! You need to provide further justification!)

$$\neg K \neg (see \wedge Twin \wedge \neg Tom)$$

This means that at round 2, the Ultra-critic chooses $F'_1 = F_1 = \{see\}$ because F'_1 does not support Tom: $\bigcap F'_1 = \bigcap\{see\} \nsubseteq Tom$ because of the state t ($t \vDash see \wedge Twin \wedge \neg Tom$).

Claimant: It is more reasonable for me to accept that Tom stole the book because I saw somebody looking just like Tom stealing a book than to accept that there exists "a twin of Tom"!

$$\{see, \neg Twin\} \in \mathcal{E}_{\mathcal{M}_2} \text{ and } B(see \wedge \neg Twin) \text{ and } B^{see \wedge \neg Twin}(Tom)$$

This means that the Claimant chooses a new argument $F_2 = \{see, \neg Twin\}$ in $\mathcal{M}_2 := \mathcal{M}_1$, which is an argument that supports Tom.

Ultra-critic: There exists a Twin of Tom!

$$!Twin$$

This means that at round 3, the Ultra-critic challenges F_2 as unsound and announces that $s \notin \bigcap F_2$ which induces an update $!(\neg \bigcap F_2)$ of the refined justification model and the plausibility model.

Update The announcement of "Twin" is taken as a public announcement, which formally will change the refined justification model and the plausibility model as described respectively in Figures 5.14 and 5.15:

So $\mathcal{M}_3 := \mathcal{M}_2|(\neg \bigcap F_2)$ where $E_{\mathcal{M}_3} = \{see, Twin\}$ with $see = \{s, t\} = e_1$ and $Twin = \{s, t\} = e_3$.

Note that after the update, $Twin$ is true in all the remaining states ($S_2|(\neg \bigcap F_2) = Twin$), so $Twin$ is an evidence set.

We have $\mathcal{E}_{\mathcal{M}_3} = \{\varnothing, \{e_1\}, \{e_3\}, \{e_1, e_3\}\}$.

The game ended Then the game ends with our agent claiming:

Claimant: Given that there exists a Twin of Tom, I have no justification left to believe Tom stole the book. So I give up: I no longer believe it!

Conclusion of the game The agent loses this round of the ultra-justification game because she cannot provide further argument for Tom. She does not even believe Tom anymore. Then the Claimant loses the game: she didn't defeasibly "know" that Tom stole the book. In this case our agent does not defeasibly know Tom since she based her belief on false evidence. Sadly enough, her initial justified belief was in fact true: even if Tom has actually a Twin, he did steal the book. The agent did not know: she only had a true justified belief.

Conclusion

We provided a formalisation of Lehrer's ultra-justification game allowing to determine if an agent really (defeasibly) knows some given proposition or if she only believes this proposition. We then proved that an agent defeasibly knows a given proposition iff she continues to believe this proposition as long as she receives only true information (pieces of evidence). We provided the rules for our game semantics for defeasible knowledge stating that an agent defeasibly knows a proposition iff she has a winning strategy in the corresponding formal ultra-justification game.

In the next part, we connect Soft Dynamic Epistemic Logic with other settings. We start by investigating the relations between *DEL* and the belief revision setting of Dynamic Doxastic Logic in Chapter 6. We show that *DDL* can internalize all the recent *DEL* developments for belief revision.

Part III

Connecting frameworks for belief revision: from Soft Dynamic Epistemic Logic to dialogues

Chapter 6

Dynamic Doxastic Logic from the perspective of Dynamic Epistemic Logic

Aim In this chapter we study Segerberg's "full *DDL*" (Dynamic Doxastic Logic) from the new perspective of "soft *DEL*" (the belief-revision-friendly version of Dynamic Epistemic Logic), as a modern semantic embodiment of the *AGM* paradigm. One of our main goals is to show that the *DDL* approach is at least as powerful as the *DEL* approach: it can internalize all the recent *DEL* developments for belief revision.

Summary: In this chapter we re-evaluate Segerberg's "full *DDL*" from the perspective of Dynamic Epistemic Logic (*DEL*), in its belief-revision-friendly incarnation. We first present an appropriately generalized and simplified version of "full *DDL*". Next, we argue that a correct version of "full *DDL*" must give up the Success Postulate for dynamic revision. We construct *AGM*-friendly versions of "full *DDL*", corresponding to various revising/contracting operations considered in the Belief Revision literature. The main points are:

- We provide a general presentation of a simplified version of "full *DDL*". The semantics is based on Segerberg's hypertheories which gives a generalization of Segerberg's onion-based semantics. We further simplify Segerberg's setting by dropping all the topological assumptions as well as all the closure assumptions on hypertheories.

- We claim that a correct version of "full *DDL*" must give up the Success Postulate for dynamic revision.

- We deal with static revision by adopting the conditional belief logic *CDL* and we develop three versions of *DDL* that internalize three of the revision operations considered in the Belief Revision literature.

- We introduce and axiomatize three *AGM*-friendly versions of belief contraction and expansion in *DDL*.

Background We introduce the background for the three main frameworks we are dealing with in this chapter:

The first framework, Dynamic Doxastic Logic (*DDL*), has been introduced and developed by Krister Segerberg in [74, 75, 76, 77, 78, 79, 73, 80]. As we briefly mentioned in Chapter 2, the main idea of Segerberg was to enhance traditional epistemic and doxastic logics with specific dynamic-modal operators for belief revision, thus linking modal logic with Belief Revision Theory (*BRT*). Looking the other way around, Segerberg's work provided *BRT* with a new syntax and formal semantics. Traditionally, the work on belief revision [1] focuses on the way in which a given theory (or belief base, consisting of sentences in a given object language) gets revised, but it does not treat "belief revision" itself as an ingredient in the object language under study (see Chapter 2 for detail). Segerberg's work opened up a new perspective by taking the very act of belief revision itself and placing it on an equal footing with the doxastic attitudes such as "knowledge" and "belief". His dynamic-modal operators describe transitions in doxastic models that model belief change. Using this setting Segerberg provides modal axioms encoding the *AGM* postulates.

The second framework, Conditional Doxastic Logic (*CDL*), has been introduced by Alexandru Baltag and Sonja Smets in [6]. For details we refer to Chapter 2 where we explained that *CDL* extends modal logic with conditional doxastic belief operators. It is important to note that the interpretation of $B^{\varphi}\psi$ is taken to be "if the agent would learn φ, then she would believe ψ was the case before the learning". Conditional beliefs capture static, purely hypothetical, revision.

The last framework, Dynamic Epistemic Logic (*DEL*), extends modal logic with dynamic operators to deal with the actual knowledge dynamics of an agent [36, 35, 5, 29]. For details we refer to Chapter 2. Here we stress that *DEL* takes epistemic models as basis and investigates how such models evolve under receipt of new information. Recently a belief-revision-friendly version of Dynamic Epistemic Logic has been developed [6, 7, 8, 11, 3, 10]. This "soft *DEL*" considers three different operations to model belief change: update, lexicographic upgrade and conservative upgrade. Each operation changes the epistemic/doxactic models differently expressing different types of belief revision. In this setting some Reduction/Recursion Laws are given to provide complete axiomatizations of the dynamic logics of these three kinds of belief revision.

Recently Johan van Benthem showed in [12] that Segerberg's Dynamic Doxastic Logic and the *DEL* tradition co-exist in the perspective of modal frame correspondence. He provides a correspondence analysis of modal logics for belief change, using recursion axioms as constraints on possible update operations. We want to add a complement to these results showing how the *DEL*-style of modelling and axiomatizing belief revision can be "internalized" in *DDL*.

6.1 Dynamic Doxastic Logic revisited

We introduce the setting of Dynamic Doxastic Logic in Chapter 2. In this section, we first present our own version of full *DDL* which is a generalized and simplified version of original full DDL. Note that in this chapter, we do not restrict ourselves to finite models. Next, we go over a main philosophical issue concerning the validity of the so-called Success Axiom in a dynamic setting. To address this, we follow the *DEL* literature in distinguishing between "static" and "dynamic" belief revision. Though it is often explained in syntactic terms (as referring to two different kinds of behaviour under revision with higher-level doxastic sentences), from a semantic point of view this distinction is in fact related to (though distinct from) the traditional dichotomy between one-step revision and iterated revision [23, 26, 62, 72].

6.1.1 General presentation

We present a generalized and simplified version of the "General Model Theory" for *DDL* introduced by Segerberg in Section 3 of [76]. The semantics is based on Segerberg's hypertheories which are families of sets of states, called fallbacks. In fact these hypertheories are generalizations of Segerberg's onions which are families of nested sets of states, called spheres in accordance to the Lewis-Grove tradition.

As a formal language to describe these models, we use the slightly extended syntax for *DDL* introduced in [80], having in addition to belief operators B and dynamic modalities $[\alpha]$, operators K for what Leitgeb and Segerberg call "nonrevisable belief" or "knowledge". We call this "irrevocable knowledge" to distinguish it from other "softer" notions of knowledge considered in the philosophical literature, namely defeasible knowledge[1]. To ensure that the K operator is factive (as expected for knowledge), we make a slight change to the definition of validity, inspired from the Moss-Parikh semantics of Topo-logic[2] [60, 61]: validity is obtained by quantifying only over pairs (s, H) of ontic states and hypertheories such that $s \in \bigcup H$.

We further simplify Segerberg's setting from [76], by dropping all the topological assumptions – Stone spaces, compactness assumptions –, as well as all the

---

[1]We refer to the introduction for more details about different notions of knowledge.

[2]Topo-logic frames $(U, \mathcal{T}, V)$ consists of a universe (set of "states") U, a family $\mathcal{T} \subseteq \mathcal{P}(U)$ of sets of states (called "opens") and a valuation V for the atomic sentences of their language. While the points $s \in U$ represent possible ontic states, the opens $V \in \mathcal{T}$ represent possible information states. When the agent's information state is V, this means that the only thing that she knows about the state of the world is that it belongs to V. Sentences are evaluated at pairs (s, V) of an ontic state $s \in U$ and information state $V \in \mathcal{T}$, with the restriction that $s \in V$ so that "knowledge" is factual. Indeed, these are information states, rather than doxastic states.

closure assumptions on hypertheories – e.g. Lewis' famous Limit Assumption[3] [52], or the assumption from [80] of closure under nonempty intersections.

The price for this generality is that the definition of belief is more complicated: we adopt the definition of B introduced by van Benthem and Pacuit [16]. But we show that, whenever hypertheories do satisfy closure under intersection, this definition boils down to Segerberg's notion of belief which is the same as Grove's definition: belief equals truth in all the states of the smallest sphere. Moreover, we show that in the special case of onions, this definition amounts to a natural generalization of Grove's definition: belief equals truth in all the states of all the spheres that are "small enough", that was already proposed in the Belief Revision literature (and which validates the same modal formulas as Grove's standard definition). Finally, in case of onions satisfying Lewis' Limit assumption, this definition boils down again to the standard Grove-Segerberg notion of belief.

Syntax The language of full *DDL* is defined as follows.

6.1.1. DEFINITION. Assume as given any object language $\mathcal{L}_{DDL}$ consisting of well-formed formulas build up from the following ingredients: propositional letters coming from a set Φ, Boolean connectives, a belief operator B, an irrevocable knowledge operator K, a set A of action terms, as well as the dynamic modalities $[\alpha]$ ("after action α") of Propositional Dynamic Logic (one for each action term $\alpha \in A$). Any such language $\mathcal{L}_{DDL}$ is called a *DDL*-language. The minimal language of full *DDL* has only the above operators[4].

Semantics We interpret $\mathcal{L}_{DDL}$ in *DDL* models.

6.1.2. DEFINITION. Let U be a set of states (a *universe*). A *hypertheory* in U is a nonempty family $H \subseteq \mathcal{P}(U)$ of nonempty subsets of U, called fallbacks. An *onion* (or sphere system) in U is a hypertheory $O \subseteq \mathcal{P}(U)$ that is nested, i.e. linearly ordered by set-inclusion: $X, Y \in O$ implies that either $X \subseteq Y$ or $Y \subseteq X$. The elements of an onion (its fallbacks) are sometimes called spheres.

We think of each $s \in U$ as an ontic state: a possible description of all the ontic (i.e. non-doxastic) facts of the world. We think of a hypertheory H as representing the agent's doxastic state. In particular, as we will see in the next section, an onion O will represent a doxastic state that satisfies the *AGM* postulates (when these postulates are appropriately stated, as axioms about static revision). Hypertheories represent the current belief state of an agent such that fallbacks are theories that can be viewed as alternative belief sets from which the agent can build a new belief state in cases where doxastic actions (contraction, expansion or revision) happen.

[3]We provide the definition of Lewis' Limit Assumption in Definition 6.1.9.
[4]Later we will add conditional belief operators to describe static revision.

6.1.3. Definition. An onion O is *standard* (or well-founded) if there is no infinite descending chain of spheres in O, i.e. there is no infinite sequence $X_1 \supset X_2 \supset X_3 \supset \ldots$, with all $X_i \in O$.

6.1.4. Definition. Given a hypertheory $H \subseteq \mathcal{P}(U)$, a family $F \subseteq H$ of fallbacks has the *finite intersection property* (f.i.p.) if every finite subfamily $F' \subseteq F$ has a non-empty intersection $\bigcap F' \neq \varnothing$. We say that a family $F \subseteq H$ of fallbacks has the *maximal f.i.p.* if F has the f.i.p. but no proper extension $F \subset G \subseteq H$ does.

Note that if O is an onion then O has itself the maximal f.i.p. and moreover O is the only family $F \subseteq O$ having the maximal f.i.p.

6.1.5. Definition. An *A-doxology* is a structure (U, D, R), where U is a universe, D is a set of hypertheories in U and $R = \{R^\alpha\}_\alpha$ is a set of binary relations $R^\alpha \subseteq D \times D$ on D, labelled with names $\alpha \in A$ coming from a given set A of action terms.

The elements $R^\alpha \in R$ are called doxastic actions, and R itself a repertoire. Note that each R^α is a binary relation between hypertheories (or onions), not between states. Intuitively, each R^α describes a specific type of change which may affect the agent's epistemic/doxastic state but which does not change the ontic state.

6.1.6. Definition. A *DDL model* $M = (U, D, R, V)$ for any *DDL* language $\mathcal{L}_{DDL}$ (with propositional letters in Φ and action terms in A) consists of an A-doxology (U, D, R) together with a valuation V, mapping every propositional letter $p \in \Phi$ to a set $V(p) \subseteq U$ of states.

6.1.7. Definition. An *onion model* is a *DDL* model (U, D, R, V) in which D consists only of onions.

6.1.8. Definition. An onion model (U, D, R, V) is *standard* if all the onions $O \in D$ are standard.

A weakening of the standardness condition, which has the disadvantage of being language-dependent is the so-called Lewis Limit Assumption.

6.1.9. Definition. An onion model (U, D, R, V), together with a semantics $\|.\|$ is said to satisfy the *Limit Assumption* if, for every formula $\varphi \in \mathcal{L}_{DDL}$ and every onion $O \in D$, we have that: $\|\varphi\| \cap \bigcup O \neq \varnothing$ implies $\bigcap\{X \in O : \|\varphi\| \cap X \neq \varnothing\} \in O$.

Standard onion models always satisfy the Limit Assumption for every language $\mathcal{L}_{DDL}$, but the converse is false. In fact, standard onion models satisfy a stronger condition, that we call the Strong Limit Assumption: for every set $P \subseteq U$ of states and every onion $O \in D$, $P \cap \bigcup O \neq \varnothing$ implies $\bigcap\{X \in O : P \cap X \neq \varnothing\} \in O$.

This means that, in a standard model, every onion intersecting a given set P contains a unique smallest sphere intersecting P.

DDL model can also satisfy the Limit Assumption.

6.1.10. Definition. A *DDL* model (U, D, R, V), together with a semantics $\|.\|$ is said to satisfy the *Limit Assumption* if, for every formula $\varphi \in \mathcal{L}_{DDL}$ and every hypertheory $H \in D$, we have that: $\|\varphi\| \cap \bigcup H \neq \varnothing$ implies $\bigcap\{X \in H : \|\varphi\| \cap X \neq \varnothing\} \in H$.

6.1.11. Definition. A *static DDL model* is a *DDL* model with $R = \varnothing$.

A semantics for $\mathcal{L}_{DDL}$ is a map that, for each *DDL* model $M = (U, D, R, V)$ and each hypertheory $H \in D$, assigns to each formula $\varphi \in \mathcal{L}_{DDL}$ some set of states $\|\varphi\|_{M,H} \subseteq \bigcup H$, and assigns to each action term $\alpha \in A$ some doxastic action $\|\alpha\|_{M,H} \in R$, in such a way that a number of conditions (to be given below) are satisfied. Our restriction to $\bigcup H$ is motivated by the intuition that the states $s \notin \bigcup H$ represent "impossible states": ontic states that are excluded by the doxastic state H. In other words, $\bigcup H$ encompasses the agent's "hard information" about the world. As a consequence, the operator K (given by quantifying over $\bigcup H$) is factive (unlike in the usual setting of *DDL*): we can think of K as representing the agent's knowledge, in the absolute sense of infallible, absolutely certain, and absolutely unrevisable knowledge. We use the notation

$$s, H \vDash_M \varphi$$

such that $s \in \bigcup H$ whenever we have $s \in \|\varphi\|_{M,H}$, and we delete the subscript(s) whenever it is possible to do this without ambiguity, writing e.g. $\|\varphi\|_H$ and $s, H \vDash \varphi$ when M is fixed, or even $\|\varphi\|$ when both M and H are fixed.

6.1.12. Definition. A semantics for $\mathcal{L}_{DDL}$ is required to satisfy the following constraints:

$$\begin{array}{lll}
s, H \vDash p & \text{iff} & s \in V(p) \\
s, H \vDash \neg\varphi & \text{iff} & s, H \nvDash \varphi \\
s, H \vDash \varphi \wedge \psi & \text{iff} & (s, H \vDash \varphi) \wedge (s, H \vDash \psi) \\
s, H \vDash B\varphi & \text{iff} & \forall F \subseteq H \text{ such that } F \text{ has the maximal f.i.p. } \exists F' \text{ finite} \subseteq F \\
 & & \forall t \in \bigcap F' \, (t, H \vDash \varphi) \\
s, H \vDash K\varphi & \text{iff} & \forall t \in \bigcup H \, (t, H \vDash \varphi) \\
s, H \vDash [\alpha]\varphi & \text{iff} & \forall H' \in D \, ((H, H') \in \|\alpha\|_H \wedge s \in \bigcup H' \implies s, H' \vDash \varphi)
\end{array}$$

Our definition of irrevocable knowledge K is essentially the same as in [80], except that our modified definition of validity entails the factivity of K, making it behave indeed like a notion of knowledge (in contrast to [80]). Our definition of belief B is a generalization of the Grove-Segerberg definition, due to van Benthem and Pacuit [16]. But it can be simplified in onion models (where it boils down to a widely used generalization of Grove's), and it can be simplified further when we have either the Limit Condition or closure under intersection (where it boils down to the Grove-Segerberg definition):

6.1.13. Proposition. *In DDL models in which the set D consists only of hypertheories H that satisfy the Limit Assumption, φ is believed iff it is true in all the "most plausible states" – i.e. the states of the smallest fallback:*

$$s, H \models B\varphi \textit{ iff } \forall t \in \bigcap H \, (t, H \models \varphi).$$

6.1.14. Proposition. *In onion models, φ is believed iff φ is true in all the states that are "plausible enough" – i.e. throughout all the spheres that are "small enough":*

$$s, O \models B\varphi \textit{ iff } \exists X \in O \, \forall t \in X \, (t, O \models \varphi).$$

Moreover, in onion models satisfying the Limit Assumption, this boils down to the usual Grove definition:

$$s, O \models B\varphi \textit{ iff } \forall t \in \bigcap O \, (t, O \models \varphi).$$

And, as a consequence, this equivalence holds in standard onion models.

For a class $\mathcal{C}$ of *DDL* models, we write $\mathcal{C} \models \varphi$ and we say that φ is valid on $\mathcal{C}$, if $\|\varphi\|_{M,H} = U$ for every model $M = (U, D, R, V) \in \mathcal{C}$ and every $H \in D$; equivalently, iff $s, H \models_M \varphi$ holds for all models $M = (U, D, R, V) \in \mathcal{C}$, all hypertheories $H \in D$ and all states $s \in \bigcup H$.

Note We can easily establish correspondences between Segerberg's onions and hypertheories and plausibility models: both semantic styles are equivalent if considered at an appropriate level of generality. Indeed any fallback H in a *DDL* model induces a corresponding relation of plausibility between states. We say that state s is at least as plausible as state t according to H, and we write $s \leq_H t$, if s belongs to all the fallbacks in H that contain t:

$$s \leq_H t \text{ iff } \forall X \in H (t \in X \Rightarrow s \in X).$$

Obviously, the plausibility relation $\leq_H$ is a preorder (reflexive and transitive relation) on the set $\bigcup H$. Moreover, if O is an onion, then $\leq_O$ is a total (i.e. connected) preorder on $\bigcup O$: for all $s, t \in \bigcup O$, we have either $s \leq_O t$ or $t \leq_O s$ (or both).

6.1.2 Static versus Dynamic belief revision

Static revision To model one-step revision, it is enough to specify the result of doxastic revision with P for every proposition P, either syntactically – as a set of sentences – or semantically – as a set of states, the ones that are most plausible after revising with P. Semantically, this can be uniformly done in three different ways by giving:

- a selection function, in Stalnaker's style;
- a family of spheres in Lewis-Grove style – i.e. an onion in the sense of Segerberg (or a hypertheory in his generalized semantics);
- a plausibility relation.

As far as modal Dynamic Doxastic Logic can tell, these three semantic styles are equivalent if considered at an appropriate level of generality.

Syntactically, one can capture static revision by specifying in AGM-style, a set $T * P$ of revised beliefs for each original set T of beliefs and each proposition P; or alternatively, one can encode static revision using conditional belief operators $B^P Q$, whose meaning is that "after revision with P, the agent will come to believe that Q was the case before the revision". The static character of this revision is reflected in the fact that, after the revision, Q is still evaluated according to the original state of affairs. In terms of Grove spheres, this is reflected in the fact that the same onion is used for evaluating Q (though not the same sphere): $B^P Q$ holds iff the smallest sphere in the current onion that intersects P is included in Q.

Dynamic revision In contrast dynamic revision involves a change of onion, or a change of plausibility relation, or a change of model. Semantically, it requires a binary relation between onions (in *DDL*-style), or between states with different plausibility (in *PDL* style), or between models (in *DEL*-style). Again, these three styles of doing doxastic dynamics are equivalent if considered at an appropriate level of generality. Syntactically, dynamic revision can be captured by the use of dynamic modalities $[*P]Q$. More precisely, $[*P]BQ$ captures the fact that Q is believed to hold after revision with P. The dynamic character is reflected in the fact that, after the revision, Q is evaluated according to the new state of affairs. In terms of Grove spheres, this is reflected in the fact that BQ is evaluated using the new onion to which the old onion is related by the dynamic binary relation R^{*P}.

Dynamic operators and reversed dynamic operators The static character of conditional belief operators $B^P Q$ can be made more explicit by expressing them in terms of dynamic operators $[*\varphi]\psi$ and the reversed dynamic operators $\langle *^{-1}\varphi\rangle\psi$. While in the Segerberg's onion semantics, dynamic operators $[*\varphi]\psi$ are the universal (Box) modalities for some binary revision relation $R^{*\varphi}$ between onions, the reversed dynamic operators $\langle *^{-1}\varphi\rangle\psi$ are the existential (Diamond) modalities for the converse relation $(R^{*\varphi})^{-1}$ (going backwards in time from the revised doxastic state to the initial, unrevised doxastic state). Then we have the following equivalence:

$$B^{\varphi}\psi \Leftrightarrow [*\varphi]B\langle *^{-1}\varphi\rangle\psi.$$

This equivalence fully captures our above explanation of static revision $B^{\varphi}\psi$, as reflecting the revised beliefs after a revision with φ about a sentence ψ's truth value before the revision.

Belief revision plans Nevertheless, we choose not to reduce static revision to dynamic revision (and its converse). Instead, we take static revision as basic, in the shape of primitive conditional belief operators $B^{\varphi}\psi$ interpreted as belief-revision plans: "if in the future I ever would have to revise with φ, I would then come to believe that ψ was true now". And we follow the *DEL* tradition by recursively reducing any instance of dynamic revision to the static revision statements (via so-called Reduction laws, or Recursion laws).

We choose this option because we think that, from a semantic point of view, static belief revision is a simpler concept than the dynamic one. Indeed, recall that to specify static revision one only needs to give one onion (together with a specific way to move between its spheres). While dynamic belief revision is given by a specific type of onion change – i.e. a specific way of moving between onions namely, a relation between onions. So in fact, dynamic belief revision does not involve only a simple revision of beliefs, but rather a revision of static belief revision plans. Indeed, to syntactically describe in full a given type of dynamic belief revision, we do not need only statements of the form $[*P]B\,Q$ (describing dynamic revision of beliefs), but rather sentences of the form $[*P]B^{R}Q$ (describing dynamic revision of static belief-revision plans).

Luckily, this distinction does not need to be iterated: since (to use van Benthem's expression) static belief revision $B^{R}Q$ pre-encodes dynamic belief revision $[*R]BQ$, it is enough to know the behaviour $[*P]B^{R}Q$ of static revision plans under dynamic revision in order to be able to calculate the result of iterated dynamic revision $[*P][*R]B\,Q$. More generally, for each specific type of doxastic dynamic revision $*$, the statement $[*P]Q$ can be recursively reduced to a statement involving only static revision operators $B^{R}Q$: these are the well-known Reduction (or Recursion) Laws, from Dynamic Epistemic Logic. Thus, dynamic revision can be straightforwardly iterated by its very semantic modelling while static belief revision is just a one-step revision of (simple) beliefs.

Static versus Dynamic revision The distinction of static versus dynamic revision is not the same as the distinction between one-step and iterated revision. The distinction is that dynamic revision fully "keeps up" with the doxastic change, while static revision looks back at the old doxastic state from the perspective of the new one. Indeed dynamic revision with higher-level doxastic sentences behaves differently than static revision.

Consider a Moore sentence[5] of the form $\varphi := p \wedge \neg Bp$ [59, 28]. An introspective agent will obviously not come to believe φ after she learns φ. Indeed, believing φ would amount to a lack of introspection since it would mean the agent believes both p and the fact that she doesn't believe p. So, after learning φ, an introspective agent will come to believe p, but not φ itself. This is correctly reflected by dynamic revision: as we will see, for any reasonable dynamic interpretation of the revision operation $*$ as a binary relation on doxastic states (onions), the formula $[*\varphi]B\varphi$ is false for any Moore sentence φ. Indeed, even if φ was true in the old doxastic state, after revision with φ the sentence $B\varphi$ is evaluated according to the new doxastic state, in which φ is false, and known to be false, hence dis-believed. In contrast, static revision with any sentence φ will always produce belief in that sentence, since after static revision, the sentence is still evaluated according to the original doxastic state: this is reflected by the conditional-belief validity $B^{\varphi}\varphi$, which is a version of the *AGM* Success Postulate $\varphi \in T * \varphi$.

This distinction is an important one, that *DDL* needs to learn from *DEL*, in order to deal correctly with higher-level doxastic sentences. Ignoring this distinction leads to the failure of the Success Postulate in the papers of Lindstrom and Rabinowicz [54, 53] on *DDL* for introspective agents, as well as in Segerberg's paper [78]. Namely, these authors assume (mistakenly, in our view) that a dynamic version of the Success postulate (in the form of the axiom $[*\varphi]B\varphi$) is desirable, or even tenable, in full *DDL* (i.e. when φ is itself a doxastic sentence). As we argue below (and as was already argued before in the *DEL* literature), this assumption is wrong.

6.1.3 Full *DDL* and the Success Postulate

Lindstrom and Rabinowicz's Semantics Lindstrom and Rabinowicz propose two solutions to the Moore paradox. We should stress that the failure of the Success Postulate affects only their first solution (in the first part of their paper [54]). There, they define a semantics for revision, which together with their standard *PDL*-like semantics for dynamic modalities, can be shown to immediately lead to a semantic failure of the Success Postulate for any (positively) introspective agent.

[5]Initially the Moore paradox is formulated in terms of knowledge. But it also works when formulated in terms of belief.

They represent propositions by sets of possible worlds in some space U. They define a topology T over U in which the closed sets are the arbitrary intersections of propositions. T is determined by the family C of all closed sets and consists of all the open sets (the subsets of U that are complements of the sets in C). They impose the compactness condition on the topology: any family of closed sets with an empty intersection includes a finite subfamily that also has an empty intersection. The propositions are represented by the clopen sets (the sets that are both closed and open).

They define a model M to be a structure $< U, Prop, w, d, b, R, V >$, where

- U is the set of all possible worlds also called total states
- $Prop$ is a Boolean set-algebra with domain U and is the set of propositions
- w is a function assigning to each total state $x \in U$ a world state $w(x)$
- d is a function assigning to each total state $x \in U$ a doxastic state $d(x)$ of the agent
- b is a function assigning to each total state $x \in U$ the set of states that are compatible with what is believed in x such that if $d(x) = d(y)$ then $b(x) = b(y)$
- R is a function that for every doxastic action term τ yields an accessibility relation $R^\tau \subseteq U \times U$
- V is a valuation.

In a Lindstrom-Rabinowicz model, formulas are evaluated at total states x, each coming with an ontic state $w(x)$ and a doxastic state $d(x)$. In their turn, doxastic states $d(x)$ are Segerberg onions (or more generally hypertheories): these are families of spheres (i.e. of closed sets of total states). If we put $b(x) := \bigcap d(x)$ for the "smallest sphere" of the onion $d(x)$, then belief is defined as usually in Grove models: $x \models B\varphi$ iff $b(x) \subseteq \|\varphi\|$.

Take now any Lindstrom-Rabinowicz model M in which the following two conditions are satisfied: (a) the agent is positively introspective with respect to some specific fact p (at all the states of the model), and (b) there exists some total state x in which the agent doesn't believe p and she doesn't believe $\neg p$. It seems clear that, no matter what additional restrictions one might want to impose on Lindstrom-Rabinowicz models, situations satisfying (a) and (b) should still be allowed[6]. So, even if we add further conditions, a model M of the above kind

[6] Even if one doesn't accept Positive Introspection as a general axiom, one certainly shouldn't exclude situations in which the agent is introspective, at least with respect to some particular fact p.

should still be in the intended class of models. As a consequence of (b), the smallest sphere $b(x) := \bigcap d(x)$ (at total state x) contains both p and $\neg p$ worlds.

In this situation, the Moore sentence $\varphi := p \wedge \neg Bp$ is semantically consistent with the agent's (semantic) beliefs. Indeed, φ is true at all the p-worlds belonging to the smallest sphere: $b(x) \cap \|\varphi\| \subseteq \|p\|$. Hence, this smallest sphere $b(x)$ has a non-empty intersection $b(x) \cap \|\varphi\| = b(x) \cap \|p\| \neq \varnothing$ with the extension $\|\varphi\|$ of φ in this model.

The Lindstrom-Rabinowicz semantic conditions, or more precisely their postulates on semantic contraction and their Levi-style definition of revision[7], ensure that in this situation a revision with φ is the same as an expansion with φ (as is also prescribed by the *AGM* theory): so, the total state y obtained after revision (i.e. such that $xR^{*\varphi}y$) is the same as the state obtained by expansion, i.e. we have $xR^{+\varphi}y$. But unlike revision (or contraction), the expansion operation is completely determined by the *AGM* axioms, which are accepted by Lindstrom and Rabinowicz, who in fact explicitly assume that the expanded state y is the unique total state satisfying the conditions $w(y) = w(x)$ (stability of ontic state) and $d(y) = d(x) + \|\varphi\| =: d(x) \cup \{X \cap \|\varphi\| : X \in d(x)\}$. This means that the smallest sphere of the new onion $d(y)$ must be $b(y) = \bigcap d(y) = \bigcap d(x) \cap \|\varphi\| = b(x) \cap \|\varphi\| = b(x) \cap \|p\| \subseteq \|p\|$. As a consequence, in the new total state y, the agent believes p: $y \vDash Bp$. Since Positive Introspection with respect to p holds in this model, we also have $y \vDash BBp$.

If the Success Postulate would also hold, in its dynamic form $x \vDash [*\varphi]B\varphi$, then by the standard *PDL* semantics for dynamic operators (accepted by Lindstrom and Rabinowicz in this part of their paper), we would have $y \vDash B\varphi$. Using the normality of the operator B (which is another immediate consequence of the Lindstrom-Rabinowicz semantic definition of belief) and the fact that $\varphi := p \wedge \neg Bp$, it follows that $y \vDash B\neg Bp$. So we have that $y \vDash (BBp \wedge B\neg Bp)$, and by normality again, we conclude that $y \vDash B(p \wedge \neg Bp)$, which by the semantic definition of B, entails that $b(y) \subseteq \|Bp \wedge \neg Bp\| = \varnothing$. But this contradicts the above-mentioned fact that $b(y) = \bigcap d(y) = \bigcap d(x) \cap \|\varphi\| = b(x) \cap \|\varphi\| = b(x) \cap \|p\| \neq \varnothing$.

This contradiction is obtained only by using the Lindstrom-Rabinowicz semantics for belief and revision, the Success Postulate, and the natural and innocuous assumptions (a) and (b) (i.e. that there occasionally may exist some agent who is introspective with respect to some fact p, while the fact p itself is currently neither believed nor disbelieved by the agent). Since the title of one of the papers presenting their setting is *Belief Change for Introspective Agents* [53], it seems to us that Lindstrom and Rabinowicz do not aim to give up even the mere possibility

[7] We recall that Levi defines the operation of revision as a composed operation since he considers revision as an operation of contraction followed by an operation of expansion : $K * p = (K - \neg p) + p$ [49] also called Levi identity (see Chapter 2).

of Positive Introspection (with respect to even just one factual statement). So it follows that they must give up the Success Postulate.

Lindstrom and Rabinowicz's Solution It is true that, in the second part of their paper [54], Lindstrom and Rabinowicz propose a second solution to the Moore paradox, their so-called bidimensional semantics, which is in fact very close to the *DEL* solution. Indeed, their rendering in English of their proposal is essentially the same as our solution: they point out that the Success Postulate makes sense for doxastic sentences φ only if it is interpreted in terms of the revised beliefs about φ's truth value before the revision. However, they formally package this solution in a different way (different from *DEL*), in order to maintain the appearance at a purely syntactic level, that the Success Postulate is maintained. Namely, they do this by adopting a bidimensional semantics in terms of pairs of states (x, y), in order to refer to both doxastic states (before and after the revision), and they radically change the *PDL* semantics of dynamic operators to a non-standard one: roughly speaking, their new semantics amounts to evaluating any doxastic expression $B\psi$ that comes in the scope of a dynamic operator $[*\varphi]$ as capturing the revised beliefs (after revision with φ) about ψ's truth value before the revision.

We fully agree with the conceptual analysis underlying the second solution of Lindstrom and Rabinowicz, but we disagree with their non-standard modification of the semantics of dynamic operators. We think dynamic modalities should be left to express what they always did: a one-way move in time, from the state before the (revision) action to the state after the action. Instead of twisting the meaning of dynamic operators, we think one should simply recognize that the Success Postulate does not and should not hold for dynamic revision with doxastic sentences.

Segerberg's Solution In most of his papers on *DDL*, Segerberg himself is cautious not to fall into the above mentioned conceptual mistake, by almost always limiting himself to basic *DDL* in which no revision with doxastic sentences is allowed. However, in [78] he proposes an axiomatic system for full *DDL*. Unfortunately, this converts a conceptual mistake into a logical error: the proposed system is not sound with respect to the proposed semantics. The reason is that the proposed Success Axiom $[*\varphi]B\varphi$ is not a validity in this semantics.

The semantic setting in [78] differs slightly from the version of *DDL* presented in our paper [4] since we follow [76, 80], in that it is actually closer to the Lindstrom-Rabinowicz setting: formulas are evaluated at states (called points) – not at pairs of a state and an onion – and so the dynamics is given via binary relations between states (similarly to the standard *PDL* semantics), rather than via relations between onions. The resulting relational frame is called a revision space. However, in this setting (from [78]), each state is assigned an onion, via an "onion determiner", which paired with a revision space gives an "onion frame".

Completeness for an axiomatic system that includes the dynamic version of the Success Axiom is claimed with respect to the class of "*AGM* onion frames" – i.e. onion frames satisfying some additional *AGM*-like semantic conditions. Introspection is not assumed by Segerberg in this setting, neither as a semantic condition nor as an axiomatic one. But it is easy to see that (Positive) Introspection is consistent with this setting: there exist *AGM* onion frames that are positively introspective. More precisely, the above counterexample (an introspective onion model in which neither p nor $\neg p$ are believed) can be easily repackaged as an *AGM* onion model in the sense of [78]. The dynamic version of the Success Axiom, when instantiated to the Moore sentence $p \wedge \neg Bp$, fails in this model. So this axiom is simply not sound.[8]

Dropping Success Postulate The lesson is that in *DDL* (as in *DEL*) we can really make sense of dynamic revision with doxastic sentences by an introspective agent only if we drop the unrestricted, dynamic version of the Success Postulate. A weakened version of this postulate can be retained either by (a) restricting it to dynamic revision with simple, Boolean, non-doxastic sentences (as in the *AGM* literature, as well as in many of Segerberg's papers), or by (b) interpreting it in terms of static revision – i.e. as a conditional-belief statement $B^{\varphi}\varphi$.

6.2 Dynamic Doxastic Logic and Conditional Doxastic Logic/Dynamic Epistemic Logic

In this section we want to show that the *DDL* approach is at least as powerful as the *DEL* approach since it can internalize all the recent *DEL* developments. We first provide a complete axiomatization of static revision using the conditional belief logic (*CDL*) and then we develop three versions of *DDL* that internalize three of the revision operations considered in the Belief Revision literature.

6.2.1 Complete axiomatization of static revision: the logic *CDL*

To capture static revision, we follow the *DEL* tradition by borrowing from conditional logic a conditional belief operator $B^{\varphi}\psi$ [6]. So we add a conditional belief operator $B^{\varphi}\psi$ in our *DDL*-language (see Definition 6.1.1). Our semantic clauses

[8]While soundness of the given axiomatic system is not explicitly claimed in [78], its completeness is claimed. But from a conceptual point of view, a completeness result (with respect to a class of frames) is of course of no use if the axioms are not sound (with respect to that same class of frames).

can be naturally extended to this enlarged language.

Notation But first, following Segerberg [76], we introduce the notation

$$H \Cap P \ := \ \{X \in H : X \cap P \neq \varnothing\}$$

for all hypertheories $H \in D$ and sets $P \subseteq U$ of states.

Moreover we generalize to any families $F \subseteq H$ of fallbacks of a hypertheory H

$$F \Cap P \ := \ \{X \in F : X \cap P \neq \varnothing\}.$$

6.2.1. Definition. The *relativization of a family $F \subseteq H$ of fallbacks (of a hypertheory H) to a set $P \subseteq U$ of states* is the family

$$F^P \ := \ \{X \cap P : X \in F \Cap P\} = \{P \cap X : X \in F, P \cap X \neq \varnothing\}.$$

Of course, this operation can be applied in particular to an hypertheory H or onion O, producing a relativized hypertheory H^P or relativized onion O^P.

6.2.2. Definition. A family $F \subseteq H$ of fallbacks has the *finite intersection property relative to P* (P-f.i.p.) if every finite subfamily (of its relativization to P) $F' \subseteq F^P$ has non-empty intersection $\bigcap F' \neq \varnothing$. We say that a family $F \subseteq H$ of fallbacks has the *maximal P-f.i.p.* if F has the P-f.i.p. but no proper extension $F \subset G \subseteq H$ has the P-f.i.p.

Observe that, if O is an onion such that $P \cap \bigcup O \neq \varnothing$, then O has itself the maximal P-f.i.p.; and moreover O is the only family $F \subseteq O$ having the maximal P-f.i.p.

When $P = \|\varphi\|_H$ for some formula φ, we write "maximal φ-f.i.p." for "maximal $\|\varphi\|_H$-f.i.p." and so on.

Semantic clause for conditional belief Now we define conditional belief by putting:

$$s, H \vDash B^\theta \varphi \text{ iff } \forall F \subseteq H \text{ such that } F \text{ has the maximal } \theta\text{-f.i.p. } \exists F' \text{ finite} \subseteq F^{\|\theta\|_H}$$
$$\forall t \in \bigcap F' \, (t, H \vDash \varphi)$$

6.2.3. Proposition. *In onion models, φ is believed conditional on θ iff φ is true in all the plausible enough states satisfying θ:*

$$s, O \vDash B^\theta \varphi \quad \textit{iff} \quad \exists X \in O^{\|\theta\|_O} \, \forall t \in X \, (t, O \vDash \varphi).$$

Moreover, in onion models satisfying the Limit Condition, this boils down to the usual Grove semantics for static revision:

$$s, O \vDash B^\theta \varphi \ \textit{iff} \ \forall t \in \bigcap O^{\|\theta\|_O} \, (t, O \vDash \varphi).$$

We provide a detailed presentation of Conditional Doxastic Logic in Chapter 2.

6.2.4. DEFINITION. The language of Conditional Doxastic Logic $\mathcal{L}_{CDL}$ is the smallest set of formulas containing the atomic sentences $p \in \Phi$, the tautological formula $\top$ and is closed under conditional belief operators $B^{\theta}\varphi$. It can be considered as a variant of the *DDL*-language, in which there are no dynamic modalities, while B and K are defined as abbreviations by putting

$$B\varphi := B^{\top}\varphi,$$

$$K\varphi =: B^{\neg\varphi}\bot$$

(where $\bot := \neg\top$).

These abbreviations are semantically equivalent to the belief and knowledge operators, as defined in the previous section.

6.2.5. THEOREM. *The following proof system CDL for Conditional Doxastic Logic is sound and complete with respect to the class of all onion models, the class of standard onion models, and the class of finite onion models:*

Necessitation Rule:	From $\vdash \varphi$ infer $\vdash B^{\psi}\varphi$
Normality:	$\vdash B^{\theta}(\varphi \to \psi) \to (B^{\theta}\varphi \to B^{\theta}\psi)$
Truthfulness of Knowledge:	$\vdash K\varphi \to \varphi$
Persistence of Knowledge:	$\vdash K\varphi \to B^{\psi}\varphi$
Full Introspection:	$\vdash B^{\psi}\varphi \to KB^{\psi}\varphi$
	$\vdash \neg B^{\psi}\varphi \to K\neg B^{\psi}\varphi$
Hypotheses are (hypothetically) accepted:	$\vdash B^{\varphi}\varphi$
Superexpansion:	$\vdash B^{\varphi\wedge\psi}\theta \to B^{\varphi}(\psi \to \theta)$
Subexpansion (=Rational Monotonicity):	$\vdash (\neg B^{\varphi}\neg\psi \wedge B^{\varphi}(\psi \to \theta)) \to B^{\varphi\wedge\psi}\theta$

(where in all the above axioms, K is just the abbreviation $K\varphi := B^{\neg\varphi}\bot$).

Figure 6.1: Proof system *CDL*

6.2.6. FACT. In [6] Baltag and Smets show that the proof system *CDL* is sound and complete with respect to their conditional doxastic models (*CDM*) and prove this setting to be equivalent to an "epistemic" version of *AGM* theory. In the epistemic version of *AGM* theory, the Triviality Postulate ($T * \varphi = \bot$ iff $\vdash \neg\varphi$) is replaced with its epistemic version: $T * \varphi = \bot$ iff $T \vdash K\neg\varphi$. Indeed this is unavoidable in the presence of any irrevocable knowledge operator K: revising with a sentence whose negation is known should lead to a contradiction.

As a consequence, onion models satisfy all the postulates of the epistemic version of *AGM*.

6.2.7. Corollary. *If we take the initial AGM theory T to be the set $T = \{\psi : s, O \vDash_M B\psi\}$ of all beliefs held in a given ontic state s and a given onion O of an onion model M, and interpret the statically-revised theory $T * \varphi$ as the set $T * \varphi = \{\psi : s, O \vDash B^{\varphi}\psi\}$ of all conditional beliefs held conditional on φ in the same state s and same onion O of the same model M, then all the postulates of the epistemic AGM theory are satisfied.*

In contrast, static revision in general *DDL* models does not satisfy the epistemic *AGM* postulates since the Subexpansion principle fails in general *DDL* models. In conclusion, general *DDL* does not support an *AGM*-type theory of belief revision, but onion models are the natural *AGM*-friendly version of *DDL*.

6.2.2 Dynamic revision in *DDL*: internalizing doxastic upgrades

In *DEL*, epistemic/doxastic models are taken as basis and evolve under new information. One actually defines in a constructive way the new epistemic/doxastic model after a given doxastic action. This constructive approach can be internalized in *DDL* models: we will use such a constructive *DDL* approach to belief revision. We give constructive definitions of binary relations between onions, that internalize three different revision operations considered in the literature. Then we adopt from *DEL* the method of using Reduction/Recursion laws to give complete axiomatizations of the dynamic logics of these three kinds of revision. Indeed, our laws are identical to the ones considered in the *DEL* literature: this is a concrete example of how the *DEL*-style of modelling and axiomatizing belief revision can be "internalized" in *DDL*.

Different revision operations One can think of many ways to change the beliefs of an agent according to the information she receives. We provide a complete description of three of them in Chapter 2 and define the corresponding *DEL* operations of: update $!\varphi$, lexicographic upgrade $\Uparrow \varphi$ and conservative upgrade $\uparrow \varphi$.

Internalization doxastic upgrades Dynamic Epistemic Logic *DEL* (in its single-agent version) for the above-mentioned three types of upgrades can now be obtained as a special case of generalized *DDL*.

6.2.8. Definition. We reuse the relativized onion notation $O^P := \{P \cap X : X \in O, P \cap X \neq \varnothing\}$ introduced in Definition 6.2.1, to define *binary relations $R^{!P}$ (for update), $R^{\Uparrow P}$ (for lexicographic upgrade) and $R^{\uparrow P}$ (for conservative upgrade) between onions $O \in D$* of some onion model (U, D, R) and sets of sets of states $O' \subseteq \mathcal{P}(U)$, as follows:

$$(O, O') \in R^{!P} \text{ iff } O' = O^P \neq \varnothing$$

$$(O,O') \in R^{\Uparrow P} \text{ iff } O' = O^P \cup \{X \cup \bigcup O^P : X \in O\}$$
$$(O,O') \in R^{\uparrow P} \text{ iff } O' = \{\bigcap O^P : \bigcap O^P \neq \varnothing\} \cup \{X \cup \bigcap O^P : X \in O\}$$

Examples We provide three examples in Figures 6.3, 6.4, 6.5 one for each revision operation. The pictures drawn are following Hans Rott's presentation [72]. The spheres of the initial sphere system are drawn as usual, as nested circles. The dots represent the states. The numbers represent the spheres of the new sphere system after the revision. Thus all regions labelled with 1 form the first sphere of the new sphere system, the regions labelled with 2 form the second sphere and so on. Finally, the regions labelled with ω contain the states that are outside the union of all the spheres of the sphere system that is, the impossible states[9].

As an example of a belief revision scenario we introduce the following story. Consider an agent and a dice. Someone throws the dice such that the agent cannot see the upper face. We have 6 possible worlds in our sphere system: `i` where 1 is the upper face, `ii` where 2 is the upper face and so on. Assume that the agent initially believes that the upper face is 3 while in reality (unknown to our agent) the upper face is 4. Besides, the agent considers that it is more likely that the upper face is 5 than 1, and that it is more likely that the upper face is 1 than 6 while she considers that it is equally likely that the upper face is 1 or 2, and that it is equally likely that it is 4 or 6. Then according to the agent: $3 < 5 < 1 \equiv 2 < 6 \equiv 4$ ($3 < 5$ is read as it is more likely that the upper face is 3 than 5 and $1 \equiv 2$ is read as it is equally likely that the upper face is 1 or 2). In other words, she considers `iii` more plausible than `v`, `v` more plausible than both `i` and `ii`, and finally `i` and `ii` more plausible than both `iv` and `vi`. We represent the corresponding sphere system in Figure 6.2. The formula $\neg\varphi$ means "the number on the upper face is even".

In Figure 6.3 we consider an example of an update scenario. Indeed we consider the case where the agent receives a piece of hard information (coming from an infallible source) saying that the number on the upper face is even. So according to our definition for update, the agent will believe that the upper face is 2. Indeed, the agent considers `ii` to be more plausible after the update.

In Figure 6.4 we consider an example of a lexicographic upgrade scenario. Indeed we consider the case where the agent receives a piece of soft information (coming from a fallible but highly trustworthy source) saying that the number on the upper face is even. So according to our definition for lexicographic upgrade,

[9]Note that from Rott's perspective, these are *subjective* impossible states. In other words theses states are the least plausible worlds in the sphere system such that the actual word can be in ω. We adopt a different perspective since in our modelling $s \in \cup H$. Our impossible states in the regions labelled with ω are ontic states excluded by the doxastic state H.

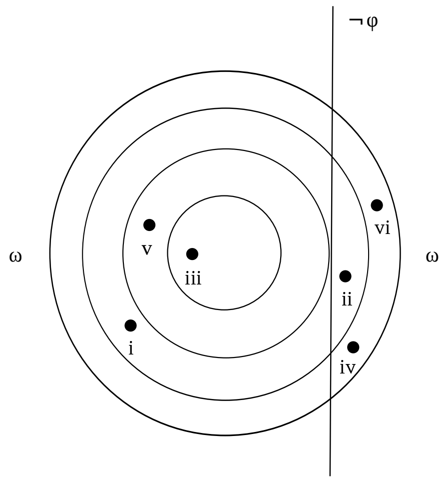

Figure 6.2: Initial sphere system

the agent will believe that the upper face is 2. Indeed, the agent considers `ii` to be more plausible after the lexicographic upgrade. If the agent still considers that it is more likely that the upper face is 3 rather than 5, and that it is more likely that the upper face is 5 rather than 1; she now also considers that both 4 and 6 are more likely than 3. According to the agent: $2 < 6 \equiv 4 < 3 < 5 < 1$. In other words, she considers `ii` more plausible than both `iv` and `vi`, `iv` and `vi` more plausible than `iii`, `iii` more plausible than `v` and finally `v` more plausible than `i`.

In Figure 6.5 we consider an example of a conservative upgrade scenario. Indeed we consider the case where the agent receives a piece of soft information (coming from a fallible and weakly trustworthy source) saying that the number on the upper face is even ($\neg\varphi$). So according to our definition for conservative upgrade, the agent will believe that the upper face is 2. Indeed, the agent considers `ii` to be more plausible after the conservative upgrade. If the agent still considers that it is more likely that the upper face is 3 rather than 5, more likely that the upper face is 5 rather than 1, and more likely that the upper face is 1 rather than 6 while she still considers that it is equally likely that the upper face is 4 or 6. According to the agent: $2 < 3 < 5 < 1 < 6 \equiv 4$. In other words, she considers `ii` more plausible than `iii`, `iii` more plausible than `v`, `v` more plausible than `i`, and finally `i` more plausible than both `iv` and `vi`.

6.2.9. Definition. We define a *DEL* onion model to be a standard onion model $M = (U, D, R)$ such that

$$R = \{R^{!P} : P \subseteq U\} \cup \{R^{\Uparrow P} : P \subseteq U\} \cup \{R^{\uparrow P} : P \subseteq U\}$$

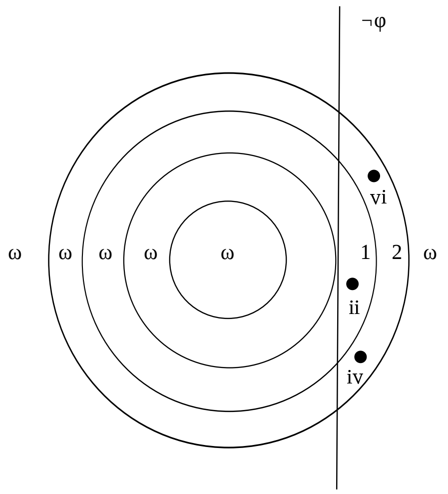

Figure 6.3: Example of an update $!\neg\varphi$

and such that D is closed under all the relations in R.

Semantic clauses The semantics is obtained by defining the interpretation maps $\|\varphi\|$ and $\|\alpha\|$ by double recursion: the static propositional clauses are as in *CDL*, the semantics of dynamic modalities is as in the generalized *DDL*, while the clauses for $\|\alpha\|$ are given by

$$\|!\varphi\| = R^{!\|\varphi\|}$$

$$\| \Uparrow \varphi\| = R^{\Uparrow\|\varphi\|}$$

$$\| \uparrow \varphi\| = R^{\uparrow\|\varphi\|}$$

6.2.10. Definition. The language of this version of *DEL* is obtained by adding to *CDL* dynamic modalities for all the above types of upgrades.

6.2.11. Theorem. *A sound and complete proof system for DEL onion models can be obtained by adding to the above proof system of CDL the van Benthem Reduction/Recursion laws [10]. We give here only the reduction laws for conditional belief:*

$$[!\varphi]B^{\psi}\theta \iff \varphi \to B^{\varphi\wedge[!\varphi]\psi}([!\varphi]\theta),$$

$$[\Uparrow\varphi]B^{\psi}\theta \iff B^{\varphi\wedge[\Uparrow\varphi]\psi}[\Uparrow\varphi]\theta \wedge \left(K^{\varphi}[\Uparrow\varphi]\neg\psi \to B^{[\Uparrow\varphi]\psi}[\Uparrow \varphi]\theta\right),$$

$$[\uparrow\varphi]B^{\psi}\theta \iff B^{\varphi}([\uparrow\varphi]\psi \to [\uparrow \varphi]\theta) \wedge \left(B^{\varphi}[\uparrow \varphi]\neg\psi \to B^{[\uparrow\varphi]\psi}[\uparrow \varphi]\theta\right),$$

where we used the abbreviation $K^{\varphi}\psi := K(\varphi \to \psi)$.

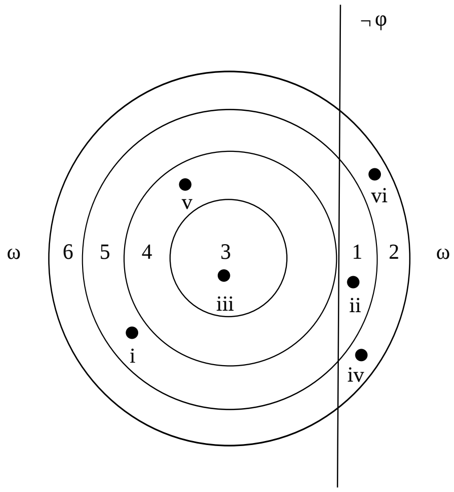

Figure 6.4: Example of a lexicographic upgrade $\Uparrow \neg\varphi$

Strongest Postcondition Modalities The standard dynamic modalities $[\alpha]\varphi$ are known in Computer Science as weakest preconditions. Indeed, they capture the weakest condition that can be imposed on an input information state (s, H) to ensure that, after performing action α in that state, φ will become true in the output-state. The dual modalities (in the sense of reversed modality) are the strongest postcondition modalities $\langle\alpha^{-1}\rangle\varphi$, capturing the weakest condition that is ensured to hold in an output-state after performing action α on an input state satisfying φ.

While standard *DEL* cannot represent strongest postconditions[10], *DDL* models contain enough information to define them, as existential (Diamond) modalities for the converse relations R_α^{-1}: equivalently, just put

$$s, H \vDash \langle\alpha^{-1}\rangle\varphi \quad \text{iff} \quad \exists H' \left((H', H) \in \|\alpha\|_H \wedge s, H' \vDash \varphi\right)$$

It is obvious that these operators are the reversed dynamic modalities, and that the same holds for their corresponding de Morgan duals: i.e. we have the validities

$$\varphi \Rightarrow [\alpha]\langle\alpha^{-1}\rangle\varphi,$$
$$\varphi \Rightarrow [\alpha^{-1}]\langle\alpha\rangle\varphi.$$

Finally, using the strongest postcondition modality for lexicographic upgrade, we can check the semantic equivalence:

$$B^\varphi\psi \Longleftrightarrow [\Uparrow \varphi]B\langle(\Uparrow \varphi)^{-1}\rangle\psi.$$

[10]But extensions of *DEL* which can define strongest postconditions have been proposed by Guillaume Aucher [2].

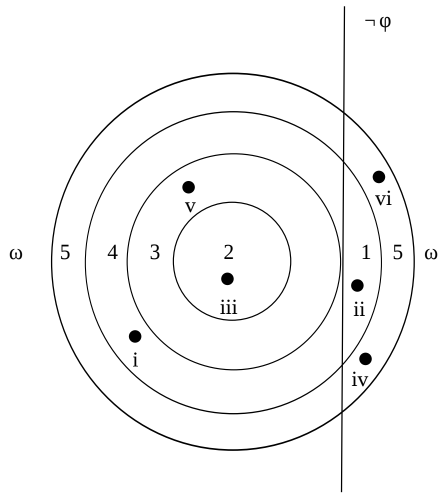

Figure 6.5: Example of a conservative upgrade $\uparrow \neg\varphi$

This equivalence confirms our interpretation of conditional beliefs $B^{\varphi}\psi$ as embodiments of static revision: the agent's revised beliefs after revision with φ about ψ's truth value before the revision.

6.3 Expansion and contraction in full *DDL*

In [76], Segerberg uses a constructive approach (similar to the one we used above for revision) for modelling expansion and contraction in *DDL*, assuming some additional conditions on the hypertheories. However these two operations do not fit the *AGM* framework. In this section we introduce and axiomatize three *AGM*-friendly versions of contraction and expansion in *DDL*.

6.3.1 Main definitions for contraction and expansion

First, Segerberg assumes some additional conditions on the hypertheories namely, that they are closed under non-empty intersections and satisfy the Strong Limit Assumption. He calls *LR* hypertheories (from Lindstrom and Rabinowicz), the hypertheories that satisfy these two conditions. Then he defines H/P, the augmentation of H by P and $H|X$, the restriction of H by X.

6.3.1. Definition. Segerberg puts, for *LR* hypertheories H and sets $P \subseteq U, X \in H$:

$$H/P := H \cup \{X \cap P : X \in H, X \cap P \neq \varnothing\},$$

$$H|X := \{Y \in H : X \subseteq Y\},$$

and requires the doxology D to be closed under these operations.

Minimal fallback Using these notations, Segerberg says that a fallback $Z \in H$ is a contraction with $P \subseteq U$ in H iff Z is a minimal fallback (with respect to inclusion) in the family $H \Cap (U - P)$ where $H \Cap (U - P) := \{X \in H : X \cap (U - P) \neq \varnothing\}$. Note that such a contraction with P in H might not exist[11], and even if it exists it might not be unique.

6.3.2. DEFINITION. Segerberg then explicitly defines an *expansion action* $+P$ and a *contraction action* $-P$ (for any given set $P \subseteq U$ of states), given by the following relation on hypertheories in D:

$$(H, H') \in R^{+P} \text{ iff } H' = H/P,$$

$$(H, H') \in R^{-P} \text{ iff } H' = H|Z \text{ for some contraction } Z \text{ with } P \text{ in } H.$$

However, these two operations do not fit the AGM framework. This was a conscious decision by Segerberg, since his aim in [76] was to give a semantics to the Lindstrom-Rabinowicz theory of contraction rather than for the AGM theory.

AGM **adaptation** In order to try to accommodate AGM, first we have to restrict the above definitions to onion models. As we saw, these are the AGM-friendly models for *DDL*. On onion models, contractions with P (as defined above) might still not exist but if they do, then they are unique as required by AGM. To ensure existence, we have to further restrict to onion models satisfying the Limit condition or (for simplicity) to the even more restricted case of standard onion models. As we will see, this restriction does ensure that Segerberg's contraction satisfies the AGM principles.

Problems with expansion But even in this case, we still have problems with Segerberg's definition of expansion. This operation does not preserve the "nestedness" property, so it does not map standard onions into onions. Moreover, there is no reasonable additional condition that would ensure that the expansion of an onion O with a set P in the sense of Segerberg, is an onion whenever $P \cap \bigcup O \neq \varnothing$. Since "onionhood" (i.e. nestedness of the hypertheories) is essential for satisfying AGM postulates, this means that one should look for a different definition for AGM expansion.

[11] Though the additional closure assumptions made by Segerberg in [76] do ensure the existence of contractions.

6.3.2 Expansion

In fact, any of the known semantic proposals for expansion (as an operation on Grove sphere models) considered in the Belief Revision literature can be internalized in *DDL*. In particular, for each of the three types of revision defined above there is a corresponding expansion action on standard onion models:

$$(O,O') \in R^{+!P} \text{ iff } (O,O') \in R^{!P} \text{ and } X \cap P \neq \varnothing \text{ for all } X \in O,$$
$$(O,O') \in R^{+\Uparrow P} \text{ iff } (O,O') \in R^{\Uparrow P} \text{ and } X \cap P \neq \varnothing \text{ for all } X \in O,$$
$$(O,O') \in R^{+\uparrow P} \text{ iff } (O,O') \in R^{\uparrow P} \text{ and } X \cap P \neq \varnothing \text{ for all } X \in O.$$

Since expansion is a special case of revision (namely the case in which the new information does not contradict any prior beliefs), the corresponding expansion modalities can be reduced to the revision ones, e.g.

$$[+!\varphi]\theta \iff (\neg B\neg\varphi \to [!\varphi]\theta),$$
$$[+\Uparrow\varphi]\theta \iff (\neg B\neg\varphi \to [\Uparrow\varphi]\theta),$$
$$[+\uparrow\varphi]\theta \iff (\neg B\neg\varphi \to [\uparrow\varphi]\theta).$$

6.3.3 Contraction

On standard onion models, contractions with P exist and are unique whenever P is not irrevocably known (i.e. whenever $(\bigcup O) \cap (U - P) \neq \varnothing$). Moreover, on standard onion models Segerberg's definition is equivalent to putting:

$$(O,O') \in R^{-P} \text{ iff } O' = O \Cap (U - P) := \{X \in O : X \cap (U - P) \neq \varnothing\}.$$

This semantic contraction operation is the operation of severe withdrawal we introduced in Chapter 3.

Axiom system for the logic of severe withdrawal We refer back to Chapter 3 for the definition of the language for the logic of severe withdrawal as well as for the proof of Theorem 6.3.3.

6.3.3. Theorem. *A sound and complete proof system for the logic SEV with the language $\mathcal{L}_{Sev2}$ over the class of onion models in DDL is given by the axioms and rules of $\mathcal{L}_{KK_D}$ plus the following reduction axioms:*

$$[-\varphi]p \iff p$$
$$[-\varphi]\neg\theta \iff \neg[-\varphi]\theta$$
$$[-\varphi](\theta \wedge \psi) \iff [-\varphi]\theta \wedge [-\varphi]\psi$$
$$[-\varphi]K\theta \iff K[-\varphi]\theta$$
$$[-\varphi]K_D\theta \iff (K_D[-\varphi]\theta \wedge (\neg K\varphi \to \neg K\neg(\neg\varphi \wedge K_D[-\varphi]\theta))$$

We cannot provide Reduction Laws in the language $\mathcal{L}_{Sev1}$ that is, if the static base language is the language of Conditional Doxastic Logic.

Objections against severe withdrawal We already mentioned some objections against severe withdrawal in Chapter 3. To these objections, we can add another one based on dynamic logic. Namely, although severe withdrawal satisfies a dynamic version of the so-called Levi identity with respect to irrevocable revision (*DEL* update)

$$R^{-\neg P}; R^{+!P} = R^{!P}$$

(where $R; R'$ is relational composition and $P \subseteq U$ is an arbitrary set of states), the corresponding Levi identities for lexicographic revision or minimal revision are not satisfied:

$$R^{-\neg P}; R^{+\Uparrow P} \neq R^{\Uparrow P},$$

$$R^{-\neg P}; R^{+\uparrow P} \neq R^{\uparrow P}.$$

Since update (irrevocable revision) is a rather implausible operation when dealing with belief change in daily life, this throws more doubt on the appropriateness of Segerberg's definition of contraction.

6.3.4 Other *AGM*-type contractions

We also introduced conservative contraction $-_c P$ and moderate contraction $-_m P$ in Chapter 3.

6.3.4. Definition. We give the formal definitions over onion models in *DDL*. First we put $O_{-P} := \bigcap O^{U-P}$ for the smallest non-empty intersection of an O-sphere with $U - P$ whenever $O^{U-P} \neq \varnothing$ (i.e. whenever $\bigcup O \nsubseteq P$), and $O_{-P} := \varnothing$ otherwise.

6.3.5. Definition. Then for any two standard onions $O, O' \in D$ we define

$$(O, O') \in R^{-_c P} \text{ iff } O' = \{X \cup O_{-P} : X \in O\},$$

$$(O, O') \in R^{-_m P} \text{ iff } O' = \{Y \cup \bigcap O : Y \in O^{U-P}\} \cup \{X \cup \bigcup O^{U-P} : X \in O\}.$$

Axiom system for the logic of conservative contraction We refer back to Chapter 3 for the definition of the language for the logic of conservative contraction as well as for the proofs of Theorem 6.3.6 and Theorem 6.3.7.

6.3.6. Theorem. *A sound and complete proof system for the logic CONS$_1$ with the language $\mathcal{L}_{Cons1}$ over the class of onion models in DDL is given by the axioms and rules of $\mathcal{L}_{CDL}$ plus the following reduction axioms:*

$$
\begin{array}{lcl}
[-_c\varphi]p & \iff & p \\
[-_c\varphi]\neg\theta & \iff & \neg[-_c\varphi]\theta \\
[-_c\varphi](\theta\wedge\psi) & \iff & [-_c\varphi]\theta\wedge[-_c\varphi]\psi \\
[-_c\varphi]B^{\psi}\theta & \iff & B([-_c\varphi]\psi\to[-_c\varphi]\theta)\wedge B^{\neg\varphi}([-_c\varphi]\psi\to[-_c\varphi]\theta) \\
 & & \wedge\big(B^{\neg\varphi}[-_c\varphi]\neg\psi\to B^{[-_c\varphi]\psi}[-_c\varphi]\theta\big)
\end{array}
$$

6.3.7. THEOREM. *A sound and complete proof system for the logic $CONS_2$ with the language $\mathcal{L}_{Cons2}$ over the class of onion models in DDL is given by the axioms and rules of $\mathcal{L}_{KK_D}$ plus the following reduction axioms:*

$$
\begin{array}{lcl}
[-_c\varphi]p & \iff & p \\
[-_c\varphi]\neg\theta & \iff & \neg[-_c\varphi]\theta \\
[-_c\varphi](\theta\wedge\psi) & \iff & [-_c\varphi]\theta\wedge[-_c\varphi]\psi \\
[-_c\varphi]K\psi & \iff & K[-_c\varphi]\psi \\
[-_c\varphi]K_D\psi & \iff & K_D[-_c\varphi]\psi\wedge B^{\neg\varphi}[-_c\varphi]\psi
\end{array}
$$

Axiom system for the logic of moderate contraction We refer back to Chapter 3 for the definition of the language for the logic of moderate contraction as well as for the proof of Theorem 6.3.8.

6.3.8. THEOREM. *A sound and complete proof system for the logic MOD_1 with the language $\mathcal{L}_{Mod1}$ over the class of onion models in DDL is given by the axioms and rules of $\mathcal{L}_{CDL}$ plus the following reduction axioms:*

$$
\begin{array}{lcl}
[-_m\varphi]p & \iff & p \\
[-_m\varphi]\neg\theta & \iff & \neg[-_m\varphi]\theta \\
[-_m\varphi](\theta\wedge\psi) & \iff & [-_m\varphi]\theta\wedge[-_m\varphi]\psi \\
[-_m\varphi]B^{\psi}\theta & \iff & B([-_m\varphi]\psi\to[-_m\varphi]\theta)\wedge B^{\neg\varphi\wedge[-_m\varphi]\psi}[-_m\varphi]\theta \\
 & & \wedge\big(K^{\neg\varphi}[-_m\varphi]\neg\psi\to B^{[-_m\varphi]\psi}[-_m\varphi]\theta\big)
\end{array}
$$

We cannot provide an axiom system for the logic MOD_2 with the language $\mathcal{L}_{Mod2}$. Indeed we do not have a reduction axiom for defeasible knowledge (see Chapter 3 for a more detailed explanation).

Advantages Conservative contraction and moderate contraction have the advantage to satisfy the dynamic versions of Levi identity for all the above-mentioned revision operators: for all sets $P \subseteq U$ of states, we have

$$R^{-_c \neg P}; R^{+!P} = R^{!P}, \qquad R^{-_m \neg P}; R^{+!P} = R^{!P},$$
$$R^{-_c \neg P}; R^{+\Uparrow P} = R^{\Uparrow P}, \qquad R^{-_m \neg P}; R^{+\Uparrow P} = R^{\Uparrow P},$$
$$R^{-_c \neg P}; R^{+\uparrow P} = R^{\uparrow P}, \qquad R^{-_m \neg P}; R^{+\uparrow P} = R^{\uparrow P}.$$

Conclusion

The *DEL* approach and the *DDL* approach are two different styles of modelling doxastic changes. In the *DEL* approach, the dynamics is external to the models. Doxastic actions are seen as model-changing actions, and represented as relations between models. Segerberg's *DDL* keeps the actual states unchanged (as ontic states) and internalizes the dynamics by representing doxastic actions as binary relations between doxastic structures (onions, hypertheories, doxastic states) living in a fixed space of possible such structures (the doxology). Again, if considered at an appropriate level of generality, these approaches are equivalent. However, there are some conceptual (and practical) differences.

The *DEL* approach is the most "open-ended", well-suited for open systems, in which there are innumerable doxastic actions that might happen. It is also the most "economical", as only the states and the doxastic structures that are currently epistemically possible are "given". Only they are represented in a given model hence, the *DEL* models can be easily visualized and drawn. It is also a "constructive" approach: the doxastic dynamics is not given in this approach but is to be constructed (in the form of various model transformers, or upgrades).

The *DDL* style keeps the states fixed and only multiplies the doxastic structure. It also brings conceptual clarity: doxastic changes are after all only changes of belief, so they shouldn't multiply the states of the world. It is an elegant and natural way to internalize doxastic changes. As shown in this chapter, it is potentially at least as expressive and powerful as the single-agent version of the *DEL* approach: all work on belief revision done in *DEL* style can be done in *DDL* style.

In the next chapter, we would like to investigate the belief dynamics in a dynamic setting namely the dialogical setting. We first present the branching-time belief revision logic $\mathcal{L}_{PLS*}$ of G. Bonanno and provide an argumentative study of this belief revision logic using the dialogical approach to logic.

Chapter 7

Bonanno's belief revision logic in a dialogical setting

Aim: In this chapter our aim is to provide an argumentative study of belief revision logic. In particular we focus on the branching-time belief revision logic of Bonanno L_{PLS*} as introduced in Definition 2.8.15. To fulfil our purpose we provide a dialogical approach to this logic L_{PLS*}.

Summary: In this chapter we motivate our choice to provide a dialogical approach to L_{PLS*} and we precisely define this dialogical setting. We provide our dialogical approach to L_{PLS*} as well as its soundness and completeness proof, establishing a formal relation with its model-theoretic approach. The main points are :

- we motivate the investigation of L_{PLS*} in an argumentative setting. We provide the main characteristics of the dialogical setting we use as well as its historical background.

- we precisely define some important dialogical notions we use in this chapter.

- we provide a dialogical approach to the logic L_{PLS*} providing the language and the rules. We apply our dialogical system to specific examples to illustrate its mechanisms.

- we prove that our dialogical approach to Bonanno's logic of belief revision is sound and complete with respect to L_{PLS*} showing that there exists a winning strategy for the Proponent in a dialogue with a thesis Δ if and only if Δ is a valid formula in L_{PLS*}.

Background: The dialogical approach to logic is a two-person game in which one party defends a proposition while the other party challenges it. The game is defined by a set of rules. Some rules stipulate how the logical constants can/have to be challenged and defended and some other rules define the process of the game itself (which player starts, can move, wins...). Two notions are fundamental in the dialogical approach to logic namely, the notion of choice which leads to the notion of strategy. Indeed players make choices among the available moves allowed by the rules. If a player wins, whatever the choices of the other player, he has a winning strategy in the corresponding dialogical game.

We have to underline that dialogical games are not played on a given model. There is no model in the dialogical approach to logic. The concept of validity has its counterpart in the concept of winning strategy. If the player defending the proposition has a winning strategy then this proposition is considered to be valid.

Paul Lorenzen and Kuno Lorenz were first to introduce the concept of formal dialogues.[1] This first dialogical approach was concerned with intuitionistic and classical logic [55]. Later on, the dialogical framework has been developed and applied to non-classical logics by one of their students Shahid Rahman [66]. In particular Shahid Rahman and Helge Rückert developed the first modal dialogues [69, 68, 45]. They introduce new sets of rules in relation with modal operators. Very recently, a dialogical approach to Dynamic Epistemic Logic has been developed [56] by Sébastien Magnier. He offers an argumentative study of Dynamic Epistemic Logic focusing mainly on Public Announcement Logic. New rules are given not only for epistemic operators but also for public announcement operators. We believe it is now time to extend the dialogical approach to belief revision logic in order to provide the first argumentative study of belief revision logic.[2] In this thesis, we choose to study the branching-time belief revision logic of Bonanno L_{PLS*} in a dialogical setting (see Chapter 2).

[1] In the spirit of our work on game semantics in this thesis in Chapter 5, let us mention that the dialogical logic founded by Lorenzen and Lorenz and the game theoretical semantics of Hintikka [42], have been compared with each other. For this comparison we refer to the work of [70] and for an analysis of the importance of the differences between these game styles in the context of the philosophical (anti)-realism debate, we refer to [58].

[2] Note that [67] provides a dialogical approach to the first logic of belief revision introduced by Bonanno in [17] (which is actually a belief expansion logic rather than a belief revision logic). We would like to emphazise that this first version of Bonanno's belief revision is very different from L_{PLS*}.

7.1 General background on the dialogical approach to logic

After motivating our interest in the argumentative study of logic, we precisely define the dialogical framework we use in concrete terms.

7.1.1 Motivation

We first have to distinguish between two different notions of dynamics namely, the dynamics of an argumentative practice and the dynamics of semantics (through dynamic modal operators). We are taking up an idea put forward by Magnier [56] where he establishes two distinct levels of dynamical changes:

- A logic is called “dynamic” because of its object language. Some operations introduced in the logical language force its dynamic nature.

- A logic becomes “dynamic” because it is implemented in an argumentative context.

So Magnier distinguishes between a dynamic logical language and a dynamic practice of logic. Magnier names the first type of dynamics “internal dynamics” and the second one “external dynamics”. The argumentative practice of logic is dynamic in an external sense because the argumentative process is dynamic in itself but not necessarily the language of the logic investigated through this argumentative practice. Thus the argumentative practice is grounded in an irreducible form of dynamics. Contrary to some dynamic modal languages such as the language of dynamic epistemic logics that can be rewritten into static epistemic languages by means of appropriate reduction axioms (see the reduction axioms in [29]). Indeed the static language pre-encodes the dynamics.

We are interested in investigating the relation between beliefs and information over time from an external dynamic perspective. We claim that this external dynamic perspective will shed some new light on this relation. In particular we will interpret this relation through the notion of choice. In an argumentative process, players challenge each other's claims. Thus players have to make choices. Implementing a belief revision logic inside an argumentative framework will produce choices about belief and information. Moreover we see now that the internal dynamics of the information operator will become more apparent in the light of this external dynamic point of view. The goal is to explore what we can learn about Bonanno's logic of belief revision L_{PLS*} from an argumentative framework. But we also carefully examine what the interpretation of the information operator through the argumentative notion of choice provides to the argumentative framework itself. Indeed the non normality of Bonanno's information operator will bring some new interesting developments in this framework.

7.1.2 Dialogues

There exist different types of dialogues. Our aim is not to present these dialogues in details, we only mention that they have different purposes and thus different sets of rules[3]. Which type of dialogue would be suitable to investigate L_{PLS*} through an argumentative practice? We have to define the main characteristics of the type of dialogues we are looking for to answer this question. First we list what kind of argumentative practice we do not want. We do not want :

– exchanges of unrelated statements,

– exchanges linked to some particular background,

– arguments to be about players (challenging players as personal attacks),

– inequality between players as different inference rules,

– infinite argumentative processes (without a winning player),

– possible or plausible conclusions (no certain conclusions).

Since we want to reconstruct L_{PLS*} into an argumentative framework, we are only interested in formal dialogues. In formal dialogues, the players are objective and impartial. They can only challenge the arguments of the other player (and not the other player himself), and they have to cooperate even if they are engaged in a competitive argumentative process because they comply with the rules. Finally we look for dialogues regulated by symmetric rules, providing them with an objective aspect. Since we want to establish a counterpart of the notion of validity of a formula, namely the notion of winning strategy, we require finite dialogues providing at least one and only one winning player. This notion of winning strategy involves the notion of competitive players. The dialogical approach to logic meets all these criteria, that's why we choose to provide a dialogical approach to L_{PLS*}.

7.1.3 The dialogical approach to logic

The dialogical approach to logic was first introduced by Lorenzen in the 1950's and then developed by Lorenz for classical and intuitionistic logic[4]. Rahman, one of Lorenz's students, has further developed the dialogical approach to logic to allow for the development and the combination of different logics in this framework (free logic, normal modal logic, non normal modal logic and so on) [66, 44]. The aim was to propose a semantics based on argumentation games as a new alternative to model theory and proof theory: the dialogical approach to logic is neither model theory nor proof theory. The main concept of this approach is "meaning as use" namely use in an argumentative process.

[3]For a complete taxonomy see [82].

[4]The most important early papers on the dialogical approach to logic are collected in [55].

Dialogical game In a dialogical game two players confront each other. The Proponent proposes a thesis that he will defend against the challenges of the Opponent who aims to find a counter-argument for it. So the game starts with the Proponent (**P**) stating a formula from some given language $\mathcal{L}$. Then the Opponent (**O**) challenges the formula and the Proponent defends: they interact by alternately choosing moves according to some rules. The notion of choice plays an essential role in the dialogical approach to logic. We will further develop this point when we will present the dialogical rules. Thus some rules are needed in order to define how players can challenge/defend logical constants but also to define when players can make a move, what kind of move they are allowed to make, when the game ends, which player wins and so on.

Dialogical rules The dialogue is a game which obeys two kinds of rules: particle rules and structural rules. Particle rules constitute the local semantics of a logic: it determines the meaning of each logical constant in terms of use (how players can/have to use them) in an argumentative process. Thus particle rules define the way in which connectives are played. These rules are symmetric that is, they are necessarily the same for the Opponent as well as the Proponent. As a consequence, the logical meaning of a given constant is independent of the players. That's why we use **X** and **Y** as variables ranging on $\{\mathbf{O}, \mathbf{P}\}$, always assuming that $\mathbf{X} \neq \mathbf{Y}$ in their definition. This symmetry of the particle rules provides an objective aspect to dialogues. Structural rules determine the global semantics of a logic: they define the way in which the dialogue proceeds.

Dialogical language A dialogical language for propositional logic is obtained from the standard propositional language by the addition of one metalogical symbol "?" standing for "challenge", and two labels **O** and **P**, standing for the players (Opponent, Proponent) of the dialogue.

Dialogical approach to modal logic Modal dialogues are developed by Rahman and Rückert [69, 68, 45]. While dialogical propositional logic investigates the meaning in terms of use in an argumentative process, dialogical modal logic contextualises this meaning in terms of use. Thus the meaning of a logical constant depends on its contextual use. Technically, dialogical modal logic needs the introduction of contextual points allowing to specify the contextual nature of the moves, i.e. the context in which the moves are made. They are the counterpart of possible worlds in the model-theoretic approach.

7.1.1. Definition. A *contextual point* is a positif integer i indexing a statement in a dialogue.

Players can choose new contextual points when challenging a modal operator according to the rules. Indeed a particle rule must be added for each modal operator to define the way in which it is played and some structural rules must be added to define what contextual points can be chosen by the players to challenge the corresponding modal operator. In that case, structural rules define the conditions of particular use of particle rules for modal operators. They can be interpreted as the counterpart of conditions imposed on frames in the model-theoretic approach. A contextual point is new if and only if it is chosen by a player in some move and there is no previous move in the same game where the contextual point is chosen. The play (challenge/defence) on a modal operator creates a chain of contextual points.

7.1.2. Definition. If a player successively challenges modal operators from the contextual point i choosing successively the contextual points j, k, $\ldots$, n, then $i.j.k....n$ is a *chain of contextual points*.

A chain of contextual points reflects the choices the players made. Thus $i.j.k$ means that the contextual point k has been chosen from the contextual point j to challenge a modal operator such that this contextual point j has been itself chosen to challenge a modal operator from the contextual point i.

7.2 Original dialogical approach to Bonanno's logic for belief revision

In this section we provide a dialogical approach to L_{PLS*} providing some important definitions and the corresponding language, particle and structural rules, based on [34] and [33]. Next, we illustrate our dialogical system for L_{PLS*} through some concrete examples. We name this dialogical approach *Dialogical Temporal Doxastic Logic* (*DTDL*).

7.2.1 *DTDL* Framework

We first define the language of *DTDL*.

7.2.1. Definition. The language of *DTDL* $\mathcal{L}_{DTDL}$ is obtained from the language of L_{PLS*} by the addition of:

- the symbols **O** and **P**,
- the symbol for challenge "?",
- two new symbols "!" and "?*" respectively for request and confirmation.

The new symbols are introduced in relation to the non normality of the information operator.

Since the language of $DTDL$ is a multimodal language, we need two different types of contextual points. Indeed we can distinguish between two types of modal operators: temporal operators ($\bigcirc^{-1}$ and $\bigcirc$) and non temporal operators (B, I, A). Then we use the contextual points i and t such that moves are made in a context (i, t). In that case, contextual points i are the counterpart of possible worlds s while t is the counterpart of instants t in the model-theoretic approach.

7.2.2. Definition. A *move* is a tuple $\langle \mathbf{X} - i, t : e \rangle$ where:

- $\mathbf{X} \in \{\mathbf{O}, \mathbf{P}\}$,
- i and t are contextual points that is, positive integers or sequences of positive integers such that (i, t) is a context,
- e is a statement of the language of $DTDL$

We now have to draw a sharp distinction between some dialogical terms.

7.2.3. Definition. We define the notions of dialogue, play, close play and terminal play:

- a *dialogical game* or *dialogue* $\mathcal{D}_\Delta$ is the set of all the possible plays for a formula Δ,
- a *play* d_Δ is a sequence of moves allowed by the rules. This sequence starts with a move $\langle \mathbf{P} - 0, 0 : \Delta \rangle$,
- a play d_Δ is *close* if and only if it contains two moves such that $\langle \mathbf{O} - i, z : p \rangle$ and $\langle P - i, t : p \rangle$[5]; $\langle \mathbf{O} - i, t : ?I_j \rangle$ and $\langle P - i, t : ?I_j^* \rangle$; $\langle \mathbf{O} - i, t : ?B_j \rangle$ and $\langle P - i, t : ?I_j^* \rangle$,
- a play d_Δ is *terminal* if there are no more moves allowed by the rules.

7.2.2 Particle rules

Particle rules define the way in which logical constants are played that is, how they should be challenged and defended. These rules are strictly the same for the Opponent and the Proponent. In other words, what matters is how a logical constant can be used regardless of the player who uses it. Thus the meaning of logical constants is given independently of the role of players (Opponent/Proponent).

[5] It is possible that $z \notin t$.

Vocabulary First we have to distinguish the dialogical terms *challenge*, *request* and *confirmation*. A player challenges a statement of the other player. A request is always about a contextual choice. A player requests the other player to confirm that he could choose a particular contextual point to challenge a modal operator in a given context. A player confirms that he could choose the required contextual point to challenge a modal operator in a given context.

Reading particle rules A particle rules involves three steps:

- **X** utters a formula,
- **Y** challenges this formula,
- **X** defends the formula.

Particle rules for standard connectives We first provide particle rules for the standard connectives in Figure 7.1.

Standard connectives	**X** Utterance	**Y** Challenge	**X** Defence
$\neg$, there is no possible defence	i,t: $\neg\ \varphi$	i,t: φ	$\otimes$
$\wedge$, the challenger chooses a conjunct	i,t: $\varphi \wedge \psi$	i,t: $?_{\wedge 1}$ or i,t: $?_{\wedge 2}$	i,t: φ respectively i,t: ψ
$\vee$, the defender chooses a disjunct	i,t: $\varphi \vee \psi$	i,t: $?_{\vee}$	i,t: φ or i,t: ψ

Figure 7.1: Particle rules for standard connectives

When **X** utters the negation of a formula, **Y** challenges by uttering the formula. There is no corresponding defence – denoted in a dialogue by the symbol $\otimes$. When **X** utters a conjunction, **Y** chooses the conjunct **X** has to defend; while when **X** utters a disjunction, **X** chooses the disjunct he wants to defend.

Particle rules for modal operators Now we provide particle rules for modal operators in Figure 7.2. The contextual points play an essential role here. Indeed they become paramount when modal operators come into the language. For the sake of clarity, we always explicitly state the modal operator challenged in the challenge itself. Thus a challenge of a belief operator looks like " $?B_j$" and a challenge of an information operator looks like " $?I_j$". The same applies to request and confirmation.

But first we have to clarify a fundamental distinction between the players of the dialogue and an agent. The players of the dialogue are the Proponent and the Opponent. These players can discuss about the beliefs of an agent and/or the information received by an agent but they are not in any case the agent in question. The Proponent and the Opponent discuss about a third person, not about their own beliefs/information.

Modal operators	**X** Utterance	**Y** Challenge	**X** Defence
$\bigcirc^{-1}$, the challenger chooses a contextual point u	i,t: $\bigcirc^{-1}\varphi$	i,t: $?\bigcirc^{-1}_{u}$	$i,t.u$: φ
$\bigcirc$, the challenger chooses a contextual point u	i,t: $\bigcirc\varphi$	i,t: $?\bigcirc_{u}$	$i,t.u$: φ
B, the challenger chooses a contextual point j	i,t: $B\varphi$	i,t: $?B_j$	$i.j,t$: φ
A, the challenger chooses a contextual point j	i,t: $A\varphi$	i,t: $?A_j$	$i.j,t$: φ
I, the challenger has the choice between two challenges	i,t: $I\varphi$	i,t $?I_j$	$i.j,t$: φ
		i,t: $!I_j$	i,t: $?I^{*}_{j}$

Figure 7.2: Particle rules for modal operators

When **X** utters a formula of the form $\bigcirc^{-1}\varphi$ in (i,t), he must be able to defend φ in any contextual point u chosen by **Y** to challenge the $\bigcirc^{-1}$ operator: in that case the context $(i,t.u)$ is called an immediate past context of the context (i,t). Indeed if a player **X** states that at the previous instant it was the case that φ, he is committed to defend φ in all immediate past contexts.

When **X** utters a formula of the form $\bigcirc\varphi$ in (i,t), he must be able to defend φ in any contextual point u chosen by **Y** to challenge the $\bigcirc$ operator: in that case the context $(i,t.u)$ is called an immediate future context of the context (i,t). Indeed if a player **X** states that at every next instant it will be the case that φ, he is committed to defend φ in all immediate future contexts.

When **X** utters a formula of the form $B\varphi$ in (i,t), he must be able to defend φ in any contextual point j chosen by **Y** to challenge the B operator. Indeed if a player **X** states that an agent believes a proposition φ, he is committed to defend φ in all contexts that this agent conceives.

When **X** utters a formula of the form $A\varphi$ in (i,t) , he must be able to defend φ in any contextual point j chosen by **Y** to challenge the A operator. Indeed if a player **X** states that it is always the case that φ, he is committed to defend φ in all contexts.

When **X** utters a formula of the form $I\varphi$ in (i,t), **Y** has the choice between two different challenges. He can choose the standard challenge: he challenges choosing a contextual point j, and then **X** must be able to defend φ in any contextual point j chosen by **Y**. Or he can choose the non-standard challenge: he chooses a contextual point j and requests **X** to confirm that this contextual point j could be chosen to challenge an information operator at (i,t), and then **X** must be able to confirm that the contextual point j chosen by **Y** can be chosen at (i,t) to challenge an information operator. Indeed if a player **X** states that an agent is informed about a proposition φ, he is committed to defend φ in all contexts of which the agent is informed and he is committed to defend that all contexts where φ holds are contexts of which the agent is informed.

The information operator is a non normal operator, that's why its particle rule is far from being standard. The non standard challenge $!I_j$ at (i,t) can be read as "show me that the contextual point j can be chosen to challenge the I operator at (i,t)". Then the corresponding defence $?I_j^*$ at (i,t) is the confirmation that indeed this contextual point j can be chosen to challenge the I operator at (i,t): "I confirm that the contextual point j can be chosen to challenge the I operator at (i,t)". This is new in the dialogical approach to logic: players no longer deal with formulas but with choices. The non normality of the I operator introduces directly the notion of choice inside the dialogue. Players discuss about their own choices, more precisely about the choices they can make. We will further develop this point in the examples (see Section 7.2.4).

Particle rules of *DTDL* The particle rules of *DTDL* consist of the particle rules for standard connectives and for modal operators. These rules determine all the possible uses of the logical constants of the language of *DTDL*. Now we have to define the structural rules of *DTDL* to determine the conditions under which some particle rules of *DTDL* can be used and how the game is played.

7.2.3 Structural rules

Structural rules regulate the process of the dialogue. We first introduce the structural rules defining how the play starts and ends, how the players can play, as well as the winning rule[6].

- ⋄ **(SR-0) (Starting rule):** Any play d_Δ of a dialogue $\mathcal{D}_\Delta$ starts with **P** uttering the thesis in an initial context (i,t). The moves of a play are numbered such that the thesis has number 0. Then **O** and **P** respectively choose a natural number n and m allowing a number of repetitions (called repetition rank). **O** and **P** can repeat the same move (challenge or defence) respectively n

[6] We provide some schemas for the reader who is not familiar with structural rules in Appendix A. We strongly recommend to first read the rules with the help of the schemas and then to read the explanations of the rules.

and m times[7].

- **(SR-1) (Game-playing rule):** Moves are made alternately by **O** and **P** according to the other rules. In any move each player may challenge any complex formula uttered by the other player, or he may defend himself against any challenge, including those which have already been defended according to his repetition rank.

- **(SR-2) (Formal rule for atomic formulas):** **P** is allowed to utter an atomic formula at (i, t) only if **O** has first uttered it at (i, z).

- **(SR-3) (Winning rule):** A player wins a play if and only if the other player cannot make a move.

Now we introduce the structural rules defining the conditions under which particle rules for modal operators can/have to be used.

- **(SR-5) (Formal rule for contextual points t):** To challenge a move as $\langle \mathbf{P} - i, t : \bigcirc\varphi \rangle$, **O** can choose any contextual point u whenever other rules allow him to do so. To challenge a move as $\langle \mathbf{P} - i, t : \bigcirc^{-1}\varphi \rangle$, **O** can choose any contextual point u provided that he has not chosen a contextual point v before to challenge a move as $\langle \mathbf{P} - i, t : \bigcirc^{-1}\varphi \rangle$.

  To challenge a move as $\langle \mathbf{O} - i, t : \bigcirc\varphi \rangle$, **P** can only choose a contextual point u already chosen by **O** to challenge a move as $\langle \mathbf{P} - i, t : \bigcirc\varphi \rangle$. To challenge a move as $\langle \mathbf{O} - i, t : \bigcirc^{-1}\varphi \rangle$, **P** can only choose a contextual point u already chosen by **O** to challenge a move as $\langle \mathbf{P} - i, t : \bigcirc^{-1}\varphi \rangle$.

  However **P** can choose the initial contextual point t to challenge a move as $\langle \mathbf{O} - i, t.u : \bigcirc\varphi \rangle$ or $\langle \mathbf{O} - i, t.u : \bigcirc^{-1}\varphi \rangle$ under some conditions:

  - **(SR-5.1)** **P** can choose the initial contextual point t to challenge a move as $\langle \mathbf{O} - i, t.u : \bigcirc\varphi \rangle$ if **O** has chosen the contextual point u to challenge a move as $\langle \mathbf{P} - i, t : \bigcirc^{-1}\varphi \rangle$.
  - **(SR-5.2)** **P** can choose the initial contextual point t to challenge a move as $\langle \mathbf{O} - i, t.u : \bigcirc^{-1}\varphi \rangle$ if **O** has chosen the contextual point u to challenge a move as $\langle \mathbf{P} - i, t : \bigcirc\varphi \rangle$.

[7]See N. Clerbout for more details [25].

- **(SR-6) (Formal rule for contextual points j):** To challenge a move as $\langle \mathbf{O} - i, t : B\varphi \rangle$, **P** can only choose a contextual point j already chosen by **O** to challenge a move as $\langle \mathbf{P} - i, t : B\varphi \rangle$; if **O** did not choose any contextual point to challenge a move as $\langle \mathbf{P} - i, t : B\varphi \rangle$, **P** can choose a new contextual point j.

  To challenge a move as $\langle \mathbf{O} - i, t : I\varphi \rangle$, **P** can only choose a contextual point j already chosen by **O** to challenge a move as $\langle \mathbf{P} - i, t : I\varphi \rangle$ or $\langle \mathbf{P} - i, t : B\varphi \rangle$.

  To challenge a move as $\langle \mathbf{O} - i, t : A\varphi \rangle$, **P** can only choose a contextual point j already chosen by **O** to challenge a move as $\langle \mathbf{P} - i, z : I\varphi \rangle$ or $\langle \mathbf{P} - i, z : B\varphi \rangle$ or $\langle \mathbf{P} - i, z : A\varphi \rangle$; or he can choose the contextual point i.

  However **P** can choose more contextual points j to challenge a move as $\langle \mathbf{O} - i, t : B\varphi \rangle$ under some conditions: let three contextual points t, u and v be such that u and v have been chosen by **O** to challenge a move as $\langle \mathbf{P} - i, t : \bigcirc\varphi \rangle$ and consider three contextual points i, j and k:

  - **(SR-6.1) P** can choose a contextual point j to challenge a move as $\langle \mathbf{O} - i, t : B\varphi \rangle$ if **O** has chosen the contextual point k to challenge a move as $\langle \mathbf{P} - i, t : B\varphi \rangle$ and to challenge a move as $\langle \mathbf{P} - i, t.u : I\varphi \rangle$/or if he has stated $\langle \mathbf{O} - i, t.u : ?I_k^* \rangle$ and if **O** has chosen the contextual point j to challenge a move as $\langle \mathbf{P} - i, t.u : B\varphi \rangle$.
  - **(SR-6.2) P** can choose a contextual point j to challenge a move as $\langle \mathbf{O} - i, t.u : B\varphi \rangle$ if **O** has chosen the contextual point j to challenge a move as $\langle \mathbf{P} - i, t : B\varphi \rangle$ and to challenge a move as $\langle \mathbf{P} - i, t.u : I\varphi \rangle$/or if he has stated $\langle \mathbf{O} - i, t.u : ?I_j^* \rangle$.
  - **(SR-6.3) P** can choose a contextual point j to challenge a move as $\langle \mathbf{O} - i, t.v : B\varphi \rangle$ if **O** has chosen the contextual point j to challenge a move as $\langle \mathbf{P} - i, t.u : B\varphi \rangle$ and if every contextual points j chosen to challenge a move as $\langle \mathbf{X} - i, t.u : I\varphi \rangle$ can also be chosen to challenge a move as $\langle \mathbf{X} - i, t.v : I\varphi \rangle$.
  - **(SR-6.4) P** can choose a contextual point j to challenge a move as $\langle \mathbf{O} - i, t.v : B\varphi \rangle$ if **O** has chosen the contextual point j to challenge a move as $\langle \mathbf{P} - i, t.u : B\varphi \rangle$ and to challenge a move as $\langle \mathbf{P} - i, t.v : I\varphi \rangle$/or if he has stated $\langle \mathbf{O} - i, t.v : ?I_j^* \rangle$.

- **(SR-7) (Request rule): Y** can choose a contextual point j and request **X** to confirm that this contextual point j could be chosen to challenge a move as $\langle \mathbf{X} - i, t : I\varphi \rangle$, only if $\langle \mathbf{O} - i.j, z : \varphi \rangle \in d_{\Delta}$.

Now we have to explain our structural rules, more precisely our formal rules: why do they make sense? What does this mean for players? What does this mean for the notion of beliefs and information? First note that players can discuss about facts, time (immediate past/future), the beliefs of an agent or the information received by an agent.

Formal rule for atomic formulas The main point of this rule is that the Proponent can only state atomic formulas in a particular context if the Opponent has already done it in some context. Indeed we saw that the Opponent tries to build a counter-argument to the thesis of the Proponent. Then he is the only one who can introduce (uttering first) atomic formulas in a context in a dialogue. The Proponent can only reuse them. But the important thing here is the contextual point i in which the Opponent states the atomic formulas, not the contextual point t. When the Opponent utters an atomic formula, in fact he states a proposition describing a fact that holds in some particular context. And we only consider here facts that do not change during the play that is, the facts described by the propositions stated by the players do not change during the discussion. For example if the fact "Earth revolves around the sun" holds in the context (i, t), it will also hold in the context (i, z) for any t and z. That's why it is sufficient that Opponent utters atomic formulas at (i, z) to be reused by Proponent at (i, t). This is the counterpart of the non-changing worlds (facts describing the world do not change over time) in the model-theoretic approach.

Formal rule for contextual points t On one hand, the main point of this rule is that Opponent can choose several contextual points u to challenge a move as $\langle \mathbf{P} - i, t : \bigcirc\varphi \rangle$ according to the other rules (his repetition rank), but only one contextual point v to challenge a move as $\langle \mathbf{P} - i, t : \bigcirc^{-1}\varphi \rangle$ operator. When players talk about (immediate) past, they only deal with one immediate past context from the actual context. Indeed they discuss about a fixed and determined (immediate) past. There are no several possibilities of (immediate) past. For example, if "at the previous instant it was the case that some peanuts lie all over the table" holds in the context (i, t), there is exactly one context $(i, t.u)$ where "it is the case that some peanuts lie all over the table" holds. On the contrary, when players discuss about the (immediate) future, they talk about an undetermined future and so several possibilities of an (immediate) future. For example, if "at every next instant it will be the case that some peanuts will be cleaned up" holds in the context (i, t), there are several possible (immediate) contexts $(i, t.u)$ where "some peanuts are actually cleaned up" holds. This is the counterpart of the branching-time frame in the model-theoretic approach.

On the other hand, the Proponent cannot introduce contextual points u for the same reason he cannot introduce atomic formulas: only the Opponent can do this because the Opponent tries to build a counter-argument to the thesis of the Proponent. Note that when players discuss about the (immediate) past and future, they have to be consistent with the notion of time. Thus they have to be consistent with respect to the choices of contextual points u they make. When a player discusses about the (immediate) past in the context (i, t) and then explicitly considers a(n) (immediate) past context $(i, t.u)$, then (i, t) is a(n) (immediate) future context of $(i, t.u)$. Conversely, if he discusses about the (immediate) future in the context (i, t) and explicitly considers a(n) (immediate) future context $(i, t.u)$, then (i, t) is a(n) (immediate) past context of $(i, t.u)$. That's why even if the initial contextual point t is not technically chosen by the Opponent, if the Opponent challenges a move as $\langle \mathbf{P} - i, t : \bigcirc\varphi \rangle$ or $\langle \mathbf{P} - i, t : \bigcirc^{-1}\varphi \rangle$ with the contextual point u, then the Proponent can choose t to challenge respectively a move as $\langle \mathbf{O} - i, t.u : \bigcirc^{-1}\varphi \rangle$ or $\langle \mathbf{O} - i, t.u : \bigcirc\varphi \rangle$.

Formal rule for contextual points j The Proponent cannot introduce contextual points j for the same reason that he cannot introduce atomic formulas and contextual points u: only the Opponent can do this because the Opponent tries to build a counter-argument to the thesis of the Proponent.

However, the Proponent can introduce a contextual point j to challenge a move as $\langle \mathbf{O} - i, t : B\varphi \rangle$ if the Opponent did not introduce such a contextual point j to challenge a move as $\langle \mathbf{P} - i, t : B\varphi \rangle$. When players discuss about beliefs, they discuss about the beliefs of an agent such that these beliefs are consistent. Thus if a player states that an agent believes a proposition φ, there should be at least one context this agent conceives (see particle rules in Section 7.2.2) where φ holds. Otherwise, the beliefs of the agent would be inconsistent. For example, if "an agent believes that the Earth revolves around the sun" holds in (i, t), there exists at least one context $(i.j, t)$ considered by the agent where "Earth revolves around the sun" holds. This is the counterpart of the seriality of beliefs in the model-theoretic approach.

Besides, the Proponent can choose a contextual point j already chosen by $\mathbf{O}$ to challenge a move as $\langle \mathbf{P} - i, t : B\varphi \rangle$, to challenge a move as $\langle \mathbf{O} - i, t : I\varphi \rangle$. In other words, the Proponent can use a contextual point j initially chosen to challenge a belief operator, to challenge an information operator in the same context. There is an interplay between the contextual points that can be chosen to challenge a belief operator and the contextual points that can be chosen to challenge an information operator in the same context. Indeed players discuss about the beliefs of an agent as well as the information she receives, such that if the agent has been informed about a proposition, she believes this proposition. That's why all contextual points chosen to challenge a belief operator in a context (i, t) can also be chosen to challenge an information operator in (i, t). For example, if "an agent is informed that the Earth revolves around the sun" holds in (i, t) and "an agent believes that the Earth revolves around the sun" holds in (i, t), then every context $(i.j, t)$ that the agent conceives is also a context of which she has been informed. This is the counterpart of the acceptance of information in the model-theoretic approach.

Finally, **P** can choose a contextual point j already chosen by **O** to challenge a move as $\langle \mathbf{P} - i, z : I\varphi \rangle$ or $\langle \mathbf{P} - i, z : B\varphi \rangle$ or $\langle \mathbf{P} - i, z : A\varphi \rangle$, to challenge a move as $\langle \mathbf{O} - i, t : A\varphi \rangle$. When players discuss about a proposition that is always the case (in a non-temporal sense), they state that whatever the context they already discussed about, the proposition is the case in that context. For example, if "it is always the case that Amsterdam is the capital city of the Netherlands" holds at (i, t), then "Amsterdam is the capital city of the Netherlands" holds in all contexts of the dialogue. Then whatever the modal operator for which the contextual point j has been introduced, it can be chosen to challenge a universal operator. Choices to challenge a universal operator are transitive and symmetric. The additional condition that **P** can choose the initial contextual point i to challenge $\langle \mathbf{O} - i, t : A\varphi \rangle$ ensures there is reflexivity. This is the counterpart of the $S5$ frame in the model-theoretic approach.

But note that once again, the contextual point t is not important here. **O** can introduce the contextual point j in the context (i, z) and **P** can choose this contextual point j to challenge a universal operator in (i, t). Indeed when players introduce contexts in a dialogue, they can discuss about them all along the dialogue. Contexts do not disappear, they are constant throughout the whole discussion. This is the counterpart of the constant worlds over time in the model-theoretic approach (see Definition 2.8.3).

Now we have to explain the four exceptions we notice with respect to the choices of contextual points j **P** can make to challenge a move as $\langle \mathbf{O} - i, t : B\varphi \rangle$. We are dealing here with the interplay between information and beliefs as well as the interplay between beliefs themselves – namely, initial beliefs and revised beliefs – that involve an interplay between the choices of the players. In other words, some choices of a player allow for other choices for the other player to be made.

The first exception states that if there exists a context the agent conceives in (i, t) and of which she is informed in $(i, t.u)$ such that $(i, t.u)$ is an immediate future context of (i, t), then all the contexts the agent conceives in $(i, t.u)$ were already conceived by the agent in (i, t). Indeed players discuss about an agent who receives a piece of information, such that if she receives an information compatible (that is, consistent) with her beliefs, then she does not add beliefs about which she is not informed. All the contexts the agent conceives after the information were already conceived before the information. That's why under these conditions, **P** can choose a contextual point j initially chosen to challenge a belief operator in $(i, t.u)$, to challenge a belief operator in (i, t). This is the counterpart of the No Add property in the model-theoretic approach.

The second exception states that if there exists a context the agent conceives in (i, t) and of which she is informed in $(i, t.u)$ such that $(i, t.u)$ is an immediate future context of (i, t), then the agent also conceives this context in $(i, t.u)$. Indeed players discuss about an agent who receives a piece of information, such that if she receives an information compatible (that is, consistent) with her beliefs, then she does not drop these beliefs. She still conceives contexts compatible with the information is received. That's why under these conditions, **P** can choose a contextual point j initially chosen to challenge a belief operator in (i, t), to challenge a belief operator in $(i, t.u)$. This is the counterpart of the No Drop property in the model-theoretic approach.

The third exception states that if an agent is informed of the same contexts in contexts $(i, t.u)$ and $(i, t.v)$ such that both contexts are immediate future contexts of (i, t), then she conceives the same contexts in $(i, t.u)$ and $(i, t.v)$. In that case, **P** can choose a contextual point j initially chosen to challenge a belief operator in $(i, t.u)$, to challenge a belief operator in $(i, t.v)$. Indeed players discuss about an agent who is consistent with respect to the information she receives. If the beliefs of the agent change and differ over time, it is only because she receives different information. This is the counterpart of the Equivalence property in the model-theoretic approach.

The last exception states that if an agent conceives a context in $(i, t.u)$ and is informed about the same context in $(i, t.v)$ such that both contexts are immediate future contexts of (i, t), then the agent also conceives this context in $(i, t.v)$. Indeed the players discuss about the beliefs of the agent such that these beliefs are rationalized with respect to the information received. This is the counterpart of the PLS property in the model-theoretic approach.

Request Rule This rule ensures that players can only choose the non-standard challenge on an information operator that is, request the other player to confirm that a particular contextual point j could be chosen to challenge a move as $\langle \mathbf{X} - i, t : I\varphi \rangle$, if the Opponent already stated that φ holds in the contextual point j[8]. Indeed a player who states that an agent is informed about a proposition φ, is committed to defend that all and only contexts where φ holds are contexts of which the agent is informed. For example, if "an agent is informed that the President is dead" holds in (i, t), then he is informed of all and only contexts where "the President is dead" holds. This is the counterpart of the non-normality of the information operator.

Structural rules of ***DTDL*** The structural rules of *DTDL* consist of (SR-0), (SR-1), (SR-2), (SR-3), (SR-4), (SR-5), (SR-5.1), (SR-5.2), (SR-6), (SR-6.1), (SR-6.2), (SR-6.3), (SR-6.4), and (SR-7).

7.2.4. DEFINITION. *DTDL* is defined by the set of particle rules and structural rules.

[8]Remember that φ must be Boolean so the contextual point t does not matter.

An argumentative interpretation When investigating the relation between beliefs and information over time from an external dynamic perspective, we interpret this relation through the notion of choice. Players make choices when they discuss about the beliefs of an agent and the information she receives. We notice a genuine interplay between the choices of the players. Under specific conditions, some choices initially made to challenge belief operators can also be made to challenge information operators; some choices initially made to challenge belief operators in a particular context can also be made to challenge belief operators in different contexts and so on. Some choices allow some other choices otherwise prohibited by the other rules. Not only the meaning of belief and information operators is defined in terms of choice, but the belief revision policy itself is defined in terms of choice. Indeed the restrictions on the possible choices players can make to challenge belief and information operators define a particular belief revision policy. In other words *DTDL* allows an argumentative interpretation of the belief revision policy of Bonanno.

7.2.4 Applications

In the Figures 7.3, 7.4, 7.5 the number in the outer column corresponds to the number of the move whereas the one in the inner column corresponds to the number of the move challenged.

Non surprising information We illustrate a play where two players discuss about an agent who receives a piece of information that does not contradict her beliefs and revises this beliefs in the light of this new information. The thesis of this play described in Figure 7.3 is the formula $\neg[\neg B\neg q \wedge Bp] \vee [\bigcirc(\neg Iq \vee B(p \wedge q))]$.

Explanations of Figure 7.3 In accordance with the starting rule **(SR-0)**, the Proponent states the thesis at move 0. At move 1, the Opponent challenges the disjunction and the Proponent chooses to defend the first disjunct at move 2. At move 3, the Opponent challenges the negation and the Proponent has no corresponding defence. Then he chooses to challenge the conjunction of move 3 choosing respectively the first conjunct at move 4 and the second conjunct at move 6 in accordance with his repetition rank $n := 2$. The Opponent defends the corresponding conjunct at moves 5 and 7. At move 8, the Proponent challenges the negation of move 5 and the Opponent counter-attacks the move 8 since he has no possible defence. So he challenges the belief operator choosing the contextual point 2 at move 9 and the Proponent defends $\neg q$ in the context $(1.2, 1)$ at move 10. The Opponent challenges the negation of move 10 and the Proponent has no possible defence. Then he chooses to change the defence against the challenge of move 1, choosing the second disjunct at move 12 in accordance with his repetition rank $n := 2$. At move 13, the Opponent challenges the $\bigcirc$ operator choosing a contextual point 2 and the Proponent defends $\neg Iq \vee B(p \wedge q)$ in the context $(1, 1.2)$.

	O			**P**	
				$1,1 : \neg[\neg B\neg q \wedge Bp] \vee [\bigcirc(\neg Iq \vee B(p \wedge q))]$	0
	$m := 1$			$n := 2$	
1	$1,1 : ?_{\vee}$	0		$1,1 : \neg[\neg B\neg q \wedge Bp]$	2
3	$1,1 : \neg B\neg q \wedge Bp$	2		$\otimes$	
5	$1,1 : \neg B\neg q$		3	$1,1 : ?_{\wedge 1}$	4
7	$1,1 : Bp$		3	$1,1 : ?_{\wedge 2}$	6
	$\otimes$		5	$1,1 : B\neg q$	8
9	$1,1 : ?B_2$	8		$1.2,1 : \neg q$	10
11	$1.2,1 : q$	10		$\otimes$	
				$1,1 : \bigcirc(\neg Iq \vee B(p \wedge q))$	12
13	$1,1 : ?\bigcirc_2$	12		$1,1.2 : \neg Iq \vee B(p \wedge q)$	14
15	$1,1.2 : ?_{\vee}$	14		$1,1.2 : \neg Iq$	16
17	$1,1.2 : Iq$	16		$\otimes$	
				$1,1.2 : B(p \wedge q)$	18
19	$1,1.2 : ?B_3$	18		$1.3,1.2 : p \wedge q$	20
21	$1.3,1.2 : ?_{\wedge 1}$	20		$1.3,1.2 : p$	26
23	$1,1.2 : ?I_2^*$		17	$1,1.2 : !I_2$	22
25	$1.3,1 : p$		7	$1,1 : ?B_3$	24

Figure 7.3: Non surprising information - Play 1

Then the Opponent challenges the disjunction and the Proponent chooses to defend the first disjunct. The Opponent challenges the negation of move 16 and the Proponent has no corresponding defence. He changes his defence against the challenge of move 15 choosing to defend the second disjunct. At move 19, the Opponent challenges the belief operator choosing the contextual point 3 and the Proponent defends $p \wedge q$ in the context $(1.3, 1.2)$. The Opponent then chooses the first conjunct when he challenges the conjunction of move 20. In accordance with the formal rule **(SR-2)**, the Proponent cannot defend now since he cannot utter first an atomic formula in a particular context. But he can challenge the information operator of move 17 choosing the non standard challenge and the contextual point 2. Indeed the Opponent has stated that q holds in the contextual point 2 at move 11 so the request rule **(SR-7)** allows him to choose this contextual point 2 for his non standard challenge. Then the Opponent defends at move 23, confirming that this contextual point 2 can be chosen to challenge an information operator in the context $(1, 1.2)$. The move 9, move 19 and move 23 allow the Proponent to challenge the belief operator of move 7 choosing the contextual point 3 in accordance with the formal rule for contextual point j **(SR-6.1)**. Indeed the Opponent has chosen the contextual point 2 to challenge a belief operator in $(1, 1)$ and has stated that this contextual point 2 can be chosen to challenge an information operator in the context $(1, 1.2)$, and he has also chosen the contextual point 3 to challenge a belief operator in the context $(1, 1.2)$. The Opponent then defends stating p in the context $(1.3, 1)$ allowing the Proponent to state p in $(1.3, 1.2)$ at move 26 to defend against the challenge of move 21 in accordance with the formal rule **(SR-2)**. In accordance with the winning rule **(SR-3)** the Proponent wins since the Opponent cannot move.

Note In move 22, the Proponent requests the Opponent to confirm that the contextual point 2 can be chosen to challenge an information operator in the context $(1, 1.2)$. In other words, the Proponent requests the Opponent to confirm a choice he could do. In moves 22 and 23, the players are actually dealing about the Opponent choices with respect to his previous choices and statements. And previously, the Opponent has stated that q holds in the contextual point 2 and that the agent is informed about q in the context $(1, 1.2)$ (moves 11 and 17). If the Opponent is consistent with himself, he must stated that $(1.2, 1.2)$ is a context of which the agent is informed. Indeed a player who states that an agent is informed about a proposition φ, is committed to defend that all and only contexts where φ holds are contexts of which the agent is informed (see Request Rule p 168). So the Opponent must confirm that 2 is an available choice to challenge an information operator in the context $(1, 1.2)$ with respect to his argumentation otherwise he contradicts himself.

What happens now if the Opponent chooses the second conjunct when he challenges the conjunction of move 20? We consider a play with the same thesis as in Figure 7.3.

Explanations of Figure 7.4 The play proceeds in the same way as in Figure 7.3 until move 21. Indeed the Opponent chooses the second conjunct when he challenges the conjunction of move 20. In that case, the Proponent challenges the information operator of move 17 choosing the standard challenge. Since the Opponent has chosen the contextual point 3 to challenge the belief operator of move 18 (move 19), the Proponent can choose this contextual point to challenge the information operator of move 17 in accordance with the formal rule for contextual point j **(SR-6)** at move 22. At move 23, the Opponent defends stating q in the context $(1.3, 1.2)$ allowing the Proponent to state q in $(1.3, 1.2)$ at move 24 to defend against the challenge of move 21 in accordance with the formal rule **(SR-2)**. In that play, in accordance with the winning rule **(SR-3)** the Proponent wins too.

In the Figures 7.3 and 7.4, the Opponent first states that the agent considers q possible (move 5) and believes p (move 7), and then is informed about q (move 17). In that case and according to the belief revision policy described by the rules

	O			**P**	
				$1,1 : \neg[\neg B\neg q \wedge Bp] \vee [\bigcirc(\neg Iq \vee B(p \wedge q))]$	0
	$m := 1$			$n := 2$	
1	$1,1 : ?_{\vee}$	0		$1,1 : \neg[\neg B\neg q \wedge Bp]$	2
3	$1,1 : \neg B\neg q \wedge Bp$	2		$\otimes$	
5	$1,1 : \neg B\neg q$		3	$1,1 : ?_{\wedge 1}$	4
7	$1,1 : Bp$		3	$1,1 : ?_{\wedge 2}$	6
	$\otimes$		5	$1,1 : B\neg q$	8
9	$1,1 : ?B_2$	8		$1.2,1 : \neg q$	10
11	$1.2,1 : q$	10		$\otimes$	
				$1,1 : \bigcirc(\neg Iq \vee B(p \wedge q))$	12
13	$1,1 : ?\bigcirc_2$	12		$1,1.2 : \neg Iq \vee B(p \wedge q)$	14
15	$1,1.2 : ?_{\vee}$	14		$1,1.2 : \neg Iq$	16
17	$1,1.2 : Iq$	16		$\otimes$	
				$1,1.2 : B(p \wedge q)$	18
19	$1,1.2 : ?B_3$	18		$1.3,1.2 : p \wedge q$	20
21	$1.3,1.2 : ?_{\wedge 2}$	20		$1.3,1.2 : q$	24
23	$1.3,1.2 : q$		17	$1,1.2 : ?I_3$	22

Figure 7.4: Non surprising information - Play 2

of *DTDL*, it is impossible for players to argue that the agent does not believe p and q after she receives the information.

Surprising information What happens if the information is surprising? We illustrate a play where two players discuss about an agent who receives a piece of information that does contradict her beliefs and revises this beliefs in the light of this new information. The thesis of this play described in Figure 7.5 is the formula $\neg[B\neg q \wedge Bp] \vee [\bigcirc(\neg Iq \vee B(p \wedge q))]$.

Explanations of Figure 7.5 The play starts and proceeds as in Figure 7.3. At move 8, the Proponent cannot challenge the moves 5 or 7 in accordance with the formal rule for contextual point j **(SR-6)**, so he changes his defence against the challenge of move 1, choosing the second disjunct in accordance with his repetition rank $n := 2$. Once again the play proceeds as in Figure 7.3. In accordance with the formal rule **(SR-2)**, the Proponent cannot defend against the challenge on move 16 since he cannot utter first an atomic formula in a particular context. The only move the Proponent can then make is to challenge the move 13 choosing the standard challenge. Since the Opponent has chosen the contextual point 3 to challenge the belief operator of move 14 (move 15), the Proponent can choose this contextual point to challenge the information operator of move 13 in accordance

	O			**P**	
				$1,1:\neg[B\neg q \wedge Bp] \vee [\bigcirc(\neg Iq \vee B(p \wedge q))]$	0
	$m := 1$			$n := 2$	
1	$1,1:?_{\vee}$	0		$1,1:\neg[B\neg q \wedge Bp]$	2
3	$1,1:B\neg q \wedge Bp$	2		$\otimes$	
5	$1,1:B\neg q$		3	$1,1:?_{\wedge 1}$	4
7	$1,1:Bp$		3	$1,1:?_{\wedge 2}$	6
				$1,1:\bigcirc(\neg Iq \vee B(p \wedge q))$	8
9	$1,1:?\bigcirc_2$	8		$1,1.2:\neg Iq \vee B(p \wedge q)$	10
11	$1,1.2:?_{\vee}$	10		$1,1.2:\neg Iq$	12
13	$1,1.2:Iq$	12		$\otimes$	
				$1,1.2:B(p \wedge q)$	14
15	$1,1.2:?B_3$	14		$1.3,1.2:p \wedge q$	16
17	$1.3,1.2:?_{\wedge 1}$	16		—	
19	$1.3,1.2:q$		13	$1,1.2:?I_3$	18

Figure 7.5: Surprising information

with the formal rule for contextual point j **(SR-6)** at move 18. At move 19, the Opponent defends stating q in the context $(1.3, 1.2)$. Then the Proponent cannot make a move.

The main difference with Figure 7.3 is that the Proponent cannot challenge the move 7 with contextual point 3 in accordance with the rules. The choices of the Opponent do not allow this choice to the Proponent. So in that play, the Proponent loses in accordance with the winning rule **(SR-3)**.

In the Figure 7.5, the Opponent first states that the agent believes $\neg q$ (move 5) and believes p (move 7), and then is informed about q (move 13). In that case and according to the belief revision policy described by the rules of $DTDL$, it is possible for the Opponent to argue that the agent does not believe p after she receives the information and then does not believe $(p \wedge q)$.

Stubborn agent What happens if the agent is stubborn? We illustrate a play where two players discuss about a stubborn agent who receives a piece of information that does not contradict her beliefs and revises this beliefs in the light of this new information such that she does not believe the information. The thesis of this play described in Figure 7.6 is the formula $\neg[\neg B\neg q \wedge Bp] \vee [\bigcirc(\neg Iq \vee (Bp \wedge \neg B\neg q))]$.

Explanations of Figure 7.6 The play starts and proceeds as in Figure 7.5. At move 17, the Opponent chooses the second conjunct when he challenges the conjunction of move 16. The Proponent defends stating $\neg q$ in the context

	O			**P**	
				$1,1:\neg[B\neg q\wedge Bp]\vee[\bigcirc(\neg Iq\vee B(p\wedge\neg q))]$	0
	$m:=1$			$n:=2$	
1	$1,1:?_{\vee}$	0		$1,1:\neg[B\neg q\wedge Bp]$	2
3	$1,1:B\neg q\wedge Bp$	2		$\otimes$	
5	$1,1:B\neg q$		3	$1,1:?_{\wedge 1}$	4
7	$1,1:Bp$		3	$1,1:?_{\wedge 2}$	6
				$1,1:\bigcirc(\neg Iq\vee B(p\wedge q))$	8
9	$1,1:?\bigcirc_2$	8		$1,1.2:\neg Iq\vee B(p\wedge\neg q)$	10
11	$1,1.2:?_{\vee}$	10		$1,1.2:\neg Iq$	12
13	$1,1.2:Iq$	12		$\otimes$	
				$1,1.2:B(p\wedge\neg q)$	14
15	$1,1.2:?B_3$	14		$1.3,1.2:p\wedge\neg q$	16
17	$1.3,1.2:?_{\wedge 2}$	16		$1.3,1.2:\neg q$	18
19	$1.3,1.2:q$				
21	$1.3,1.2:q$		13	$1,1.2:?I_3$	20

Figure 7.6: Stubborn agent

(1.3, 1.2). At move 19, the Opponent challenges the negation stating q in the context (1.3, 1.2). The only move the Proponent can then make is to challenge the move 13 choosing the standard challenge. Since the Opponent has chosen the contextual point 3 to challenge the belief operator of move 14 (move 15), the Proponent can choose this contextual point to challenge the information operator of move 13 in accordance with the formal rule for contextual point j **(SR-6)** at move 20. At move 21, the Opponent defends stating q in the context (1.3, 1.2). Then the Proponent cannot make a move.

In the Figure 7.6, the Opponent first states that the agent believes $\neg q$ (move 5) and believes p (move 7), and then is informed about q (move 13). In that case and according to the belief revision policy described by the rules of *DTDL*, it is possible for the Opponent to argue that the agent does not believe $\neg q$ after she receives the information and then does not believe $(p\wedge\neg q)$.

Irrational agent What happens if the agent is irrational? We illustrate a play where two players discuss about an irrational agent who receives a piece of information that does not contradict her beliefs and revises this beliefs in the light of this new information in an irrational way. Here we interpret "irrational way" as changing her beliefs regardless the information received. The thesis of this play described in Figure 7.7 is the formula $\neg[\neg\bigcirc\neg(Ip\wedge Bq)]\vee[\bigcirc(\neg Ip\vee B\neg q)]$.

	O			**P**	
				$1,1 : \neg[\neg\bigcirc\neg(Ip \wedge Bq)] \vee [\bigcirc(\neg Ip \vee B\neg q)]$	0
	$m := 1$			$n := 2$	
1	$1,1 : ?_{\vee}$	0		$1,1 : \neg[\neg\bigcirc\neg(Ip \wedge Bq)]$	2
3	$1,1 : \neg\bigcirc\neg(Ip \wedge Bq)$	2		$\otimes$	
	$\otimes$		3	$1,1 : \bigcirc\neg(Ip \wedge Bq)$	4
5	$1,1 : ?\bigcirc_2$	4		$1,1.2 : \neg(Ip \wedge Bq)$	6
7	$1,1.2 : Ip \wedge Bq$	6		$\otimes$	
9	$1,1.2 : Ip$		7	$1,1.2 : ?_{\wedge 1}$	8
11	$1,1.2 : Bq$		7	$1,1.2 : ?_{\wedge 2}$	10
				$1,1 : \bigcirc(\neg Ip \vee B\neg q)$	12
13	$1,1 : ?\bigcirc_3$	12		$1,1.3 : \neg Ip \vee B\neg q$	14
15	$1,1.3 : ?_{\vee}$	10		$1,1.3 : \neg Ip$	16
17	$1,1.3 : Ip$	16		$\otimes$	
				$1,1.3 : B\neg q$	18
19	$1,1.3 : ?B_2$	18		$1.2,1.3 : \neg q$	20
21	$1.2,1.3 : q$				
23	$1.2,1.3 : p$		17	$1,1.3 : ?I_2$	22
25	$1,1.2 : ?I_2^*$		9	$1,1.2 : !I_2$	24
27	$1.2,1.2 : q$		11	$1,1.2 : ?B_2$	26

Figure 7.7: Irrational agent

Explanations of Figure 7.7 At move 1, the Opponent challenges the disjunction and the Proponent chooses to defend the first disjunct at move 2. At move 3, the Opponent challenges the negation and the Proponent has no corresponding defence. Then he counter-attacks and challenges the negation of move 3. At move 5 the Opponent challenges the $\bigcirc$ operator choosing a contextual point 2 and the Proponent defends $\neg(Ip \wedge Bq)$ in the context $(1, 1.2)$. The Opponent challenges the negation at move 7 and the Proponent counter-attacks challenging the conjunction. He chooses respectively the first conjunct at move 8 and the second conjunct at move 10. The Opponent defends the corresponding conjunct at move 9 and 11. The Proponent cannot challenge these moves so he decides to change his defence against the challenge of move 1, choosing the second disjunct at move 12 in accordance with his repetition rank $n := 2$. At move 13 the Opponent challenges the $\bigcirc$ operator choosing a contextual point 3 and the Proponent defends $\neg Ip \vee B\neg q$ in the context $(1, 1.3)$. At move 15, the Opponent challenges the disjunction and the Proponent chooses to defend the first disjunct at move 16. At move 17, the Opponent challenges the negation and the Proponent has no corresponding defence. He decides to change his defence against the challenge of move 15, choosing the second disjunct at move 18 in accordance with his repetition rank $n := 2$. At move 19, the Opponent challenges the belief operator choosing the contextual point 2 and the Proponent defends $\neg q$ in the context $(1.2, 1.3)$. The Opponent then challenges the negation of move 20 stating q in the context $(1.2, 1.3)$. The Proponent challenges the information operator of move 17 choosing the standard challenge and the contextual point 2 in accordance with the formal rule for contextual point j **(SR-6)**. Indeed the Opponent has chosen the contextual point 2 to challenge a belief operator in $(1, 1.3)$ at move 19. Since the Opponent defends stating p in $(1.2, 1.3)$, the Proponent can challenge the information operator of move 9 choosing the contextual point 2 for his non standard challenge at move 24 in accordance with the request rule **(SR-7)**. Then the Opponent defends at move 25, confirming that this contextual point 2 can be chosen to challenge an information operator in the context $(1, 1.2)$. The move 19, move 22 and move 25 allow the Proponent to challenge the belief operator of move 11 choosing the contextual point 2 in accordance with the formal rule for contextual point j **(SR-6.3)**. Indeed the Opponent has chosen the contextual point 2 to challenge a belief operator in $(1, 1.3)$ and so contextual point 2 can be chosen by the Proponent to challenge an information operator in $(1, 1.3)$, and the Opponent has also stated that this contextual point 2 can be chosen to challenge an information operator in the context $(1, 1.2)$. The Opponent then defends stating q in the context $(1.2, 1.2)$. In accordance with the winning rule **(SR-3)** the Proponent loses since he cannot move.

Note In move 24, the Proponent requests the Opponent to confirm that the contextual point 2 can be chosen to challenge an information operator in the context $(1, 1.2)$ that is, he requests the Opponent to confirm a choice he could do. In moves 24 and 25, the players are actually dealing about the Opponent choices with respect to his previous choices and statements: the Opponent has stated that the agent is informed about p in the context $(1, 1.2)$ and that p holds in the contextual point 2 (moves 9 and 23). If the Opponent is consistent with himself, he must stated that $(1.2, 1.2)$ is a context of which the agent is informed (see Request Rule). So he must confirm that 2 is an available choice to challenge an information operator in the context $(1, 1.2)$ with respect to his argumentation otherwise he contradicts himself.

In the Figure 7.7, the Opponent first states that the agent is informed about p (move 9) and believes q (move 11) in a particular context and that this agent is also informed about p in another context (move 17). In that case and according to the belief revision policy described by the rules of *DTDL*, it is possible for the Opponent to argue that the agent does not believe $\neg q$ after she receives the information in the second context.

Winning strategy In the Figures 7.3 and 7.4, we provide two different plays of the dialogue $\mathcal{D}_{\neg[\neg B\neg q\wedge Bp]\vee[\bigcirc(\neg Iq\vee B(p\wedge q))]}$ corresponding to different choices of the Opponent when he challenges the conjunction of move 20. Whatever he chooses the first –Figure 7.3– or the second –Figure 7.4– conjunct, the Proponent wins the play. In other words, the Proponent can win whatever the choices of the Opponent if he plays optimally that is, if he makes the choices allowing his victory (among the available ones).

Being able to win whatever the choices of the player **Y** means not only winning a play $d_\Delta \in \mathcal{D}_\Delta$ but also all the possible plays for Δ that is, winning $\mathcal{D}_\Delta$. In other words, this means that **X** has a winning strategy.

7.2.5. Definition. A player has a *winning strategy* if he can win whatever the choices of the other player.

This notion of winning strategy is the counterpart of the notion of validity in the model-theoretic approach. Since *DTDL* is a dialogical approach to L_{PLS*}, there exists a correspondence between having a winning strategy for a formula φ in *DTDL* and being valid in L_{PLS*} for the same formula φ.

7.3 Soundness and Completeness for *DTDL*

We prove that *DTDL* is sound and complete with respect to L_{PLS*} showing that there exists a winning strategy for the Proponent in $\mathcal{D}_\Delta$ iff Δ is a valid formula in L_{PLS*}.

We start with one hypothesis.

7.3.1. Hypothesis. *Both Players always play the best move that is, they are ideal players able to choose the best move to win the play. Thus we can always consider plays where the Opponent chooses* $m := 1$ *and the Proponent chooses* $n := 2$ *as repetition ranks. Indeed if the the Opponent plays in an optimal way, it is sufficient for him to have* $m := 1$ *because if he has a winning strategy and follows it, he does not have to change his defences or challenges. In Theorem 7.3.16, we show that the repetition ranks* $m := 1$ *and* $n := 2$ *are optimal respectively for the Opponent and the Proponent.*

7.3.1 Soundness

We prove that our dialogical approach is sound with respect to L_{PLS*} showing that if the Proponent has a winning strategy in $\mathcal{D}_\Delta$ then the formula Δ is valid in L_{PLS*}. We prove the contrapositive : we prove that if there exists a model satisfying $\neg\Delta$ then the Proponent cannot win any play with Δ as thesis.

We start from one hypothesis.

7.3.2. HYPOTHESIS. *A dialogical move* $\langle \boldsymbol{X} - i, t : \varphi \rangle$ *means that :*

- $\mathcal{M}, (i,t) \vDash \varphi$ *if* $\boldsymbol{X} = \boldsymbol{O}$
- $\mathcal{M}, (i,t) \vDash \neg\varphi$ *if* $\boldsymbol{X} = \boldsymbol{P}$

A dialogical move $\langle \boldsymbol{X} - i, t : ?I_j^* \rangle$ *means that :*

- $j \in I_t(i)$ *if* $\boldsymbol{X} = \boldsymbol{O}$
- $j \notin I_t(i)$ *if* $\boldsymbol{X} = \boldsymbol{P}$

We first need to prove that our particle rules preserve satisfaction that is, our Hypothesis 7.3.2 is preserved after the use of any particle rule.

7.3.3. LEMMA. *Given a branching time belief revision model* $\mathcal{M}$*, all the particle rules preserve satisfiability.*

7.3.4. PROOF. We show that our 8 particle rules preserve satisfiability.

◇ Particle rule for negation:
if $\langle \mathbf{X} - i, t : \neg\varphi \rangle \in d_\Delta$
then $\langle \mathbf{Y} - i, t : \varphi \rangle \in d_\Delta$

1. if $\mathbf{X} = \mathbf{O}$, by Hypothesis 7.3.2, $\mathcal{M}, (i,t) \vDash \neg\varphi$ iff $\mathcal{M}, (i,t) \nvDash \varphi$ (by Definition 2.8.11).
2. if $\mathbf{X} = \mathbf{P}$, by Hypothesis 7.3.2, $\mathcal{M}, (i,t) \vDash \neg\neg\varphi$ iff $\mathcal{M}, (i,t) \vDash \varphi$ (by Definition 2.8.11).

◇ Particle rule for conjunction:
if $\langle \mathbf{X} - i, t : \varphi_1 \wedge \varphi_2 \rangle \in d_\Delta$
then $\langle \mathbf{Y} - i, t : ?_{\wedge 1} \rangle \in d_\Delta$, or $\langle \mathbf{Y} - i, t : ?_{\wedge 2} \rangle \in d_\Delta$
so $\langle \mathbf{X} - i, t : \varphi_1 \rangle \in d_\Delta$, or $\langle \mathbf{X} - i, t : \varphi_2 \rangle \in d_\Delta$

1. if $\mathbf{X} = \mathbf{O}$, $\mathbf{P}$ can change his challenge since $n := 2$, then by Hypothesis 7.3.2, $\mathcal{M}, (i,t) \vDash (\varphi_1 \wedge \varphi_2)$ iff $\mathcal{M}, (i,t) \vDash \varphi_1$ and $\mathcal{M}, (i,t) \vDash \varphi_2$ (by Definition 2.8.11).
2. if $\mathbf{X} = \mathbf{P}$, $\mathbf{O}$ cannot change his challenge since $m := 1$, then by Hypothesis 7.3.2, $\mathcal{M}, (i,t) \vDash \neg(\varphi_1 \wedge \varphi_2)$ iff $\mathcal{M}, (i,t) \vDash \neg\varphi_1$ or $\mathcal{M}, (i,t) \vDash \neg\varphi_2$ (by Definition 2.8.11).

◇ Particle rule for disjunction:
if $\langle \mathbf{X} - i, t : \varphi_1 \vee \varphi_2 \rangle \in d_\Delta$
then $\langle \mathbf{Y} - i, t : ?_\vee \rangle \in d_\Delta$
so $\langle \mathbf{X} - i, t : \varphi_1 \rangle \in d_\Delta$, or $\langle \mathbf{X} - i, t : \varphi_2 \rangle \in d_\Delta$

1. if $\mathbf{X} = \mathbf{O}$, $\mathbf{O}$ cannot change his defence since $m := 1$, then by Hypothesis 7.3.2, $\mathcal{M}, (i, t) \vDash (\varphi_1 \vee \varphi_2)$ iff $\mathcal{M}, (i, t) \vDash \varphi_1$ or $\mathcal{M}, (i, t) \vDash \varphi_2$ (by Definition 2.8.11).

2. if $\mathbf{X} = \mathbf{P}$, $\mathbf{P}$ can change his defence since $n := 2$, then by Hypothesis 7.3.2, $\mathcal{M}, (i, t) \vDash \neg(\varphi_1 \vee \varphi_2)$ iff $\mathcal{M}, (i, t) \vDash \neg\varphi_1$ and $\mathcal{M}, (i, t) \vDash \neg\varphi_2$ (by Definition 2.8.11).

⋄ Particle rule for $\bigcirc^{-1}$ operator :
if $\langle \mathbf{X} - i, t : \bigcirc^{-1}\varphi \rangle \in d_\Delta$
then $\langle \mathbf{Y} - i, t : ?\bigcirc_u^{-1} \rangle \in d_\Delta$
so $\langle \mathbf{X} - i, t.u : \varphi \rangle \in d_\Delta$

1. if $\mathbf{X} = \mathbf{O}$, then by Hypothesis 7.3.2, $\mathcal{M}, (i, t) \vDash \bigcirc^{-1}\varphi$ iff $\mathcal{M}, (i, t.u) \vDash \varphi$ for every u such that $u \rightsquigarrow t$ (by Definition 2.8.11).

2. if $\mathbf{X} = \mathbf{P}$, then by Hypothesis 7.3.2, $\mathcal{M}, (i, t) \vDash \neg\bigcirc^{-1}\varphi$ iff $\mathcal{M}, (i, t.u) \vDash \neg\varphi$ for at least one u such that $u \rightsquigarrow t$ (by Definition 2.8.11).

⋄ Particle rule for $\bigcirc$ operator :
if $\langle \mathbf{X} - i, t : \bigcirc\varphi \rangle \in d_\Delta$
then $\langle \mathbf{Y} - i, t : ?\bigcirc_u \rangle \in d_\Delta$
so $\langle \mathbf{X} - i, t.u : \varphi \rangle \in d_\Delta$

1. if $\mathbf{X} = \mathbf{O}$, then by Hypothesis 7.3.2, $\mathcal{M}, (i, t) \vDash \bigcirc\varphi$ iff $\mathcal{M}, (i, t.u) \vDash \varphi$ for every u such that $t \rightsquigarrow u$ (by Definition 2.8.11).

2. if $\mathbf{X} = \mathbf{P}$, then by Hypothesis 7.3.2, $\mathcal{M}, (i, t) \vDash \neg\bigcirc\varphi$ iff $\mathcal{M}, (i, t.u) \vDash \neg\varphi$ for at least one u such that $t \rightsquigarrow u$ (by Definition 2.8.11).

⋄ Particle rule for B operator :
if $\langle \mathbf{X} - i, t : B\varphi \rangle \in d_\Delta$
then $\langle \mathbf{Y} - i, t : ?B_j \rangle \in d_\Delta$
so $\langle \mathbf{X} - i.j, t : \varphi \rangle \in d_\Delta$

1. if $\mathbf{X} = \mathbf{O}$, then by Hypothesis 7.3.2, $\mathcal{M}, (i, t) \vDash B\varphi$ iff $\mathcal{M}, (i.j, t) \vDash \varphi$ for every $j \in B_t(i)$ (by Definition 2.8.11).

2. if $\mathbf{X} = \mathbf{P}$, then by Hypothesis 7.3.2, $\mathcal{M}, (i, t) \vDash \neg B\varphi$ iff $\mathcal{M}, (i.j, t) \vDash \neg\varphi$ for at least one $j \in B_t(i)$ (by Definition 2.8.11).

⋄ Particle rule for A operator :
if $\langle \mathbf{X} - i, t : A\varphi \rangle \in d_\Delta$
then $\langle \mathbf{Y} - i, t : ?A_j \rangle \in d_\Delta$
so $\langle \mathbf{X} - i.j, t : \varphi \rangle \in d_\Delta$

1. if $\mathbf{X} = \mathbf{O}$, then by Hypothesis 7.3.2, $\mathcal{M}, (i, t) \vDash A\varphi$ iff $\mathcal{M}, (i.j, t) \vDash \varphi$ for every $j \in S$ (by Definition 2.8.11).

2. if $\mathbf{X} = \mathbf{P}$, then by Hypothesis 7.3.2, $\mathcal{M}, (i,t) \vDash \neg A\varphi$ iff $\mathcal{M}, (i.j,t) \vDash \neg\varphi$ for at least one $j \in S$ (by Definition 2.8.11).

⋄ Particle rule for I operator :
if $\langle \mathbf{X} - i, t : I\varphi \rangle \in d_\Delta$
then $\langle \mathbf{Y} - i, t : ?I_j \rangle \in d_\Delta$, or $\langle \mathbf{Y} - i, t : !I_j \rangle \in d_\Delta$
so $\langle \mathbf{X} - i.j, t : \varphi \rangle \in d_\Delta$, or $\langle \mathbf{X} - i, t : ?I_j^{*e} \rangle \in d_\Delta$

1. if $\mathbf{X} = \mathbf{O}$, $\mathbf{P}$ can change his challenge since $n := 2$, then by Hypothesis 7.3.2, $\mathcal{M}, (i,t) \vDash I\varphi$ iff $\mathcal{M}, (i.j,t) \vDash \varphi$ for every $j \in I_t(i)$; and $\mathcal{M}, (i.j,t) \vDash \varphi$ and $j \in I_t(i)$ (by Definition 2.8.11).

2. if $\mathbf{X} = \mathbf{P}$, $\mathbf{O}$ cannot change his challenge since $m := 1$, then by Hypothesis 7.3.2, $\mathcal{M}, (i,t) \vDash \neg I\varphi$ iff $\mathcal{M}, (i.j,t) \vDash \neg\varphi$ for at least one $j \in I_t(i)$; or $\mathcal{M}, (i.j,t) \vDash \varphi$ and $j \notin I_t(i)$ (by Definition 2.8.11).

□

7.3.5. Lemma. *The Proponent wins d_Δ iff he states an atomic formula or a context choice confirmation.*

7.3.6. Proof. Note that the structural rule **SR-3** states that a player wins iff he plays the last move of the play[9].

1. If the Proponent wins d_Δ, then his last move is $\langle \mathbf{P} - i, t : e \rangle$ such that e is an atomic formula or a context choice confirmation.

   7.3.7. Hypothesis. *The Proponent wins d_Δ such that the last move is $\langle \boldsymbol{P}-i, t : e \rangle$, where e is not an atomic formula or a context choice confirmation.*

   From Hypothesis 7.3.7 it follows that :

   (a) $\langle \mathbf{P} - i, t : e \rangle$ is a defence of the Proponent where e is a complex formula. Then the Opponent can challenge that formula, consequently the previous move of the Proponent is not the last one, contradicting the Hypothesis 7.3.7 ; or

   (b) $\langle \mathbf{P} - i, t : e \rangle$ is a challenge of the Proponent. Then the Opponent is always able to produce a defence (consequently the previous move of the Proponent is not the last one, contradicting the Hypothesis 7.3.7) unless it is a challenge against a negation (since there is no possible defence in that case). In that case :

[9] See structural rule **SR-3** in Section 7.2.

i. either e is an atomic formula, contradicting the Hypothesis 7.3.7,
ii. or e is not an atomic formula but a complex formula, and the Opponent is always able to produce a counter-attack (consequently the previous move of the Proponent is not the last one, contradicting the Hypothesis 7.3.7).

2. If the Proponent states an atomic formula or a context choice confirmation then he wins d_Δ with the corresponding move.

7.3.8. HYPOTHESIS. *The Proponent states an atomic formula or a context choice confirmation in a move $\alpha \in d_\Delta$, but he does not win d_Δ.*

From Hypothesis 7.3.8 it follows that :

(a) there exists a move β of the Opponent immediately following α (by **SR-1** and **SR-3**).

(b) Since there is no possible challenge on atomic formula or context choice confirmation, β cannot be a challenge on α.

(c) β cannot be a defence against α. Indeed α should be a challenge, consequently it would be a challenge on a negation (this is the only case where the challenge amounts to state a formula) and there is no possible defence in that case.

(d) Consequently β must be a challenge or a defence against a previous move of the Proponent.

◇ If β is a challenge :
  - there exists a move γ of the Proponent challenged by β.
  - After γ, the Opponent had the choice between β and the move δ immediately following γ in d_Δ.
  - But since $m := 1$ and δ is already a challenge on γ, β cannot be another challenge on this move : β cannot be a challenge on γ.

  Consequently, if β is a challenge on a previous move of the Proponent, this challenge is on a move ε preceding γ.

◇ If β is a defence :
  - there exists a move γ of the Proponent of which β is the defence.
  - After γ, the Opponent had the choice between β and the move δ immediately following γ in d_Δ.
  - But since $m := 1$ and δ is already a defence against γ, β cannot be another defence against this move : β cannot be a challenge against γ.

Consequently, if β is a defence against a previous move of the Proponent, this defence is against a move ε preceding γ.

This reasoning can be applied until the start of d_Δ showing that β cannot be a challenge or a defence against ε but has to be (on the same grounds as above) a challenge or a defence against a previous move ζ and so on. Finally, β would be a challenge on Δ or a defence against the first challenge of the Proponent but since $m := 1$, the actual choice and β cannot both belong to d_Δ.

Consequently, β cannot be a challenge or a defence against a previous move of the Proponent. In other words, (d) leads to a contradiction. Then it follows that (a) leads to a contradiction : there does not exist a move β following α, contradicting – in accordance with the structural rule **SR-3** – the defeat of the Proponent in d_Δ (Hypothesis 7.3.8).

□

Soundness Theorem

7.3.9. THEOREM. *If the Proponent wins d_Δ with the rules of DTDL, then Δ is a valid formula in L_{PLS*}.*

7.3.10. PROOF. We prove soundness by showing the contrapositive that is, we show that if there exists one $(\mathcal{M}, (i.t))$ such that $\neg\Delta$ is satisfiable in $(\mathcal{M}, (i, t))$, then the Proponent loses d_Δ. It follows from Lemma 7.3.3 and Lemma 7.3.5 that if $\neg\Delta$ is satisfiable then the Proponent loses :

7.3.11. HYPOTHESIS. *Let a play d_Δ be such that $\neg\Delta$ is satisfiable in $(\mathcal{M}, (i, t))$ and the Proponent wins d_Δ.*

1. By **SR-3** and Hypothesis 7.3.11, it follows that the Proponent plays the last move.

2. By (1) the Proponent plays the last move in d_Δ. By Lemma 7.3.5, the last move of the Proponent is an atomic formula or a context choice confirmation. By **SR-2**, the Proponent can state an atomic formula only if this atomic formula has been stated by the Opponent first. By **SR-5** and **SR-5.2** The Proponent can only state a context choice confirmation if the corresponding context has been chosen by the Opponent first to challenge a B or I operator.

3. From (2) and Definition 7.2.3, d_Δ is close.

4. By Lemma 7.3.3 et (3) it follows that it exists a branching time belief revision model $\mathcal{M}$ such that $\mathcal{M}, (i, t) \vDash p$ and $\mathcal{M}, (i, t) \vDash \neg p$, or such that $j \in I_t(i)$ and $j \notin I_t(i)$, which is a contradiction.

Consequently, if the Proponent wins d_Δ, there is no branching time belief revision model satisfying $\neg\Delta$.

□

7.3.2 Completeness

We prove that *DTDL* is complete with respect to L_{PLS*} showing that if Δ is valid in L_{PLS*} then the Proponent has a winning strategy in $\mathcal{D}_\Delta$ with the rules of *DTDL*. We prove the contrapositive : we prove that if the Proponent loses d_Δ with the rules of *DTDL* then Δ is not valid in L_{PLS*}.

Note that we still assume Hypothesis 7.3.1.

We start providing two definitions.

7.3.12. DEFINITION. An *extended dialogue* $\mathfrak{D}_\Delta$ is a play d_Δ where the Proponent can challenge modal operators as many times as he needs. In other words, repetition ranks do not concern modal operators anymore[10]. A branching time belief revision model $\mathcal{M}$ can be built from an extended dialogue: $\mathcal{M}$ is defined as $< T, \rightsquigarrow, S, \{B_t, I_t\}_{t\in T}, V >$, where[11]:

- $T = \{t$ such that $\langle \mathbf{X} - i, t : \varphi\rangle \in \mathfrak{D}_\Delta\}$, or $\langle \mathbf{X} - i, s : ?\Box_t\rangle \in \mathfrak{D}_\Delta\}$ for $\Box$ any kind of temporal operator $(\bigcirc, \bigcirc^{-1})$,
- $t^{\rightsquigarrow} = \{u$ such that $\langle \mathbf{X} - i, t : ?\bigcirc_u\rangle \in \mathfrak{D}_\Delta\}$,
- $S = \{i$ such that $\langle \mathbf{X} - i, t : \varphi\rangle \in \mathfrak{D}_\Delta\}$, or $\langle \mathbf{X} - h, t : ?\Box_i\rangle \in \mathfrak{D}_\Delta\}$ for $\Box$ any kind of non temporal operator (A, B, I),
- $B_t(i) = \{j$ such that $\langle \mathbf{X} - i, t : ?B_j\rangle \in \mathfrak{D}_\Delta\}$,
- $I_t(i) = \{j$ such that $\langle \mathbf{X} - i, t : ?I_j\rangle \in \mathfrak{D}_\Delta\}$ or $\langle \mathbf{O} - i, t : ?I_j^*\rangle \in \mathfrak{D}_\Delta\}$,
- $V_p = \{i$ such that $\langle \mathbf{O} - i, t : p\rangle \in \mathfrak{D}_\Delta\}$.

It seems that an extended dialogue can then be infinite. However, we noticed in Definition 2.8.7 that the set of states S of branching-time belief revision frames is finite as well as the set $t^{\rightsquigarrow}$ of all immediate successors of an instant t for all instants t. Then we can consider a finite number of choices of contextual points to challenge modal operators. So we can only consider finite extended dialogues.

[10]In Theorem 7.3.16, we show that if the Proponent can win, $n := 2$ is enough to win d_Δ that is, he does not need to challenge all contextual points.

[11]Note that we cast all the relations in terms of maps.

7.3.13. Definition. The *length* of φ is defined as:

- $len(p) = 1$
- $len(\neg\varphi) = 1 + len(\varphi)$
- $len(\varphi \wedge \psi) = 1 + len(\varphi) + len(\psi)$
- $len(\bigcirc^{-1}\varphi) = 2 + len(\varphi)$
- $len(\bigcirc\varphi) = 2 + len(\varphi)$
- $len(B\varphi) = 1 + len(\varphi)$
- $len(I\varphi) = 1 + len(\varphi)$
- $len(A\varphi) = 1 + len(\varphi)$

7.3.14. Lemma. *If $\mathfrak{D}_\Delta$ is terminal and the Proponent loses $\mathfrak{D}_\Delta$, then it exists a model $(\mathcal{M}, (i,t))$ such that:*

- $\langle \mathbf{O} - i, t : \varphi \rangle \in \mathfrak{D}_\Delta$ means $\mathcal{M}, (i,t) \vDash \varphi$, and
- $\langle \mathbf{P} - i, t : \varphi \rangle \in \mathfrak{D}_\Delta$ means $\mathcal{M}, (i,t) \vDash \neg\varphi$.
- $\langle \mathbf{O} - i, t : ?I_j^* \rangle \in \mathfrak{D}_\Delta$ means $j \in I_t(i)$, and
- $\langle \mathbf{P} - i, t : ?I_j^* \rangle \in \mathfrak{D}_\Delta$ means $j \notin I_t(i)$.

7.3.15. Proof. We proceed by induction on the length of φ. The basic case is about atomic formulas.

1. Base: $\varphi := p$
If $\langle \mathbf{X} - (i,t) : p \rangle \in \mathfrak{D}_\Delta$, then either:

1. $\mathbf{X} = \mathbf{O}$ and $\mathcal{M}, (i,t) \vDash p$ (Definition 7.3.12); or
2. $\mathbf{X} = \mathbf{P}$ and consequently, in accordance with Lemma 7.3.5, the Proponent wins $\mathfrak{D}_\Delta$ since he states an atomic formula, contradicting Lemma 7.3.14.

2. Induction Hypothesis:
If $len(\varphi) \leq n$ then if $\langle \mathbf{X} - i, t : \varphi \rangle \in \mathfrak{D}_\Delta$ and the Proponent loses $\mathfrak{D}_\Delta$, there exists a model $(\mathcal{M}, (i,t))$ such that:

- $\langle \mathbf{O} - i, t : \varphi \rangle \in \mathfrak{D}_\Delta$ means $\mathcal{M}, (i,t) \vDash \varphi$, and
- $\langle \mathbf{P} - i, t : \varphi \rangle \in \mathfrak{D}_\Delta$ means $\mathcal{M}, (i,t) \vDash \neg\varphi$.

– $\langle \mathbf{O} - i, t : ?I_j^* \rangle \in \mathfrak{D}_\Delta$ means $j \in I_t(i)$, and

– $\langle \mathbf{P} - i, t : ?I_j^* \rangle \in \mathfrak{D}_\Delta$ means $j \notin I_t(i)$.

3. Inductive step :
Let us assume that $len(\varphi) = n + 1$. We consider 8 different cases, one for each logical constant in our dialogical language.

Case 1 : $\varphi := \neg\psi$
If $\langle \mathbf{X} - i, t : \neg\psi \rangle \in \mathfrak{D}_\Delta$, then :
$\langle \mathbf{Y} - i, t : \psi \rangle \in \mathfrak{D}_\Delta$.

1. $\mathbf{Y} = \mathbf{O}$ then $\mathcal{M}, (i,t) \vDash \psi$ (by Induction Hypothesis – H. I.) ; or
2. $\mathbf{Y} = \mathbf{P}$ then either :
   (a) $\psi \notin \Phi$: $\langle \mathbf{P} - i, t : \psi \rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (i,t) \vDash \neg\psi$ (by H. I.) ; or
   (b) $\psi \in \Phi$:
      i. $\langle \mathbf{O} - i, z : \psi \rangle \notin \mathfrak{D}_\Delta$ for any contextual point z,[12] so $i \notin V_\psi$ then $\mathcal{M}, (i,t) \vDash \neg\psi$ (by Definition 7.3.12),
      ii. $\langle \mathbf{O} - i, z : \psi \rangle \in \mathfrak{D}_\Delta$ and $\langle \mathbf{P} - i, t : \psi \rangle \in \mathfrak{D}_\Delta$, then by Lemma 7.3.5 the Proponent wins $\mathfrak{D}_\Delta$, contradicting the hypothesis of Lemma 7.3.14.

Case 2 : $\varphi := \psi \wedge \chi$
If $\langle \mathbf{X} - i, t : \psi \wedge \chi \rangle \in \mathfrak{D}_\Delta$, then :
$\langle \mathbf{Y} - i, t : ?_{\wedge 1} \rangle \in \mathfrak{D}_\Delta$ or $\langle \mathbf{Y} - i, t : ?_{\wedge 2} \rangle \in \mathfrak{D}_\Delta$.

1. If $\mathbf{X} = \mathbf{O}$, the Proponent can change his challenge since $n := 2$. Consequently :
   $\langle \mathbf{O} - i, t : \psi \rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (i,t) \vDash \psi$ (by H. I.) ; and
   $\langle \mathbf{O} - i, t : \chi \rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (i,t) \vDash \chi$ (by H. I.)
   iff $\mathcal{M}, (i,t) \vDash \psi \wedge \chi$ (by Definition 2.8.11).
2. If $\mathbf{X} = \mathbf{P}$, $\mathbf{O}$ can only challenge once since $m := 1$. We only deal with the case where the Opponent challenges the first conjunct[13].
   (a) If $\psi \notin \Phi$ then :
   $\langle \mathbf{P} - i, t : \psi \rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (i,t) \vDash \neg\psi$ (by H. I.)
   iff $\mathcal{M}, (i,t) \vDash \neg(\psi \wedge \chi)$ (by Definition 2.8.11).

[12]Indeed it is possible that $z \neq t$.
[13]The same reasoning can be applied in the case where the Opponent challenges the second conjunct.

(b) If $\psi \in \Phi$ then either:

i. $\langle \mathbf{O} - i, z : \psi \rangle \notin \mathfrak{D}_\Delta$ for any contextual point z, so $i \notin V_\psi$ then $\mathcal{M}, (i,t) \vDash \neg\psi$ (by Definition 7.3.12)
iff $\mathcal{M}, (i,t) \vDash \neg(\psi \wedge \chi)$ (by Definition 2.8.11); or

ii. $\langle \mathbf{O} - i, z : \psi \rangle \in \mathfrak{D}_\Delta$ and $\langle \mathbf{P} - i, t : \psi \rangle \in \mathfrak{D}_\Delta$, then by Lemma 7.3.5 the Proponent wins $\mathfrak{D}_\Delta$, contradicting the hypothesis of Lemma 7.3.14.

Case 3: $\varphi := \psi \vee \chi$
If $\langle \mathbf{X} - i, t : \psi \vee \chi \rangle \in \mathfrak{D}_\Delta$, then:
$\langle \mathbf{Y} - i, t : ?_\vee \rangle \in \mathfrak{D}_\Delta$.

1. If $\mathbf{X} = \mathbf{O}$, $\mathbf{O}$ can only produce one defence since $m := 1$. Consequently:

   $\langle \mathbf{O} - i, t : \psi \rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (i,t) \vDash \psi$ (by H. I.); or

   $\langle \mathbf{O} - i, t : \chi \rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (i,t) \vDash \chi$ (by H. I.)

   iff $\mathcal{M}, (i,t) \vDash (\psi \vee \chi)$ (by Definition 2.8.11).

2. If $\mathbf{X} = \mathbf{P}$, the Proponent can change his defence since $n := 2$.

   (a) If $\psi \notin \Phi$ and $\chi \notin \Phi$ then:
   $\langle \mathbf{P} - i, t : \psi \rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (i,t) \vDash \neg\psi$ (by H. I.); and
   $\langle \mathbf{P} - i, t : \chi \rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (i,t) \vDash \neg\chi$ (by H. I.)
   iff $\mathcal{M}, (i,t) \vDash \neg(\psi \vee \chi)$ (by Definition 2.8.11).

   (b) If $\psi \in \Phi$ and $\chi \notin \Phi$ then either:

   i. $\langle \mathbf{O} - i, z : \psi \rangle \notin \mathfrak{D}_\Delta$ for any contextual point z, so $i \notin V_\psi$ then $\mathcal{M}, (i,t) \vDash \neg\psi$ (by Definition 7.3.12) and
   $\langle \mathbf{P} - i, t : \chi \rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (i,t) \vDash \neg\chi$ (by H. I.)
   iff $\mathcal{M}, (i,t) \vDash \neg(\psi \vee \chi)$ (by Definition 2.8.11); or

   ii. $\langle \mathbf{O} - i, z : \psi \rangle \in \mathfrak{D}_\Delta$ and $\langle \mathbf{P} - i, t : \psi \rangle \in \mathfrak{D}_\Delta$, then by Lemma 7.3.5 the Proponent wins $\mathfrak{D}_\Delta$, contradicting the hypothesis of Lemma 7.3.14.[14]

   (c) If $\psi \in \Phi$ and $\chi \in \Phi$ then either:

   i. $\langle \mathbf{O} - i, z : \psi \rangle \notin \mathfrak{D}_\Delta$ for any contextual point z, so $i \notin V_\psi$ then $\mathcal{M}, (i,t) \vDash \neg\psi$ (by Definition 7.3.12)
   iff $\mathcal{M}, (i,t) \vDash \neg(\psi \vee \chi)$ (by Definition 2.8.11); or

   ii. $\langle \mathbf{O} - i, z : \psi \rangle \in \mathfrak{D}_\Delta$ and $\langle \mathbf{P} - i, t : \psi \rangle \in \mathfrak{D}_\Delta$, then by Lemma 7.3.5 the Proponent wins $\mathfrak{D}_\Delta$, contradicting the hypothesis of Lemma 7.3.14.[15]

---

[14] The same reasoning can be applied in the case where $\psi \notin \Phi$ and $\chi \in \Phi$.

[15] We only show the reasoning for ψ. The same reasoning can be applied for χ.

Case 4 : $\varphi := \bigcirc^{-1}\psi$

If $\langle \mathbf{X} - i, t : \bigcirc^{-1}\psi\rangle \in \mathfrak{D}_\Delta$, then:

$\langle \mathbf{Y} - i, t : ?\bigcirc^{-1}_u\rangle \in \mathfrak{D}_\Delta$ for all contextual points u.

1. If $\mathbf{X} = \mathbf{O}$, then:

   $\langle \mathbf{O} - i, t.u : \psi\rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (i,u) \vDash \psi$ (by H. I.). By Hypothesis Lemma 7.3.14, $\mathfrak{D}_\Delta$ is terminal, consequently:

   $\langle \mathbf{O} - i, t.v : \psi\rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (i,v) \vDash \psi$ (by H. I.), and

   $\vdots$

   $\langle \mathbf{O} - i, t.w : \psi\rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (i,w) \vDash \psi$ (by H. I.)

   for all contextual points u that respect **SR-4, SR-4.1** and **SR-4.2**

   iff $\mathcal{M}, (i,t) \vDash \bigcirc^{-1}\psi$ (by Definition 2.8.11).

2. If $\mathbf{X} = \mathbf{P}$, then:

   (a) If $\psi \notin \Phi$ then:
   $\langle \mathbf{P} - i, t.u : \psi\rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (i,u) \vDash \neg\psi$ (by H. I.)
   for at least one contextual point u that respect **SR-4**
   iff $\mathcal{M}, (i,t) \vDash \neg\bigcirc^{-1}\psi$ (by Definition 2.8.11).

   (b) If $\psi \in \Phi$ then either:

      i. $\langle \mathbf{O} - i, z : \psi\rangle \notin \mathfrak{D}_\Delta$ for any contextual point z,[16] so $i \notin V_\psi$ then $\mathcal{M}, (i,u) \vDash \neg\psi$ (by Definition 7.3.12)
      iff $\mathcal{M}, (i,t) \vDash \neg\bigcirc^{-1}\psi$ (by Definition 2.8.11); or

      ii. $\langle \mathbf{O} - i, z : \psi\rangle \in \mathfrak{D}_\Delta$ and $\langle \mathbf{P} - i, t.u : \psi\rangle \in \mathfrak{D}_\Delta$ then by Lemma 7.3.5 the Proponent wins $\mathfrak{D}_\Delta$, contradicting the hypothesis of Lemma 7.3.14.

Case 5 : $\varphi := \bigcirc\psi$

If $\langle \mathbf{X} - i, t : \bigcirc\psi\rangle \in \mathfrak{D}_\Delta$, then:

$\langle \mathbf{Y} - i, t : ?\bigcirc_u\rangle \in \mathfrak{D}_\Delta$ for all contextual points u.

1. If $\mathbf{X} = \mathbf{O}$, then:

   $\langle \mathbf{O} - i, t.u : \psi\rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (i,u) \vDash \psi$ (by H. I.). By Hypothesis Lemma 7.3.14, $\mathfrak{D}_\Delta$ is terminal, consequently:

   $\langle \mathbf{O} - i, t.v : \psi\rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (i,v) \vDash \psi$ (by H. I.), and

   $\vdots$

   $\langle \mathbf{O} - i, t.w : \psi\rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (i,w) \vDash \psi$ (by H. I.)

   for all contextual points u that respect **SR-4, SR-4.1** and **SR-4.2**

   iff $\mathcal{M}, (i,t) \vDash \bigcirc\psi$ (by Definition 2.8.11).

[16] Indeed it is possible that $z \neq u$.

2. If $\mathbf{X} = \mathbf{P}$, then:

   (a) If $\psi \notin \Phi$ then:
   $\langle \mathbf{P} - i, t.u : \psi \rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (i, u) \vDash \neg\psi$ (by H. I.)
   for at least one contextual point u that respect **SR-4**
   iff $\mathcal{M}, (i, t) \vDash \neg\bigcirc\psi$ (by Definition 2.8.11).

   (b) If $\psi \in \Phi$ then either:

      i. $\langle \mathbf{O} - i, z : \psi \rangle \notin \mathfrak{D}_\Delta$ for any contextual point z, so $i \notin V_\psi$ then $\mathcal{M}, (i, u) \vDash \neg\psi$ (by Definition 7.3.12)
      iff $\mathcal{M}, (i, t) \vDash \neg\bigcirc\psi$ (by Definition 2.8.11); or

      ii. $\langle \mathbf{O} - i, z : \psi \rangle \in \mathfrak{D}_\Delta$ and $\langle \mathbf{P} - i, t.u : \psi \rangle \in \mathfrak{D}_\Delta$ then by Lemma 7.3.5 the Proponent wins $\mathfrak{D}_\Delta$, contradicting the hypothesis of Lemma 7.3.14.

Cas 6 : $\varphi := B\psi$
If $\langle \mathbf{X} - i, t : B\psi \rangle \in \mathfrak{D}_\Delta$ then:
$\langle \mathbf{Y} - i, t : ?B_j \rangle \in \mathfrak{D}_\Delta$ for all contextual points j.

1. If $\mathbf{X} = \mathbf{O}$, then:

   $\langle \mathbf{O} - i.j, t : \psi \rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (j, t) \vDash \psi$ (by H. I.). By Hypothesis of Lemma 7.3.14, $\mathfrak{D}_\Delta$ is terminal, consequently:

   $\langle \mathbf{O} - i.k, t : \psi \rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (k, t) \vDash \psi$ (by H. I.), and

   $\vdots$

   $\langle \mathbf{O} - i.l, t : \psi \rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (l, t) \vDash \psi$ (by H. I.)

   for all contextual points j that respect **SR-5, SR-5.4, SR-5.5, SR-5.6, SR-5.7** and **SR-5.8**

   iff $\mathcal{M}, (i, t) \vDash B\psi$ (by Definition 2.8.11).

2. If $\mathbf{X} = \mathbf{P}$, then:

   (a) If $\psi \notin \Phi$ then:
   $\langle \mathbf{P} - i.j, t : \psi \rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (j, t) \vDash \neg\psi$ (by H. I.)
   for at least one contextual point j that respect **SR-5**
   iff $\mathcal{M}, (i, t) \vDash \neg B\psi$ (by Definition 2.8.11).

   (b) If $\psi \in \Phi$ then either:

      i. $\langle \mathbf{O} - i.j, z : \psi \rangle \notin \mathfrak{D}_\Delta$ for any contextual point z,[17] so $j \notin V_\psi$ then $\mathcal{M}, (j, t) \vDash \neg\psi$ (by Definition 7.3.12)
      iff $\mathcal{M}, (i, t) \vDash \neg B\psi$ (by Definition 2.8.11); or

[17] Indeed it is possible that $z \neq t$.

ii. $\langle \mathbf{O} - i.j, z : \psi \rangle \in \mathfrak{D}_\Delta$ and $\langle \mathbf{P} - i.j, t : \psi \rangle \in \mathfrak{D}_\Delta$ then by Lemma 7.3.5 the Proponent wins $\mathfrak{D}_\Delta$, contradicting the hypothesis of Lemma 7.3.14.

Cas 7 : $\varphi := A\psi$

If $\langle \mathbf{X} - i, t : A\psi \rangle \in \mathfrak{D}_\Delta$ then :
$\langle \mathbf{Y} - i, t : ?A_j \rangle \in \mathfrak{D}_\Delta$ for all contextual points j.

1. If $\mathbf{X} = \mathbf{O}$, then :

   $\langle \mathbf{O} - i.j, t : \psi \rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (j, t) \models \psi$ (by H. I.). By Hypothesis of Lemma 7.3.14, $\mathfrak{D}_\Delta$ is terminal, consequently :

   $\langle \mathbf{O} - i.k, t : \psi \rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (k, t) \models \psi$ (by H. I.), and

   $\vdots$

   $\langle \mathbf{O} - i.l, t : \psi \rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (l, t) \models \psi$ (by H. I.)

   for all contextual points j that respect **SR-5, SR-5.1, SR-5.3**

   iff $\mathcal{M}, (i, t) \models A\psi$ (by Definition 2.8.11).

2. If $\mathbf{X} = \mathbf{P}$, then :

   (a) If $\psi \notin \Phi$ then :
   $\langle \mathbf{P} - i.j, t : \psi \rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (j, t) \models \neg\psi$ (by H. I.)
   for at least one contextual point j that respect **SR-5**
   iff $\mathcal{M}, (i, t) \models \neg A\psi$ (by Definition 2.8.11).

   (b) If $\psi \in \Phi$ then either :

   i. $\langle \mathbf{O} - i.j, z : \psi \rangle \notin \mathfrak{D}_\Delta$ for any contextual point z, so $j \notin V_\psi$ then $\mathcal{M}, (j, t) \models \neg\psi$ (by Definition 7.3.12)
   iff $\mathcal{M}, (i, t) \models \neg A\psi$ (by Definition 2.8.11) ; or

   ii. $\langle \mathbf{O} - i.j, z : \psi \rangle \in \mathfrak{D}_\Delta$ and $\langle \mathbf{P} - i.j, t : \psi \rangle \in \mathfrak{D}_\Delta$ then by Lemma 7.3.5 the Proponent wins $\mathfrak{D}_\Delta$, contradicting the hypothesis of Lemma 7.3.14.

Cas 8 : $\varphi := I\psi$

If $\langle \mathbf{X} - i, t : I\psi \rangle \in \mathfrak{D}_\Delta$ then :
$\langle \mathbf{Y} - i, t : ?I_j \rangle \in \mathfrak{D}_\Delta$ or $\langle \mathbf{Y} - i, t : !I_j \rangle \in \mathfrak{D}_\Delta$ for all contextual points j.

1. If $\mathbf{X} = \mathbf{O}$, the Proponent can change his challenge since $n := 2$. Consequently :

   $\langle \mathbf{O} - i.j, t : \psi \rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (j, t) \models \psi$ (by H. I.). By Hypothesis of Lemma 7.3.14, $\mathfrak{D}_\Delta$ is terminal, consequently :

   $\langle \mathbf{O} - i.k, t : \psi \rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (k, t) \models \psi$ (by H. I.), and

   $\vdots$

$\langle \mathbf{O} - i.l, t : \psi \rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (l,t) \models \psi$ (by H. I.)

for all contextual points j that respect **SR-5, SR-5.2** ; and

$\langle \mathbf{O} - i, t : ?I^*_j \rangle \in \mathfrak{D}_\Delta$ then $j \in I_t(i)$ (by H. I.)

for all contextual points j that respect **SR-5, SR-5.2, SR-5.X** that is, $\mathcal{M}, (i.j, t) \models \psi$

iff $\mathcal{M}, (i,t) \models I\psi$ (by Definition 2.8.11).

2. If $\mathbf{X} = \mathbf{P}$, **O** can only challenge once since $m := 1$.

   (a) If $\psi \notin \Phi$ then :
   $\langle \mathbf{P} - i.j, t : \psi \rangle \in \mathfrak{D}_\Delta$ then $\mathcal{M}, (j,t) \models \neg\psi$ (by H. I.)
   for at least one contextual point j that respect **SR-5**
   iff $\mathcal{M}, (i,t) \models \neg I\psi$ (by Definition 2.8.11).

   (b) If $\psi \in \Phi$ then either :

      i. $\langle \mathbf{O} - i.j, z : \psi \rangle \notin \mathfrak{D}_\Delta$ for any contextual point z, so $j \notin V_\psi$ then $\mathcal{M}, (j,t) \models \neg\psi$ (by Definition 7.3.12)
      iff $\mathcal{M}, (i,t) \models \neg I\psi$ (by Definition 2.8.11) ; or

      ii. $\langle \mathbf{O} - i.j, z : \psi \rangle \in \mathfrak{D}_\Delta$ and $\langle \mathbf{P} - i.j, t : \psi \rangle \in \mathfrak{D}_\Delta$ then by Lemma 7.3.5 the Proponent wins $\mathfrak{D}_\Delta$, contradicting the hypothesis of Lemma 7.3.14 ; or

   (a) $\langle \mathbf{O} - i, t : ?I_j \rangle \notin \mathfrak{D}_\Delta$ and $\langle \mathbf{O} - i, t : ?B_j \rangle \notin \mathfrak{D}_\Delta$ and $\langle \mathbf{O} - i, t : ?I^*_j \rangle \notin \mathfrak{D}_\Delta$ then $j \notin I_t(i)$ for any contextual points j that respect **SR-5, SR-5.X** that is, $\mathcal{M}, (i.j, t) \models \psi$ (by Definition 7.3.12)
   iff $\mathcal{M}, (i,t) \models \neg I\psi$ (by Definition 2.8.11) ; or

   (b) $\langle \mathbf{O} - i, t : ?I_j \rangle \in \mathfrak{D}_\Delta$ or $\langle \mathbf{O} - i, t : ?B_j \rangle \in \mathfrak{D}_\Delta$ or $\langle \mathbf{O} - i, t : ?I^*_j \rangle \in \mathfrak{D}_\Delta$, and $\langle \mathbf{P} - i, t : ?I^*_j \rangle \in \mathfrak{D}_\Delta$ then by Lemma 7.3.5 the Proponent wins $\mathfrak{D}_\Delta$, contradicting the hypothesis of Lemma 7.3.14.

□

In Lemma 7.3.14, we assumed players can change their challenges or defences on modal operators as many time as needed. This allows to establish a symmetric link between the moves in $\mathfrak{D}_\Delta$ and a model $\mathcal{M}$ satisfying Δ since it allows to check every possible situations of a model. In the next theorem, we show that it is enough to consider a play d_Δ where $m := 1$ and $n := 2$.

7.3.16. THEOREM. *The repetition ranks $m := 1$ and $n := 2$ are enough to check if there exists a winning strategy for the Proponent in $\mathfrak{D}_\Delta$.*

7.3.17. PROOF. We proceed by induction on the length of φ in the scope of modal operators.

1. $\bigcirc^{-1}$ operator :
   (a) **Base :** We show that if the Proponent wins $\mathfrak{D}_{\Delta}$, then a repetition rank 1 is enough for a challenge on a formula $\bigcirc^{-1}\varphi$ of the Opponent where $len(\varphi) = 1$:
   $\langle \mathbf{O} - i, t : \bigcirc^{-1}\varphi \rangle \in \mathfrak{D}_{\Delta}$
   $\langle \mathbf{P} - i, t : ?\bigcirc_u^{-1} \rangle \in \mathfrak{D}_{\Delta}$ or $\langle \mathbf{P} - i, t : ?\bigcirc_v^{-1} \rangle \in \mathfrak{D}_{\Delta}$ for any contextual points u and v. However these contextual points have to respect **SR-4, SR-4.1** and **SR-4.2** that is, $u = v$. The Proponent can only choose one single instant to challenge a P operator whatever his repetition rank is.
   (b) **Induction Hypothesis :** If the Proponent wins $\mathfrak{D}_{\Delta}$ then a repetition rank 1 is enough for a challenge on a formula $\bigcirc^{-1}\varphi$ of the Opponent if $len(\varphi) = n$.
   (c) **Inductive step :** The same reasoning as the basic case can be applied to show that if the Proponent wins $\mathfrak{D}_{\Delta}$, a repetition rank 1 is enough for a challenge on a formula $\bigcirc^{-1}\varphi$ of the Opponent where $len(\varphi) \geq n + 1$.

2. $\bigcirc$ operator :
   (a) **Base :** We show that if the Proponent wins $\mathfrak{D}_{\Delta}$, then a repetition rank 1 is enough for a challenge on a formula $\bigcirc\varphi$ of the Opponent where $len(\varphi) = 1$:
   $\langle \mathbf{O} - i, t : \bigcirc\varphi \rangle \in \mathfrak{D}_{\Delta}$
   $\langle \mathbf{P} - i, t : ?\bigcirc_u \rangle \in \mathfrak{D}_{\Delta}$ or $\langle \mathbf{P} - i, t : ?\bigcirc_v \rangle \in \mathfrak{D}_{\Delta}$ for any contextual points u and v, so
   $\langle \mathbf{O} - i, t.u : \varphi \rangle \in \mathfrak{D}_{\Delta}$ or $\langle \mathbf{O} - i, t.v : \varphi \rangle \in \mathfrak{D}_{\Delta}$. The Proponent can then state the atomic formula φ in the contextual point i at every instant (by **SR-2**). The Proponent only needs to challenge once whatever the contextual point he chooses.
   (b) **Induction Hypothesis :** If the Proponent wins $\mathfrak{D}_{\Delta}$ then a repetition rank 1 is enough for a challenge on a formula $\bigcirc\varphi$ of the Opponent if $len(\varphi) = n$.
   (c) **Inductive step :** We show that if the Proponent wins $\mathfrak{D}_{\Delta}$, then a repetition rank 1 is enough for a challenge on a formula $\bigcirc\varphi$ of the Opponent where $len(\varphi) \geq n + 1$:
   $\langle \mathbf{O} - i, t : \bigcirc\varphi \rangle \in \mathfrak{D}_{\Delta}$

$\langle \mathbf{P} - i, t : ?\bigcirc_u \rangle \in \mathfrak{D}_\Delta$ or $\langle \mathbf{P} - i, t : ?\bigcirc_v \rangle \in \mathfrak{D}_\Delta$ for any contextual points u and v, so

$\langle \mathbf{O} - i, t.u : \varphi \rangle \in \mathfrak{D}_\Delta$ or $\langle \mathbf{O} - i, t.v : \varphi \rangle \in \mathfrak{D}_\Delta$. Since the Proponent wins $\mathfrak{D}_\Delta$, we only need to consider the choice u or v leading him to win.

3. B operator :

    (a) **Base :** We show that if the Proponent wins $\mathfrak{D}_\Delta$, then a repetition rank 1 is enough for a challenge on a formula $B\varphi$ of the Opponent where $len(\varphi) = 1$:
    $\langle \mathbf{O} - i, t : B\varphi \rangle \in \mathfrak{D}_\Delta$
    $\langle \mathbf{P} - i, t : ?B_j \rangle \in \mathfrak{D}_\Delta$ or $\langle \mathbf{P} - i, t : ?B_k \rangle \in \mathfrak{D}_\Delta$ for any contextual points j and k, so
    $\langle \mathbf{O} - i.j, t : \varphi \rangle \in \mathfrak{D}_\Delta$ or $\langle \mathbf{O} - i.k, t : \varphi \rangle \in \mathfrak{D}_\Delta$. The Proponent can then state the atomic formula φ in the contextual point j or k at any instant (by **SR-2**). Since he wins $\mathfrak{D}_\Delta$, we only need to consider the choice j or k leading him to win.

    (b) **Induction Hypothesis :** If the Proponent wins $\mathfrak{D}_\Delta$ then a repetition rank 1 is enough for a challenge on a formula $B\varphi$ of the Opponent if $len(\varphi) = n$.

    (c) **Inductive step :** We show that if the Proponent wins $\mathfrak{D}_\Delta$, then a repetition rank 1 is enough for a challenge on a formula $B\varphi$ of the Opponent where $len(\varphi) \geq n + 1$:
    $\langle \mathbf{O} - i, t : B\varphi \rangle \in \mathfrak{D}_\Delta$
    $\langle \mathbf{P} - i, t : ?B_j \rangle \in \mathfrak{D}_\Delta$ or $\langle \mathbf{P} - i, t : ?B_k \rangle \in \mathfrak{D}_\Delta$ for any contextual points j and k, so
    $\langle \mathbf{O} - i.j, t : \varphi \rangle \in \mathfrak{D}_\Delta$ or $\langle \mathbf{O} - i.k, t : \varphi \rangle \in \mathfrak{D}_\Delta$. Since the Proponent wins $\mathfrak{D}_\Delta$, we only need to consider the choice j or k leading him to win.

4. A operator :

    (a) **Base :** We show that if the Proponent wins $\mathfrak{D}_\Delta$, then a repetition rank 1 is enough for a challenge on a formula $A\varphi$ of the Opponent where $len(\varphi) = 1$:
    $\langle \mathbf{O} - i, t : A\varphi \rangle \in \mathfrak{D}_\Delta$
    $\langle \mathbf{P} - i, t : ?A_j \rangle \in \mathfrak{D}_\Delta$ or $\langle \mathbf{P} - i, t : ?A_k \rangle \in \mathfrak{D}_\Delta$ for any contextual points j and k, so
    $\langle \mathbf{O} - i.j, t : \varphi \rangle \in \mathfrak{D}_\Delta$ or $\langle \mathbf{O} - i.k, t : \varphi \rangle \in \mathfrak{D}_\Delta$. Then Proponent can then state the atomic formula φ in the contextual point j or k at any instant (by **SR-2**). Since he wins $\mathfrak{D}_\Delta$, we only need to consider the choice j or k leading him to win.

(b) **Induction Hypothesis :** If the Proponent wins $\mathfrak{D}_\Delta$ then a repetition rank 1 is enough for a challenge on a formula $A\varphi$ of the Opponent if $len(\varphi) = n$.

(c) **Inductive step :** We show that if the Proponent wins $\mathfrak{D}_\Delta$, then a repetition rank 1 is enough for a challenge on a formula $A\varphi$ of the Opponent where $len(\varphi) \geq n + 1$:
$\langle \mathbf{O} - i, t : A\varphi \rangle \in \mathfrak{D}_\Delta$
$\langle \mathbf{P} - i, t : ?A_j \rangle \in \mathfrak{D}_\Delta$ or $\langle \mathbf{P} - i, t : ?A_k \rangle \in \mathfrak{D}_\Delta$ for any contextual points j and k, so
$\langle \mathbf{O} - i.j, t : \varphi \rangle \in \mathfrak{D}_\Delta$ or $\langle \mathbf{O} - i.k, t : \varphi \rangle \in \mathfrak{D}_\Delta$. Since the Proponent wins $\mathfrak{D}_\Delta$, we only need to consider the choice j or k leading him to win.

5. I operator :

(a) **Base :** We show that if the Proponent wins $\mathfrak{D}_\Delta$, then a repetition rank 1 is enough for a challenge on a formula $I\varphi$ of the Opponent where $len(\varphi) = 1$:
$\langle \mathbf{O} - i, t : I\varphi \rangle \in \mathfrak{D}_\Delta$
$\langle \mathbf{P} - i, t : ?I_j \rangle \in \mathfrak{D}_\Delta$ or $\langle \mathbf{P} - i, t : ?I_k \rangle \in \mathfrak{D}_\Delta$, or $\langle \mathbf{P} - i, t : !I_j \rangle \in \mathfrak{D}_\Delta$ or $\langle \mathbf{P} - i, t : !I_k \rangle \in \mathfrak{D}_\Delta$ for any contextual points j and k, so
$\langle \mathbf{O} - i.j, t : \varphi \rangle \in \mathfrak{D}_\Delta$ or $\langle \mathbf{O} - i.k, t : \varphi \rangle \in \mathfrak{D}_\Delta$, or $\langle \mathbf{O} - i, t : ?I^*_j \rangle \in \mathfrak{D}_\Delta$ or $\langle \mathbf{O} - i, t : ?I^*_k \rangle \in \mathfrak{D}_\Delta$. The Proponent can then state the atomic formula φ in the contextual point j or k at any instant (by **SR-2**), or he can then use the contextual point j or k at t to challenge a I operator. Since he wins $\mathfrak{D}_\Delta$, we only need to consider the choice leading him to win.

(b) **Induction Hypothesis :** If the Proponent wins $\mathfrak{D}_\Delta$ then a repetition rank 1 is enough for a challenge on a formula $I\varphi$ of the Opponent if $len(\varphi) = n$.

(c) **Inductive step :** We show that if the Proponent wins $\mathfrak{D}_\Delta$, then a repetition rank 1 is enough for a challenge on a formula $I\varphi$ of the Opponent where $len(\varphi) \geq n + 1$:
$\langle \mathbf{O} - i, t : I\varphi \rangle \in \mathfrak{D}_\Delta$
$\langle \mathbf{P} - i, t : ?I_j \rangle \in \mathfrak{D}_\Delta$ or $\langle \mathbf{P} - i, t : ?I_k \rangle \in \mathfrak{D}_\Delta$, or $\langle \mathbf{P} - i, t : !I_j \rangle \in \mathfrak{D}_\Delta$ or $\langle \mathbf{P} - i, t : !I_k \rangle \in \mathfrak{D}_\Delta$ for any contextual points j and k, so
$\langle \mathbf{O} - i.j, t : \varphi \rangle \in \mathfrak{D}_\Delta$ or $\langle \mathbf{O} - i.k, t : \varphi \rangle \in \mathfrak{D}_\Delta$, or $\langle \mathbf{O} - i, t : ?I^*_j \rangle \in \mathfrak{D}_\Delta$ or $\langle \mathbf{O} - i, t : ?I^*_k \rangle \in \mathfrak{D}_\Delta$. Since the Proponent wins $\mathfrak{D}_\Delta$, we only need to consider the choice leading him to win.

□

If the Proponent wins $\mathfrak{D}_{\Delta}$, a repetition rank 1 is enough for a challenge on a modal formula $\Box\varphi$ such that $\Box$ is any $(\bigcirc^{-1}, \bigcirc, B, A, I)$ of the Opponent, whatever the length of φ is. A repetition rank 2 is only required for Proponent to challenge a conjunction or change his defence against disjunction. Then we can deal with d_{Δ} instead of $\mathfrak{D}_{\Delta}$ for the completeness theorem.

Completeness Theorem

7.3.18. THEOREM. *If Δ is a valid formula in $\mathcal{L}_{PLS*}$, then the Proponent wins d_{Δ} with the rules of DTDL.*

7.3.19. PROOF. We prove completeness by showing the contrapositive that is, we show that if the Proponent loses d_{Δ} then Δ is not a valid formula in L_{PLS*}. By Lemma 7.3.14 and Theorem 7.3.16, if the Proponent loses d_{Δ}, then there exists a branching time belief revision model $(\mathcal{M}, (i.t))$ satisfying $\neg\Delta$. Consequently, there exists a model $(\mathcal{M}, (i.t))$ such that Δ is not satisfiable in $(\mathcal{M}, (i.t))$ and then, Δ is not a valid formula in L_{PLS*}. $\Box$

Conclusion

DTDL allows an argumentative interpretation of the belief revision policy of Bonanno. The notion of belief and information as well as their relation is then interpreted in terms of choice in a dialogical framework. We pointed out the interpretation of the interplay between information and beliefs as well as the interplay between the beliefs themselves – namely, initial beliefs and revised beliefs – as the interplay between the choices of the players. In particular we showed the interplay between the choices of contextual points to challenge belief and/or information operators. We noticed that this is the interplay that defines a particular belief revision policy, namely the belief revision policy of Bonanno. Finally, we underlined the originality of the interpretation of the information operator through the notion of choice. Indeed this notion of choice is directly implemented in the dialogue in the sense that players discuss explicitly the choices they can/should make with respect to their previous arguments.

Part IV

Conclusion

Chapter 8

Conclusion

Answering the questions

In the first chapter we have provided the reader with a list of questions, which have guided the work in this thesis. Some of these questions have received clear answers while others have generated more refined questions or even new questions. Looking back, it is time to take stock of what we have achieved so far.

In Part I, Chapter 2 we explored the literature on belief revision theory and presented the reader with an overview of different ways in which one can model belief change. We introduced the first formal (and syntactic) approach provided to deal with belief expansion, belief contraction and belief revision. Next, we turned to the semantic approaches to belief revision. We focused on the logics modelling belief revision in a static (Conditional Doxastic Logic) or dynamic way (Dynamic Epistemic Logic, Dynamic Doxastic Logic). We also presented a new approach dealing with evidential dynamics (Evidence Logics). Finally, we investigated temporal doxastic logics (branching-time temporal belief revision) modelling belief revision over time. What we have covered in Chapter 2 does not exhaust all the literature on this topic but provides the formal details of several important approaches that were used in the later chapters.

In Part II, we further developed the setting of Soft Dynamic Epistemic Logic. In Chapter 3, we did justice to the notion of belief contraction. Indeed belief revision is not the only notion that is worth considering. We explored in this chapter three different notions of contraction in the framework of *DEL*, clarifying the mechanism of each of these operations. In Chapter 4, we provided the new setting of justification models, general enough to encompass some existing formalisms. Indeed justification models subsume plausibility models, counting and weighting models as well as evidence models. We defined the usual epistemic and doxastic notions as well as the notions of evidence and justification in those justification models. In Chapter 5, we used this new setting to create a game semantics allowing to determine if an agent really (defeasibly) knows some given proposition or if she only believes this proposition. The game semantics we provided gives a qualitative formalization of Keith Lehrer's philosophical account in which defeasible knowledge is interpreted as undefeated justified acceptance.

In Part III, we connected the setting of Soft Dynamic Epistemic Logic with two other settings. In Chapter 6, we showed that if considered at an appropriate level of generality, the setting of Dynamic Doxastic Logic and the setting of Dynamic Epistemic Logic are in fact equivalent. Moreover, we showed that *DDL* is potentially at least as expressive and powerful as the single-agent version of *DEL*: all work on belief revision done in *DEL* style can be done in *DDL* style. In Chapter 7, we provided an argumentative interpretation of the belief revision logic of Bonanno. We interpreted the notion of belief and information as well as their relation in terms of choice in a dialogical framework.

Finally let us focus on the three important notions which we dealt with in this thesis: Knowledge, Argumentation and Dialogue. From our point of view these three concepts are inseparably linked together. We investigated the notion of Knowledge, trying to solve the eternal debate about what is the correct definition of knowledge. We choose to define (defeasible) knowledge using the notions of truth and justification: an agent knows a proposition if and only if she has a "correct" justification (i.e. a true undefeatable argument) for this proposition. It is sufficient for the agent to have at least one such correct justification to be allowed to say "I know". Checking if an agent really knows a proposition or only believes this proposition entails that this agent argues that she has a correct justification for this proposition. Here comes in our second notion of Argumentation. The agent has to provide arguments to justify her knowledge. Justifying a claim emerges only in argumentative contexts in which there are (at least) two agents. As long as an agent is alone she can claim and argue that she knows something when she is a total ignorant. It is only when an (omniscient) Opponent objectively challenges her arguments to check whether they are sound and convincing that we can check that the agent has a correct justification for her knowledge. As such we end up with our last notion of Dialogue, bringing all three notions together.

Asking new questions

Philosophers have been looking for answers to key questions since antiquity but many of these questions have not been answered, on the contrary what often happens in a philosophical debate is that questions get refined and reformulated and very often new questions arise. This thesis is not different in this respect, while looking for answers we have come to realize that there are many more questions that lay on our path. In the remainder of this section we list new questions and indicate new directions of work, all building further on the results obtained in this thesis.

We start with the last chapter in which we provided a dialogical setting for Bonanno's branching time logic. We stressed earlier on that Bonanno's branching time belief revision logic restricts the main information operator I to Boolean formulas, which means that the agent is facing only new factual information but not any information about her own beliefs and higher-order beliefs. In many scenarios, revision with higher-order beliefs is important so the question about extending Bonanno's setting to higher-order belief revision should be addressed. However, lifting Bonanno's Boolean restriction on the I-operator will have strong consequences for his belief revision setting, in particular the current frame conditions and axioms need then to be revised. One of the necessary revisions refers to Bonanno's belief acceptance axiom (similar to the AGM success postulate), which can no longer be maintained in a dynamic context that holds on to the consistency of (higher-order) beliefs (see Section 6.1.3 for more details). We believe that constructing a revised framework for a branching time belief revision logic is important. Ideally, such a revised framework should keep track of two things: how beliefs evolve over time as well as the conditional beliefs of an agent at a previous moment in time (before the belief changing action took place). Looking for a temporal belief revision setting with those two ingredients, brings us immediately close to the work that has recently been developed in the context of soft DEL and its connections to doxastic temporal logics for multi-agent belief revision, allowing also revision with higher-order information. Here we mention the work of van Benthem and Dégremont on doxastic-temporal models H with a total plausibility pre-order in [27, 13] for which they show that if these doxastic-temporal models H satisfy a number of conditions (i.e. the property of propositional stability, synchronicity, bisimulation invariance, preference propagation and preference revelation) then there exists a total plausibility model $\mathcal{M}$ and a sequence of total plausibility event models such that H is isomorphic to the forest generated by the priority update of $\mathcal{M}$ by this sequence of events, and vice versa.[1] In the light of this investigation, it would be interesting to study any exact correspondence there is between the temporal belief revision axioms given by Bonanno (and their adaptations to a higher-order belief setting) and the bridge principles (such as "propositional stability", "preference propagation" and so on) of van Benthem and Dégremont. In the end one hopes for a more general characterization of the specific classes of doxastic-temporal models resulting from

[1] This result (in the way we have referred to it here) has been worked out for so-called uniform protocols in [13] but we do note that this result can be (and has also been) extended to the setting of state-based protocols, i.e. the case in which "the set of executable sequences of events forming our current informational process, varies from state to state" (see [13] p.15) and to the case where the models are equipped with a partial plausibility relation (see also [27]).

specific belief revision, (or expansion or contraction) protocols. Such a result will go beyond the already expressed formal relation between the I operator and the public announcement operator in *PAL* (see Zvesper [83] and Section 2.8 for more details.) for belief expansion in which case it is easy to match the principles of Bonanno (No Add, No Drop, Qualitative Acceptance) to the principles of Zvesper (No Miracles, Perfect Recall, Uniform Announcement).[2] The work in this direction would bring an interesting contribution to the further alignment of two research directions (temporal belief revision logic and dynamic epistemic logics for belief change). Note that we have recently started working in this direction together with S. Gosh and S. Smets.

In Chapter 7 we offered an argumentative study of belief revision logic, providing the dialogical meaning of information and belief operators. The next step would be to investigate the dialogical meaning of other epistemic and doxastic operators (defeasible knowledge, conditional belief, strong belief) as well as the dynamic operators of Soft *DEL*. A dialogical approach to *DEL* (in particular to *PAL*) has already been provided in [56]. This approach provides the dialogical meaning of the update operator: the dialogical rules for the update operator entail a restriction on the choices of the players (restricting the number of available choices for both players). These results give rise to the following question: what would be the dialogical meaning of the upgrade operators? What would be the impact of the dialogical rules for the upgrade operators on the choices of the players? Next the notions of evidence and justification will be worthwhile to investigate as well in this setting. We think that it would be very interesting to build an argumentative setting allowing to deal with players trying to provide some justifications for their knowledge/beliefs as well as pieces of evidence for or against a particular claim P. Some recent work [56, 57] started to investigate the connections between (Public Announcement) Logic and Law from the point of view of dialogues. These investigation do study the similarity of the notions of proof and evidence both in Law and Logic. The setting of Law reveals how the dialogical interpretation of the public announcement operator is worthwhile in the context of a legal trial. It allows to introduce the notion of proof but also to determine who has to bear the burden of proof. Following these very recent results in [56, 57], it would be interesting to provide a formal framework allowing

[2] We introduce these principles in Chapter 2. The axioms Perfect Recall (PR), No Miracles (MN) on the one hand and No Drop (ND), No Add (NA), Qualitative Acceptance (QA) on the other hand express something very similar. PR states that the agents have a perfect memory and ND states that an agent does not drop her beliefs if the incoming information is not surprising. NM states that the only way an agent can change her mind is when it is triggered by an announcement and NA states that an agent does not add a new belief about which she is not informed if the incoming information is not surprising. Finally QA states that if the information received does not contradict the initial beliefs, an agent believes it and we know that Zvesper only considers the case where the incoming information is not surprising.

to deal with the notions of justification, evidence, upgrade, knowledge and belief in an argumentative framework.

In Chapter 5 we have given a formal foundation of the close relationship between "belief revision" and the definition of "knowledge" in K. Lehrer's defeasibility account of knowledge. What we have shown in this thesis is that the tools of Soft *DEL* are very powerful and can be used to make several philosophical ideas formally precise. In particular we have paid special attention to Lehrer's ultra-justification game. However, Lehrer's philosophy also describes how an agent can justify her (possibly false) beliefs via a so-called personal justification game. While we have not covered it in this thesis, it is possible to extend our formal setting to a game semantics for the case of personal justification. In Lehrer's personal justification games, an agent has to prove that her beliefs fit with her background knowledge and beliefs. So we think that all this agent has to prove in terms of our justification setting, is that her beliefs are consistent and that they do support her claim. In the case of a personal justification game, the agent does not have to provide a sound justification for her claim, but only an argument for it. The work in Chapter 5 opens the door to provide game-semantic accounts for many more concepts reaching beyond the notions of truth and (single-agent) knowledge. Defeasible knowledge is just the tip of the iceberg, there are other epistemic and doxastic attitudes that are worth analysing and making precise using the tools we have provided. In particular our game semantical framework can maybe be extended to combine different epistemic attitudes, defining a game for agents having multiple attitudes (as we actually do in real life). A further challenge would be to look at a multi-player game against a critic, this is interesting when analysing different notions of group knowledge. In this context it would be interesting to look at the notion of common knowledge and analyse it in the philosophical context of Lewis' work on conventions [51]. In joint work with A. Baltag and S. Smets we want to explore these directions of work further.

A further direction of research would connect our game semantics to the study of argumentation theory. Argumentation theory provides a framework representing arguments and the relations between these arguments (an argument can attack or defend another argument) as well as formal methods to define what arguments are justified. This kind of framework is called *argumentation system* and has been introduced in [30]. Here we are only interested in the systems dealing with an abstract notion of argument. These systems are called *abstract argumentation systems*[3]. Abstract argumentation systems have been developed by [30] and more recently by [38]. An abstract argumentation system represents relations between arguments through formal models called attack graphs - or Dung framework. Solving an attack graph means finding the justified argument in that attack

[3]See [9, 65] for a more detailed presentation of argumentation theory.

graph. There exist several methods to determine which arguments are justified in an attack graph corresponding to different notions of "best" argument. Different sets of criteria - corresponding to these different notions - a justified argument must satisfy are provided. Here we are only interested in the notion of grounded set according to which a justified argument is an unattacked argument. Our idea is to provide a game semantics for defeasible knowledge using this argumentation setting. We want to use attack graphs as graphic representation of available arguments as well as the relations between these arguments. Thus they would represent the different choices available for both players during an ultra-justification game.

In chapter 3 we have conducted the study of belief contraction in *DEL*. Indeed we wanted to study how an agent contracts her beliefs after receiving a new information. Does the agent lose a lot of information after contracting her beliefs? Is she willing to prefer to believe the inverse of the information she received ? Or did she keep her initial beliefs as much as possible, contracting only what is necessary to give up the belief contradicting the information she receives? The answer we gave to these questions within *DEL* is that it all depends on the trust the agent has in the source of the information. In this thesis we noted that the belief contraction operation that was easiest to formalize is the conservative contraction operation. The conservative contraction operation corresponds to the case where the agent is careful and contracts as little as possible her beliefs, keeping as much as possible her initial beliefs. What does it mean? We know that the three contraction operations we studied are *AGM* friendly so they respect *AGM* postulates, i.e. the rationality constraints that should be imposed upon belief contraction. We think that our results mean that, among the notions of contraction studied in this thesis, the most rational way to contract the beliefs is the conservative way. Of course, in the literature other notions of contraction have been studied that are less *AGM* friendly. It would be interesting to analyse these notions as well within the framework of *DEL* and compare them with the results we have presented here. In joint work with A. Baltag and S. Smets we will start exploring this direction of work further.

*
* *

Looking beyond the horizon The answers in this thesis (as well as the new questions we want to address in the future) belong to the area of formal epistemology. However, we are well aware that other approaches in formal epistemology use quantitative methods to model belief change. While our work is of a pure qualitative nature, we do think that exploring the connections to quantitative frameworks may lead us to new insights. Going from qualitative to quantitative methods is only one way to look further, another road to explore (as some of our new questions indicate) brings us to study belief change in a multi-agent setting. One idea would be to bring in multiple agents who can reason about each other's beliefs.

However another idea would be to argue that the true nature of (even single-agent) belief change can only be revealed in an inherently multi-agent context, and maybe our game semantics ultimately does hint in this direction. The roles of the two players are essential in this respect: we need a Believer and a Critic exactly as indicated in the previous chapters. If we take this second idea a step further, it ties in with one of van Benthem's views on analysing belief dynamics itself as a form of multi-agent preference merge (we refer here to a presentation of van Benthem, titled "The Social Choice Behind Belief Revision" held at the workshop "Dynamic Logic Montreal" in 2007). Hence, viewing belief revision from a multi-agent angle opens a whole new philosophical perspective which, in a larger framework, can be supported by the approach we presented in this thesis.

And last but not least, we are well aware of the fact that we live in a world in flux, it is not only our beliefs that change, our knowledge that gets updated but also the environment changes. In this thesis we have restricted ourselves to study belief change in a world in which the environment (or the facts themselves) do not change. Other work in the belief revision literature lifts this restriction and similarly, the formal tools of *DEL* have been extended to study fact-changes. Hence looking beyond the horizon we have drawn for ourselves means that we should study belief change, knowledge and its justification in a changing environment.

While playing with knowledge and belief, it is fair to say that the game has only just begun.

Appendix A

Structural rules

We provide some schemas for the reader who is not familiar with structural rules. For a detailed explanation of the rules, see Chapter 7, Section 7.2.3. We only present the rules allowing Proponent to state formulas or choose contextual points. The line above the dash describes the moves required in order to allow the move under the dash. If more than one move is required, the moves are listed one above the other. If the line above the dash is crossed out, it means that there should be no such move to allow the move under the dash. For any contextual points t, u and z and contextual points i, j and l:

(SR-2) $$\frac{\langle \mathbf{O} - i, z : p\rangle}{\langle \mathbf{P} - i, t : p\rangle}$$

(SR-5) $$\frac{\langle \mathbf{O} - i, t : ?\bigcirc_u\rangle}{\langle \mathbf{P} - i, t : ?\bigcirc_u\rangle}$$

$$\frac{\langle \mathbf{O} - i, t : ?\bigcirc_u^{-1}\rangle}{\langle \mathbf{P} - i, t : ?\bigcirc_u^{-1}\rangle}$$

(SR-5.1) $$\frac{\langle \mathbf{O} - i, t : ?\bigcirc_u^{-1}\rangle}{\langle \mathbf{P} - i, t.u : ?\bigcirc_t\rangle}$$

(SR-5.2) $$\frac{\langle \mathbf{O} - i, t : ?\bigcirc_u\rangle}{\langle \mathbf{P} - i, t.u : ?\bigcirc_t^{-1}\rangle}$$

(SR-6) $$\frac{\langle \mathbf{O} - i, t : ?B_j\rangle}{\langle \mathbf{P} - i, t : ?B_j\rangle}$$

$$\frac{\cancel{\langle \mathbf{O} - i, t : ?B_l\rangle}}{\langle \mathbf{P} - i, t : ?B_j\rangle} \quad \text{for any new } j$$

$$\frac{\langle \mathbf{O} - i, t : ?I_j \rangle \quad \text{or} \quad \langle \mathbf{O} - i, t : ?B_j \rangle}{\langle \mathbf{P} - i, t : ?I_j \rangle}$$

$$\frac{\langle \mathbf{O} - i, z : ?I_j \rangle \text{ or } \langle \mathbf{O} - i, z : ?B_j \rangle \text{ or } \langle \mathbf{O} - i, z : ?A_j \rangle}{\langle \mathbf{P} - i, t : ?A_j \rangle}$$

$$\frac{\langle \mathbf{O} - i, t : A\varphi \rangle}{\langle \mathbf{P} - i, t : ?A_i \rangle}$$

Now let four contextual points t, u, v and z be such that u and v have been chosen by **O** to challenge a move as $\langle \mathbf{P} - i, t : \bigcirc\varphi \rangle$ and consider three contextual points i, j and k:

(SR-6.1)
$$\frac{\begin{array}{l}\langle \mathbf{O} - i, t.u : ?B_j \rangle \\ \langle \mathbf{O} - i, t : ?B_k \rangle \\ \langle \mathbf{O} - i, t.u : ?I_k \rangle \quad \text{or} \quad \langle \mathbf{O} - i, t.u : ?I_k^* \rangle\end{array}}{\langle \mathbf{P} - i, t : ?B_j \rangle}$$

(SR-6.2)
$$\frac{\begin{array}{l}\langle \mathbf{O} - i, t : ?B_j \rangle \\ \langle \mathbf{O} - i, t.u : ?I_j \rangle \quad \text{or} \quad \langle \mathbf{O} - i, t.u : ?I_j^* \rangle\end{array}}{\langle \mathbf{P} - i, t.u : ?B_j \rangle}$$

(SR-6.3)
$$\frac{\begin{array}{l}\langle \mathbf{O} - i, t.u : ?B_j \rangle \\ \langle \mathbf{O} - i, t.u : ?I_j \rangle \quad \text{or} \quad \langle \mathbf{O} - i, t.u : ?I_j^* \rangle \\ \langle \mathbf{O} - i, t.v : ?I_j \rangle \quad \text{or} \quad \langle \mathbf{O} - i, t.v : ?I_j^* \rangle\end{array}}{\langle \mathbf{P} - i, t.v : ?B_j \rangle}$$

(SR-6.4)
$$\frac{\begin{array}{l}\langle \mathbf{O} - i, t.u : ?B_j \rangle \\ \langle \mathbf{O} - i, t.v : ?I_j \rangle \quad \text{or} \quad \langle \mathbf{O} - i, t.v : ?I_j^* \rangle\end{array}}{\langle \mathbf{P} - i, t.v : ?B_j \rangle}$$

(SR-7)
$$\frac{\begin{array}{l}\langle \mathbf{X} - i, t : I\varphi \rangle \\ \langle \mathbf{O} - i.j, z : \varphi \rangle\end{array}}{\langle \mathbf{Y} - i, t : !I_j \rangle}$$

Bibliography

[1] C.E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50:510–530, 1985.

[2] G. Aucher. DEL-sequents for progression. *Journal of Applied Non-Classical Logics*, 21(3-4):289–321, 2011.

[3] A. Baltag, J. van Benthem, and L.S. Moss. Epistemic logic and information update. *Handbook of the Philosophy of Information, Elsevier Science Publishers, Amsterdam*, pages 361–456, 2008.

[4] A. Baltag, V. Fiutek, and S. Smets. DDL as an "internalization" of dynamic belief revision. *Krister Segerberg on Logic of Actions. Outstanding Contributions to Logic*, 1, 2014.

[5] A. Baltag, L.S. Moss, and S. Solecki. The logic of common knowledge, public announcements, and private suspicions. In I. Gilboa, editor, *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK 98)*, pages 43–56, 1998.

[6] A. Baltag and S. Smets. Conditional doxastic models: a qualitative approach to dynamic belief revision. *Electronic Notes in Theoretical Computer Science*, 165:5–21, 2006.

[7] A. Baltag and S. Smets. Dynamic belief revision over multi-agent plausibility models. In W. van der Hoek and M. Wooldridge, editors, *Proceedings of LOFT'06*, pages 11–24. University of Liverpool, Liverpool, 2006.

[8] A. Baltag and S. Smets. A qualitative theory of dynamic interactive belief revision. In G. Bonanno, W. van der Hoek, and M. Wooldridge, editors, *Logic and the Foundations of Game and Decision Theory (LOFT 7)*, Texts in Logic and Games, pages 9–58. Amsterdam University Press, 2008.

[9] P. Baroni and M. Giacomin. Semantics of abstract argument systems. In *Argumentation in Artificial Intelligence*, pages 25–44. Springer, 2009.

[10] J. van Benthem. Dynamic logic for belief revision. *Journal of Applied Non-classical Logics*, 17(2):129–155, 2007.

[11] J. van Benthem. *Logical Dynamics of Information and Interaction.* Cambridge University Press, 2011.

[12] J. van Benthem. Two logical faces of belief revision. *Krister Segerberg on Logic of Actions. Outstanding Contributions to Logic*, 1, 2014.

[13] J. van Benthem and C. Dégremont. Bridges between dynamic doxastic and doxastic temporal logics. In *Logic and the Foundations of Game and Decision Theory–LOFT 8*, pages 151–173. Springer, 2010.

[14] J. van Benthem, D. Fernández-Duque, and E. Pacuit. Evidence and plausibility in neighborhood structures. *Annals of Pure and Applied Logic*, 2013.

[15] J. van Benthem, J. Gerbrandy, T. Hoshi, and E. Pacuit. Merging frameworks for interaction. *Journal of Philosophical Logic*, 38(5):491–526, 2009.

[16] J. van Benthem and E. Pacuit. Dynamic logics of evidence-based beliefs. *Studia Logica*, 99(1):61 – 92, 2011.

[17] G. Bonanno. A simple modal logic for belief revision. *Synthese*, 147:193–228, 2005.

[18] G. Bonanno. Axiomatic characterization of the agm theory of belief revision in a temporal logic. *Artificial Intelligence*, 171:144–160, 2007.

[19] G. Bonanno. A sound and complete temporal logic for belief revision. *Dialogues, Logics and other Strange Things. Essays in Honour of Shahid Rahman*, 147:193–228, 2007.

[20] G. Bonanno. Temporal belief revision rationalized by plausibility orderings. *Working Paper*, 2007.

[21] G. Bonanno. Belief revision in a temporal framework. *New Perspectives on Games and Interaction*, 4:45–80, 2009.

[22] G. Bonanno. Belief change in branching time: AGM-consistency and iterated revision. *Working Paper*, 2010.

[23] C. Boutilier. Iterated revision and minimal change of conditional beliefs. *Journal of Philosophical Logic*, 25(3):263–305, 1996.

[24] A. Brandenburger and H.J. Keisler. An impossibility theorem on beliefs in games. *Studia Logica*, 84(2):211–240, 2006.

[25] N. Clerbout. *Étude de quelques sémantiques dialogiques. Concepts fondamentaux et éléments de métathéorie.* PhD thesis, Université de Lille, 2014.

[26] A. Darwiche and J. Pearl. On the logic of iterated belief revision. *Artificial intelligence*, 89(1):1–29, 1997.

[27] C. Dégremont. *PhD Thesis: The Temporal Mind. Observations on the logic of belief change in interactive systems.* Universiteit van Amsterdam, 2010.

[28] H. van Ditmarsch, W. van der Hoek, and P. Ilievr. Everything is knowable–how to get to know whether a proposition is true. *Theoria*, 78(2):93–114, 2012.

[29] H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic epistemic logic.* Synthese Library. Springer, 2007.

[30] P.H. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and < i> n</i>-person games. *Artificial intelligence*, 77(2):321–357, 1995.

[31] E. Fermé and R. Rodriguez. A brief note about Rott contraction. *Logic Journal of IGPL*, 6(6):835–842, 1998.

[32] M.J. Fischer and R.E. Ladner. Propositional dynamic logic of regular programs. *Journal of computer and system sciences*, 18(2):194–211, 1979.

[33] V. Fiutek. A dialogical approach of iterated belief revision. In C. Barès, S. Magnier, and F. Salguero, editors, *Logic of Knowledge. Theory and Applications*, pages 141–157. College Publications, London, 2012.

[34] V. Fiutek, H. Rückert, and S. Rahman. A dialogical semantics for Bonanno's system of belief revision. In *Construction*, pages 315–334. College Publication, London, 2010.

[35] J. Gerbrandy. Dynamic epistemic logic. In L.S. Moss et al., editor, *Logic, Language and Information, vol. 2.* CSLI Publications, Stanford University, 1999.

[36] J. Gerbrandy and W. Groeneveld. Reasoning about information change. *Journal of Logic, Language and Information*, 6:147–169, 1997.

[37] E. Gettier. Is justified true belief knowledge? *Analysis*, 23:121–123, 1963.

[38] D. Grossi. Fixpoints and iterated updates in abstract argumentation. In *KR*, 2012.

[39] A. Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17:157–170, 1988.

[40] W.L. Harper. Rational conceptual change. In *Proceedings of the Meeting of the Philosophy of Scienee Association*, pages 462–494, 1975.

[41] J. Hintikka. *Knowledge and Belief.* Cornell University Press, Ithaca, NY, 1962.

[42] J. Hintikka and G. Sandu. Game-theoretical semantics. *Handbook of Logic and Language*, pages 361–411, 1996.

[43] I.L. Humberstone. The modal logic of "all and only". *Notre Dame Journal of Formal Logic*, 28(2):177–188, 1987.

[44] L. Keiff. Dialogical logic. `http://plato.stanford.edu/entries/logic-dialogical`, 2009.

[45] L. Keiff and S. Rahman. On how to be a dialogician. In *Logic, thought and action*, pages 359–408. Springer, 2005.

[46] K. Lehrer. *Theory of Knowledge.* Routledge, London, 1990.

[47] K. Lehrer. *Theory of Knowledge.* Westview Press, United States, 2000.

[48] H.J. Levesque. All I know: a study in autoepistemic logic. *Artificial Intelligence*, 42(2):263–309, 1990.

[49] I. Levi. Subjunctives, dispositions and chances. *Synthese*, 34(4):423–455, 1977.

[50] I. Levi. *Mild contraction: Evaluating loss of information due to loss of belief.* Clarendon Oxford, 2004.

[51] D. Lewis. *Convention — A Philosophical Study.* Harvard University Press, Cambridge, Massachusetts, 1969.

[52] D.K. Lewis. *Counterfactuals.* Blackwell Publishing, Oxford, 1973.

[53] S. Lindström and W. Rabinowicz. Belief change for introspective agents. *Spinning Ideas, Electronic Essays Dedicated to Peter Gärdenfors on His Fiftieth Birthday*, 1999.

[54] S. Lindström and W. Rabinowicz. DDL unlimited: dynamic doxastic logic for introspective agents. *Erkenntnis*, 50:353–385, 1999.

[55] K. Lorenz and P. Lorenzen. *Dialogische logik.* Wissenschaftliche Buchgesellschaft Darmstadt, 1978.

[56] S. Magnier. *Approche dialogique de la dynamique épistémique et de la condition juridique*. College publications, 2013.

[57] S. Magnier. *Epistemic dynamics, Dialogues and Legal Reasoning*. Springer LAR, To appear.

[58] M. Marion. Game semantics and the manifestation thesis. In *The Realism-Antirealism Debate in the Age of Alternative Logics*, pages 141–168. Springer, 2012.

[59] G.E. Moore. Moore's paradox. *GE Moore: Selected writings*, pages 207–212, 1993.

[60] L.S. Moss and R. Parikh. Topological reasoning and the logic of knowledge. In Y. Moses, editor, *Proceedings of TARK IV*, pages 95–105, San Francisco, 1992. Morgan Kaufmann Publishers.

[61] L.S. Moss, R. Parikh, and C. Steinsvold. Topology and epistemic logic. In *Handbook of Spatial Logics*, pages 299–341. Springer, 2007.

[62] A.C. Nayak, M. Pagnucco, and P. Peppas. Dynamic belief revision operators. *Artificial Intelligence*, 146(2):193–228, 2003.

[63] M. Pagnucco and H. Rott. Severe withdrawal (and recovery). *Journal of Philosophical Logic*, 28(5):501–547, 1999.

[64] J.A. Plaza. Logics of public communications. In M.L. Emrich, M.S. Pfeifer, M. Hadzikadic, and Z.W. Ras, editors, *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*, pages 201–216, 1989.

[65] H. Prakken and G. Vreeswijk. Logics for defeasible argumentation. *Handbook of philosophical logic*, 4(5):219–318, 2002.

[66] S. Rahman. Non-normal dialogics for a wonderful world and more. In *The age of alternative logics*, pages 311–334. Springer, 2006.

[67] S. Rahman. Abduction, belief-revision and non-normality. In *Abduction and the Process of Scientific Discovery*, pages 13–25. Lisboa: CFCUL/Publidisa, 2007.

[68] S. Rahman and H. Rückert. Dialogische modallogik (für T, B, S4, und S5). *Logique et analyse*, 167(168):243–282, 1999.

[69] S. Rahman and H. Rückert. New perspectives in dialogical logic. *Synthese*, 127, 2001.

[70] S. Rahman and T. Tulenheimo. From games to dialogues and back. In *Games: Unifying Logic, Language, and Philosophy*, pages 153–208. Springer, 2009.

[71] H. Rott. Stability, strength and sensitivity: Converting belief into knowledge. *Erkenntnis*, 61:469–493, 2004.

[72] H. Rott. Shifting priorities: Simple representations for twenty-seven iterated theory change operators. In *Towards mathematical philosophy*, pages 269–296. Springer, 2009.

[73] K. Segerberg. The basic dynamic doxastic logic of AGM. In *The Goldblatt Variations*, volume 1, pages 76–107. Uppsala Prints and Preprints in Philosophy, 1991.

[74] K. Segerberg. Belief revision from the point of view of doxastic logic. *Bulletin of the IGPL*, 3:535–553, 1995.

[75] K. Segerberg. A general framework for the logic of theory change. *Bulletin of the Section of Logic*, 25:2–8, 1996.

[76] K. Segerberg. Proposal for a theory of belief revision along the lines of lindström and rabinowicz. *Fundamenta Informaticae*, 32(2):183–191, 1997.

[77] K. Segerberg. Irrevocable belief revision in dynamic doxastic logic. *Notre Dame Journal of Formal Logic*, 39(3), 1998.

[78] K. Segerberg. A completeness proof in full DDL. *Philosophical Crumbs: Essays Dedicated to Ann-Mari Henschen-Dahlqvist on the Occasion of her Seventy-fifth Birthday, Uppsala Philosophical Studies*, 49:195–207, 1999.

[79] K. Segerberg. Two traditions in the logic of belief: bringing them together. In H. J. Ohlbach and U. Reyle, editors, *Logic, Language, and Reasoning*, pages 135–147, Dordrecht, 1999. Kluwer Academic Publishers.

[80] K. Segerberg and H. Leitgeb. Dynamic doxastic logic - why, how and where to? *Synthese*, 155(167-190), 2007.

[81] G.H. von Wright. *An Essay in Modal Logic.* North-Holland Amsterdam, Amsterdam, 1951.

[82] D.N. Walton and C.W. Krabbe. *Commitment in dialogue: Basic concepts of interpersonal reasoning.* SUNY press, 1995.

[83] J.A. Zvesper. How to keep changing your mind, dynamically. In S. Ju J. van Benthem and F. Veltman, editors, *A Meeting of the Minds, Proceedings of the workshop on Logic, Rationality and Interaction, Texts in Computer Science*, volume 8, pages 291–306, 2007.

Index

Samenvatting

Dit proefschrift draagt bij aan de ontwikkeling van de Soft Dynamisch Epistemische Logica. De Soft Dynamisch Epistemische Logica werd geïntroduceerd om met een aantal informationele fenomenen, zoals geloofsherziening, om te gaan. In dit proefschrift breiden we de Soft Dynamisch Epistemische Logica uit naar "geloof contractie". Op deze wijze creëren we een kader voor de studie van geloofsverandering. Dit onderzoek naar geloofsverandering draagt ook bij aan de studie van het kennis-concept. Eén van de belangrijkste uitdagingen in de formele epistemologie vandaag de dag is het formeel vastleggen van wat een juiste definitie van kennis is. Om dit aan te pakken, moeten we de concepten van "evidence" en "justified true belief" formeel definiëren. In dit proefschrift breiden we de Soft Dynamisch Epistemische Logica zodanig uit dat het toegepast kan worden op de concepten van "evidence" en "justification". We ontwikkelen in deze context een speltheoretische semantiek voor de notie van "defeasible knowledge". Op deze wijze geven we een nieuwe formalisatie aan K. Lehrer's notie van kennis in termen van "undefeated justified acceptance". Dit kader biedt een nieuw perspectief voor de analyse van epistemologische problemen zoals het Gettier-probleem. Dit proefschrift geeft ook het verband weer tussen de Soft Dynamisch Epistemische Logica en twee andere aanpakken die in de literatuur werden bestudeerd. Eén van deze andere aanpakken is de Dynamisch Doxastische Logica van K. Segerberg. Een belangrijk deel van het werk in dit proefschrift gaat over de vergelijking van deze twee benaderingen en daarbij brengen we de verschillen en overeenkomsten in kaart. Ten slotte verbinden we ons werk aan de argumentatieve studie van geloofsherziening waarbij we de dynamiek van geloof onderzoeken in een dynamisch argumentatief kader.

Hoofdstuk 2 presenteert een aantal verschillende benaderingen van geloofsverandering die in de literatuur werden bestudeerd.

In Hoofdstuk 3 bestuderen we drie noties van geloofscontractie die we we uit de literatuur hebben gekozen: extreme contractie, gematigde contractie en conservatieve contractie. We definiëren de overeenkomstige operaties van geloofscontractie als operaties in plausibiliteitsmodellen en we geven een axiomatisatie voor elk van deze operaties in de stijl van *DEL*.

In Hoofdstuk 4 bestuderen we "justification"-modellen waarin we de informatie en bewijsstukken van een agent kunnen modelleren. We geven een formele definitie van een geldig (waar) argument en van wat een rechtvaardiging (of "justification") is.

In Hoofdstuk 5 analyseren we de informele kennis-theorie van K. Lehrer in het kader waarin we een oplossing zoeken voor het Gettier probleem. We stellen in dit hoofdstuk ook een speltheoretische semantiek voor zodanig dat we Lehrer's notie van "defeasible knowledge" formeel kunnen onderbouwen. Het ultra-rechtvaardigingsspel laat ons toe om op formele wijze te bepalen als een agent deze vorm van "defeasible knowledge" heeft van een propositie (of als hij deze propositie gelooft maar niet kent): een agent ("the Claimant") heeft "defeasible knowledge" van een propositie P als en slechts dan als zij een strategie heeft om het ultra-rechtvaardigingsspel voor propositie P te winnen.

Hoofdstuk 6 onderzoekt de relatie tussen de Dynamisch Doxastische Logica en de Dynamisch Epistemische Logica waarin we "full *DDL*" bestuderen vanuit het perspectief van "Soft *DEL*". We bekijken verschillende versies van *DDL*, waarbij de aandacht gaat naar verschillende operaties voor geloofsherzieningen evenals verschillende operaties van geloofsuitbreiding en -contractie. We tonen aan dat *DDL* minstens even krachtig is als *DEL*.

In Hoofdstuk 7 geven we een uiteenzetting over G. Bonanno's logica voor "branching-time belief revision" en bestuderen we deze logica in een argumentatief kader. We rechtvaardigen onze keuze om Bonanno's logica voor geloofsherziening te onderzoeken via de aanpak van de dialoog-logica. We geven in dit hoofdstuk de taal en de regels van het dialogisch systeem voor geloofsherziening. We richten onze aandacht op de dialogische interpretatie van de concepten van geloof en informatie.

Abstract

This thesis contributes to the development of Soft Dynamic Epistemic Logic. Soft Dynamic Epistemic Logic has been introduced to deal with a number of informational phenomena, including belief revision. The work in this thesis extends the scope of Soft Dynamic Epistemic Logic to belief contraction, providing as such a framework which can now deal with belief change. This study of belief change contributes also to the study of the notion of knowledge. Nowadays, one of the main challenges in formal epistemology is to formally capture what is a correct definition of knowledge. To tackle this issue we need to be able to formally define the notions of evidence and justified true belief. In this thesis, we extend Soft Dynamic Epistemic Logic such that it can indeed deal with the notions of evidence and justification. In this context we provide a game semantics for "defeasible knowledge", offering a new formalization of K. Lehrer's concept of knowledge in terms of "undefeated justified acceptance". This setting provides a new perspective for analysing epistemological problems such that the Gettier problem. This thesis also connects Soft Dynamic Epistemic Logic to two different approaches that have been studied in the literature. One of these other approaches is Dynamic Doxastic Logic, as introduced by K. Segerberg. An important part of the work we have done, compares Dynamic Doxastic Logic to Soft Dynamic Epistemic Logic. This comparison makes it possible to investigate what are the differences and the similarities between these two approaches. Finally we connect our work to the argumentative study of belief revision, offering an investigation of belief dynamics in a dynamic argumentative setting.

Chapter 2 presents a number of different settings of belief change that have been studied in the literature.

Chapter 3 introduces three notions of belief contraction that we choose from the literature: severe withdrawal, moderate contraction and conservative contraction. We define the corresponding belief contracting operations as operations on total plausibility models and axiomatize each of them in *DEL* style.

In chapter 4 we introduce the new framework of justification models as a general setting to model the information and evidence an agent has. We formally define what is a sound (true) argument and what is a justification.

Chapter 5 introduces the informal theory of knowledge of K. Lehrer as a solution to the Gettier problem and proposes a game semantics that formalises the notion of defeasible knowledge of K. Lehrer. Our ultra-justification game formally determines if an agent defeasibly knows a proposition (or merely believes but does not know this proposition): an agent (the Claimant) defeasibly knows a proposition P iff she has a winning strategy in the ultra-justification game corresponding to the claim P.

Chapter 6 compares Dynamic Doxastic Logic and Dynamic Epistemic Logic, studying full *DDL* from the perspective of Soft *DEL*. We provide several versions of *DDL* internalizing different belief revision operations, as well as several operations of expansion and contraction, showing that the *DDL* approach is at least as powerful as the *DEL* approach.

Chapter 7 introduces the branching-time belief revision logic of G. Bonanno and provides an argumentative study of this belief revision logic. We use the dialogical approach to logic and provide the language as well as the rules of our dialogical system of belief revision. We focus on the dialogical interpretation of the notions of belief and information.

Résumé

Cette thèse contribue au développement de la Logique Épistémique Dynamique dite flexible (*Soft Dynamic Epistemic Logic*). La Logique Épistémique Dynamique flexible a été introduite afin de capturer un certain nombre de phénomènes informationnels, incluant la révision de croyances. Cette thèse étend la portée de la Logique Épistémique Dynamique flexible à la contraction de croyances, fournissant ainsi une structure capable de traiter plus généralement du changement de croyances. Cette étude du changement de croyances contribue également à l'étude de la notion de savoir. L'un des principaux défis de l'épistémologie formelle contemporaine est de capturer formellement ce qu'est une définition correcte du savoir. Pour répondre à ce problème, nous avons besoin de définir formellement les notions d'évidence et de croyance vraie justifiée. Dans cette thèse, nous étendons la Logique Épistémique Dynamique flexible de telle sorte qu'elle puisse désormais capturer les notions d'évidence et de justification. Dans ce contexte, nous proposons une sémantique des jeux pour le savoir dit défaisable (*defeasible knowledge*), offrant ainsi une nouvelle formalisation du concept de savoir de K. Lehrer en termes d'acceptation justifiée indéfaisable ("undefeated justified acceptance"). Ce cadre fournit une nouvelle perspective pour l'analyse des problèmes épistémologiques tels que le problème de Gettier. Cette thèse connecte également la Logique Épistémique Dynamique flexible à deux approches différentes qui ont été étudiées dans la littérature. Une de ces approches est la Logique Doxastique Dynamique (*Dynamic Doxastic Logic*) introduite par K. Segerberg. Une part importante du travail que nous avons produit, compare la Logique Doxastique Dynamique à la Logique Épistémique Dynamique flexible et permet de mettre en évidence les différences et les similitudes de ces deux approches. Finalement, nous proposons une approche argumentative de la dynamique des croyances.

Le chapitre 2 présente un certain nombre de différentes logiques de changement de croyances qui ont été étudiées dans la littérature.

Le chapitre 3 introduit trois notions de contraction de croyances que nous avons choisies dans la littérature: la réduction sévère (*severe withdrawal*), la

contraction modérée (*moderate contraction*) et la contraction conservatrice (*conservative contraction*). Nous définissons les opérations correspondantes de contraction de croyances comme des opérations sur les modèles de plausibilité totaux (*total plausibility models*) et axiomatisons chacune d'entre elles dans le style de la Logique Épistémique Dynamique.

Dans le chapitre 4 nous introduisons un nouveau cadre, celui des modèles de justification en tant que cadre général permettant de modéliser l'information et les évidences que l'agent possède. Nous définissons formellement ce qu'est un argument vrai et ce qu'est une justification.

Le chapitre 5 introduit la théorie informelle du savoir de K. Lehrer en tant que solution au problème de Gettier et propose une sémantique des jeux qui formalise la notion de savoir défaisable de K. Lehrer. Notre jeu de l'ultra-justification (*ultra-justification game*) détermine formellement si un agent connait – au sens défaisable du terme – une proposition (ou simplement croit mais ne connait pas cette proposition): un agent (que nous appelons le Claimant) connait – au sens défaisable du terme – une proposition P si et seulement si il a une stratégie de victoire dans le jeu de l'ultra-justification correspondant à la revendication P.

Le chapitre 6 compare la Logique Doxastique Dynamique et la Logique Épistémique Dynamique flexible, étudiant la Logique Doxastique Dynamique dite complète (*full DDL*) à partir de la perspective de la Logique Épistémique Dynamique flexible. Nous fournissons plusieurs versions de la Logique Doxastique Dynamique internalisant différentes opérations de révision de croyances, ainsi que plusieurs versions d'expansion et de contraction de croyances, montrant que l'approche de la Logique Doxastique Dynamique est au moins aussi puissante que l'approche de la Logique Épistémique Dynamique flexible.

Le chapitre 7 introduit la logique de révision de croyances de temps branché de G. Bonanno et fournit une étude argumentative de cette logique de révision de croyances. Nous utilisons l'approche dialogique de la logique et fournissons le langage ainsi que les règles de notre système dialogique de révision de croyances. Nous nous concentrons sur l'interprétation dialogique des notions de croyances et d'information.

Titles in the ILLC Dissertation Series:

ILLC DS-2009-01: **Jakub Szymanik**
Quantifiers in TIME and SPACE. Computational Complexity of Generalized Quantifiers in Natural Language

ILLC DS-2009-02: **Hartmut Fitz**
Neural Syntax

ILLC DS-2009-03: **Brian Thomas Semmes**
A Game for the Borel Functions

ILLC DS-2009-04: **Sara L. Uckelman**
Modalities in Medieval Logic

ILLC DS-2009-05: **Andreas Witzel**
Knowledge and Games: Theory and Implementation

ILLC DS-2009-06: **Chantal Bax**
Subjectivity after Wittgenstein. Wittgenstein's embodied and embedded subject and the debate about the death of man.

ILLC DS-2009-07: **Kata Balogh**
Theme with Variations. A Context-based Analysis of Focus

ILLC DS-2009-08: **Tomohiro Hoshi**
Epistemic Dynamics and Protocol Information

ILLC DS-2009-09: **Olivia Ladinig**
Temporal expectations and their violations

ILLC DS-2009-10: **Tikitu de Jager**
"Now that you mention it, I wonder...": Awareness, Attention, Assumption

ILLC DS-2009-11: **Michael Franke**
Signal to Act: Game Theory in Pragmatics

ILLC DS-2009-12: **Joel Uckelman**
More Than the Sum of Its Parts: Compact Preference Representation Over Combinatorial Domains

ILLC DS-2009-13: **Stefan Bold**
Cardinals as Ultrapowers. A Canonical Measure Analysis under the Axiom of Determinacy.

ILLC DS-2010-01: **Reut Tsarfaty**
Relational-Realizational Parsing

ILLC DS-2010-02: **Jonathan Zvesper**
Playing with Information

ILLC DS-2010-03: **Cédric Dégremont**
The Temporal Mind. Observations on the logic of belief change in interactive systems

ILLC DS-2010-04: **Daisuke Ikegami**
Games in Set Theory and Logic

ILLC DS-2010-05: **Jarmo Kontinen**
Coherence and Complexity in Fragments of Dependence Logic

ILLC DS-2010-06: **Yanjing Wang**
Epistemic Modelling and Protocol Dynamics

ILLC DS-2010-07: **Marc Staudacher**
Use theories of meaning between conventions and social norms

ILLC DS-2010-08: **Amélie Gheerbrant**
Fixed-Point Logics on Trees

ILLC DS-2010-09: **Gaëlle Fontaine**
Modal Fixpoint Logic: Some Model Theoretic Questions

ILLC DS-2010-10: **Jacob Vosmaer**
Logic, Algebra and Topology. Investigations into canonical extensions, duality theory and point-free topology.

ILLC DS-2010-11: **Nina Gierasimczuk**
Knowing One's Limits. Logical Analysis of Inductive Inference

ILLC DS-2010-12: **Martin Mose Bentzen**
Stit, Iit, and Deontic Logic for Action Types

ILLC DS-2011-01: **Wouter M. Koolen**
Combining Strategies Efficiently: High-Quality Decisions from Conflicting Advice

ILLC DS-2011-02: **Fernando Raymundo Velazquez-Quesada**
Small steps in dynamics of information

ILLC DS-2011-03: **Marijn Koolen**
The Meaning of Structure: the Value of Link Evidence for Information Retrieval

ILLC DS-2011-04: **Junte Zhang**
System Evaluation of Archival Description and Access

ILLC DS-2011-05: **Lauri Keskinen**
Characterizing All Models in Infinite Cardinalities

ILLC DS-2011-06: **Rianne Kaptein**
Effective Focused Retrieval by Exploiting Query Context and Document Structure

ILLC DS-2011-07: **Jop Briët**
Grothendieck Inequalities, Nonlocal Games and Optimization

ILLC DS-2011-08: **Stefan Minica**
Dynamic Logic of Questions

ILLC DS-2011-09: **Raul Andres Leal**
Modalities Through the Looking Glass: A study on coalgebraic modal logic and their applications

ILLC DS-2011-10: **Lena Kurzen**
Complexity in Interaction

ILLC DS-2011-11: **Gideon Borensztajn**
The neural basis of structure in language

ILLC DS-2012-01: **Federico Sangati**
Decomposing and Regenerating Syntactic Trees

ILLC DS-2012-02: **Markos Mylonakis**
Learning the Latent Structure of Translation

ILLC DS-2012-03: **Edgar José Andrade Lotero**
Models of Language: Towards a practice-based account of information in natural language

ILLC DS-2012-04: **Yurii Khomskii**
Regularity Properties and Definability in the Real Number Continuum: idealized forcing, polarized partitions, Hausdorff gaps and mad families in the projective hierarchy.

ILLC DS-2012-05: **David García Soriano**
Query-Efficient Computation in Property Testing and Learning Theory

ILLC DS-2012-06: **Dimitris Gakis**
Contextual Metaphilosophy - The Case of Wittgenstein

ILLC DS-2012-07: **Pietro Galliani**
The Dynamics of Imperfect Information

ILLC DS-2012-08: **Umberto Grandi**
Binary Aggregation with Integrity Constraints

ILLC DS-2012-09: **Wesley Halcrow Holliday**
Knowing What Follows: Epistemic Closure and Epistemic Logic

ILLC DS-2012-10: **Jeremy Meyers**
Locations, Bodies, and Sets: A model theoretic investigation into nominalistic mereologies

ILLC DS-2012-11: **Floor Sietsma**
Logics of Communication and Knowledge

ILLC DS-2012-12: **Joris Dormans**
Engineering emergence: applied theory for game design

ILLC DS-2013-01: **Simon Pauw**
Size Matters: Grounding Quantifiers in Spatial Perception