

University of Groningen

Evidence-Based Beliefs in Many-Valued Modal Logics

David Santos, Yuri

DOI:
[10.33612/diss.155882457](https://doi.org/10.33612/diss.155882457)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
David Santos, Y. (2021). *Evidence-Based Beliefs in Many-Valued Modal Logics*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.155882457>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Evidence-Based Beliefs
in Many-Valued Modal Logics

Yuri David Santos

Copyright © 2020 by Yuri David Santos

The copyright holder grants any entity the right to use this work **for any purpose**, without any conditions, unless such conditions are required by law.

This research was funded by Ammodo KNAW project “Rational Dynamics and Reasoning”.



university of
 groningen

Evidence-Based Beliefs in Many-Valued Modal Logics

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijmenga
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Thursday 14 January 2021 at 14.30 hours

by

Yuri David Santos

born on 12 November 1988
in Santos, Brazil

Supervisors

Prof. B.P. Kooi

Prof. L.C. Verbrugge

Assessment Committee

Prof. D. Grossi

Prof. S. J. L. Smets

Prof. O. Roy

Acknowledgements

First, I would like to thank my advisors Barteld Kooi and Rineke Verbrugge for the great guidance they have offered me during these years. I consider myself extremely lucky in terms of advisors. They both are very kind and have enormous knowledge in their fields, and I am always impressed about how fast they can see things and make complicated inferences in a matter of milliseconds. Rineke is one of the most responsible people I have ever met in my life, I cannot imagine her promising something and not fulfilling it, and this was a crucial source of confidence and support for me during this project. She has incredibly vast and varied knowledge and experience, and that contributed a lot to improve the quality of this thesis. If you find some grammar errors or clumsy phrasings in this acknowledgements section, it is because Rineke has not read it before I printed it. Rineke always had positive words to tell, even when my drafts needed a lot of improvement. I also thank her for a lot of tips that do not concern the project, such as housing, job seeking and many other things. Now this project would not have existed in the first place if it was not for Barteld, who won the Ammodo prize for his groundbreaking work in Dynamic Epistemic Logic, and single-handedly funded this and other PhDs. Barteld is fun and relaxed, and that is great because it allowed me to not feel pressured or stressed during the project. I am grateful for the freedom with which he guided me and for the confidence he had in me, always very positive and motivating me to pursue my ideas. He helped me with things from lending books to fixing my bike's flat tire, and of course by reading many of my texts and providing me with ideas from his extensive logical expertise. Barteld has a very sharp mind, and it is hard for me to comprehend how someone his age can have so many accomplishments. I feel truly blessed for having been advised by two of the best logicians of our time.

Another person that I would also like to thank is Allard Tamminga, who participated in this project during the first years, and gave me some

very important contributions, such as the idea of using four-valued logic on the problem of logical omniscience and limited reasoners. Allard gave me many cultural suggestions and helped me to get to know the best things to do in my first years in Groningen.

I have to thank the university staff, especially Marga Hids – who unfortunately is leaving the Faculty of Philosophy – but also Elina Sietsema, Siepie Blom, Marion Vorrink, Hauke de Vries and Sarah van Wouwe. I am also grateful to the RUG, which is an excellent research university, for the opportunity of doing this PhD.

I would also like to thank professors Davide Grossi, Sonja Smets and Olivier Roy, who kindly agreed to participate in the reading committee of this thesis. Other professors from the RUG that I need to thank are: Gerard Renardel (who gave an excellent course on modal logic), Jan-Willem Romeijn and Bart Verheij.

I thank my friend Merel for translating my samenvatting to Dutch, and all my friends who have been of vital importance during these years in the Netherlands, but especially: Ebrar, Elli, Faik, Karmen, Oscar, Robson, Stipe, Thiago and Mari. Each of you have been present in the best and the worst moments of my life in many different ways during these years (and I hope will still be in the years to come), and for that I am deeply grateful to all of you. I would also like to send a hello to my old-time friends, especially Caio, Ignasi, Igor, Kelvin and Rafael; I am glad we have kept in touch despite the distance. Thanks also to my project and/or office mates Ana, Burcu, Hang, Xingchi and Yanjun; good friends Abby, Corne, César, Manvi and Mi Xue; my lunch-time and AI friends Hermine, Katja, Stephan, Hamed, Hamid, Vishal, Aniket, Atefeh and Yuzhe; and philosophy colleagues Diego, Eric, Fré, Job, Kritika, Lieuwe, Petar, Sanne and Sjoerd. Thank you all for the good company.

I also want to thank my former master's advisor, Renata Wassermann, who found and suggested me to apply for this PhD position, and my friend Guilherme Furtado who has insisted that I applied, even though I was reluctant at the moment to move to such a distant country. Without these two, I would not have started this PhD.

Last, I would like to thank my family, with special thanks to Victor, João and Dineia.

A Seila, Luiz, Benedita, Rosa, Jaconias e Antônio.

Contents

1	Introduction	1
1.1	Historical Remarks	2
1.2	Logical Omniscience	4
1.3	Many-valued Logics and Evidence	5
1.4	Overview	5
1.5	Prerequisite Knowledge	7
1.6	Publications	7
2	A Multi-Agent Four-Valued Dynamic Epistemic Logic	9
2.1	Introduction	9
2.2	Four-Valued Epistemic Logic	12
2.2.1	Intended Readings of Formulas	14
2.2.2	Validity and Entailment	15
2.2.3	Basic Properties of FVEL	16
2.3	Tableaux	22
2.4	Correspondence Results	28
2.5	Public Announcements	34
2.5.1	Reduction Validities	36
2.5.2	Tableaux	45
2.6	Related Work	46
2.6.1	Logics of Evidence	46
2.6.2	Other Many-Valued Modal Logics	48
2.7	Conclusions and Future Work	50
3	Consolidations: Turning Evidence into Belief	53
3.1	Introduction	53
3.2	Preliminaries	55
3.2.1	Aspects of Evidence	55
3.2.2	Principles of Rationality	57

3.3	A Consolidation Operation	60
3.3.1	Definitions	60
3.3.2	Other Cluster Consolidations	63
3.3.3	Properties	66
3.3.4	A Unified Language for Evidence and Beliefs	69
3.4	Equivalence Between Evidence Models	70
3.4.1	From B&P to FVEL models	72
3.4.2	From FVEL to B&P models	72
3.4.3	Evaluating the conversions	74
3.5	Comparing Consolidations	79
3.6	Conclusion	81
4	Social Consolidations: Evidence and Peerhood	83
4.1	Introduction	83
4.2	Syntax and Semantics	84
4.2.1	Syntax	84
4.2.2	Semantics	84
4.3	Rationality Conditions for Consolidations	86
4.3.1	Epistemic Autonomy versus Epistemic Authority . .	86
4.3.2	Rationality Postulates	87
4.4	Social Consolidations	93
4.4.1	Preliminaries	93
4.4.2	Consolidation Policies	96
4.5	Dynamics	103
4.6	Related Work	110
4.7	Conclusions and Future Work	111
5	Iterative Social Consolidations: Private Evidence	113
5.1	Introduction	113
5.2	Logical Language	114
5.2.1	Syntax	114
5.2.2	Semantics	115
5.3	Iterative Social Consolidations: Preliminaries	116
5.4	Consolidation Policies	118
5.4.1	Policy I: Monotonic Belief Diffusion	118
5.4.2	Policy II: Unstable Consolidations	120
5.4.3	Policy III: Ignoring Unstable Peers	126
5.4.4	Other Policies	130
5.5	Conclusions and Future Work	130

6	Conclusions	133
6.1	The Logic FVEL	133
6.2	Consolidations	134
6.3	The Magic Word: Rationality	138
6.4	The Mainstream View	139
6.5	Closing Thoughts	140
	Bibliography	141
	Summary	155
	Samenvatting	157
	Resumo	161

Chapter 1

Introduction

With more than half of the world’s population having access to the internet¹ and the advent of social media platforms, the way information spreads and is consumed has shifted drastically in the last decades. As many as 42% of news consumers in countries like Chile, Brazil and Malaysia prefer to get informed via social media, whereas in countries such as the US, Canada and Australia this figure is about 25%.² While this new way of sharing information might have some advantages, it also facilitates the spread of misinformation. When asked whether they are worried about what is “real or fake” on the internet, a majority of people in Brazil (85%), the UK (70%) and the US (67%) answered positively, whereas in other countries the figure is lower but still significant: for example, 38% for Germany and 31% for the Netherlands.³

Amidst such a deluge of information of variable quality, perhaps one of the most important skills to have in the twenty-first century is to bear proper “epistemic machinery”: to know how to fetch, filter and aggregate information, separate reliable from unreliable sources, combine the pieces of evidence appropriately and draw sensible conclusions from it.⁴ And all this has to be done with a limited amount of time and cognitive resources.

¹ Data from the UN agency for ICT (ITU). Accessed via: <https://news.itu.int/itu-statistics-leaving-no-one-offline/>

² Digital News Report 2019, Reuters Institute. Accessed via: https://ora.ox.ac.uk/objects/uuid:18c8f2eb-f616-481a-9dff-2a479b2801d0/download_file?file_format=pdf&safe_filename=reuters_institute_digital_news_report_2019.pdf&type_of_work=Report

³Ibid.

⁴ For a special issue of *Synthese* on evidence amalgamation in the sciences, see Fletcher, Landes, and Poellinger (2018).

Even for those who are aware of these hurdles, this is not an easy task, as information can be manipulated on purpose by bad actors (Weatherall, O'Connor, and Bruner, 2018) or distorted by the arrangement of the agents within social network structures (Stewart, Mosleh, Diakonova, Arechar, Rand, and Plotkin, 2019). The consequences of holding false beliefs go far beyond being punished on the individual level for being gullible and clueless. As members of society, our false beliefs often result in bad decisions – individual and collective – which in turn can cause all kinds of large-scale damage: from the election of corrupt leaders to ecological disaster. Ultimately, apparently inoffensive false beliefs, such as that the Earth is flat or that species are not subject to natural selection, reveal an inadequate underlying epistemic machinery, which will result in more severe consequences when employed for the judgement of more serious matters.

This thesis, however, does not try to give practical advice on how to form reliable beliefs. The issues mentioned above serve as the underlying motivation for our work, but we explore them from a logical and multi-agent systems perspective, occasionally drawing from sources in the epistemology literature. Our models focus especially on the issue of choosing a belief attitude (or doxastic attitude) towards atomic propositions, by considering the existence of evidence for and against such propositions, and the evidence or beliefs possessed by peers. This endeavor, therefore, has significant intersections with areas of study such as social choice theory (Arrow, 1951; Gibbard, 1973) and opinion aggregation (Baltag, Christoff, Rendsvig, and Smets, 2016b; Endriss and Grandi, 2017; Dietrich and List, 2016). The way evidence is depicted in the next chapters is quite simple, but we will see that even in those scenarios there are *rationality constraints* that should be respected.

1.1 Historical Remarks

Despite plenty of evidence showing that we humans have difficulties in performing certain logical reasoning tasks,⁵ logic – in the broad sense of the word – is rooted in human intuition and in the way the human mind works. Besides Greece, some forms of logical studies have appeared already in ancient times in India (Matilal, 1999) and China (see Zhang and Liu (2007)). Even before the birth of logic as a discipline, the Socratic method, seen in

⁵ For example, see the famous card selection task of Wason (1968), but also a response in Stenning and van Lambalgen (2001).

the writings of Socrates' pupil Plato, already shows how logic has been extensively (but implicitly) used since the beginning of philosophy. Logic is an integral part of philosophy, science and sound reasoning in general. Even some grammatical constructs found in natural language reflect basic logical concepts, as shown in Zhang and Liu (2007) and evidenced by the numerous attempts at formalising natural language. However, it is only in the 4th century BC with Plato's student and Alexander's tutor, Aristotle, that logic is officially inaugurated as a field of study in itself, in a collection of six works later called the *Organon* by his successors. The word *organon* means *tool*, *instrument*. Logic is a tool used for separating valid from invalid inferences.

Thousands of years later, Boole, De Morgan, Cantor (1878), Frege (1893), Whitehead and Russell (1910), Gödel (1930), Hilbert and Bernays (1934) laid the groundwork in set theory and the foundations of mathematics that led to modern symbolic/mathematical logic. Logic developed fast during the twentieth century. Lewis and Langford (1932) introduced the famous S1-S5 systems of modal logic. Saul Kripke began his work on the semantics of modal logic while still in high school. Starting in 1959, as a 19-year-old undergraduate student, he published a series of papers (Kripke, 1959, 1963a,b) introducing his *possible worlds* semantics (also known as Kripke semantics), which is widely used for modal logic nowadays – including this thesis. Von Wright (1951) brought important contributions to philosophical and modal logic, and his student Hintikka (1962) is considered one of the founders of formal epistemic/doxastic logics, the logics of knowledge and belief, respectively. These are the types of logic in which we are mostly interested here. Hintikka and Beth (1955) independently developed semantic tableaux, the type of proof system for modal (and first-order) logic used in this thesis.

Epistemic and doxastic logics have been quite successful, especially within computer science applications such as model checking, cryptography and security (Bieber, 1990; Syverson and van Oorschot, 1994; Dechesne and Wang, 2007; Boureau, Cohen, and Lomuscio, 2009; Balliu, Dam, and Le Guernic, 2011), but they have also been widely studied by philosophers (Chisholm, 1963; Stalnaker, 1984; Williamson, 2002). Epistemic logic as a tool helps us find out what agents *can* (or *could*, in theory) know about the world, about their own knowledge, and about the knowledge of others. That is why it is very useful for cyber security applications, where the mere knowability of something might represent a potential threat.

1.2 Logical Omniscience

The knowledge/knowability distinction mentioned above is connected to a well-known problem (or feature) of epistemic logic, the so-called *logical omniscience* (Hintikka, 1979). Epistemic logic has a built-in assumption that agents know all the logical consequences of what they know, which includes all logical truths. Formally, if $\models \varphi \rightarrow \psi$, then $\models K_i\varphi \rightarrow K_i\psi$, where $K_i\chi$ denotes that *agent i knows χ* . With this idealisation, epistemic logic shows to be more a logic of *knowability* than of *knowledge*. But the question of what are the appropriate underlying logics of knowledge and belief – if any – remains under debate. This is known as the problem of normativity of logic (see MacFarlane (2004)).

A brief explanation of why logical omniscience holds is the following. An agent i is said to know φ ($K_i\varphi$) if and only if φ holds in all worlds that i considers possible – denote this set of possible worlds by X . But, according to possible worlds semantics, what holds in any given world is closed under logical consequence: if ψ and $\psi \rightarrow \chi$ hold in a world w , then χ also holds there. So, if $K_i\varphi$, then φ holds in all worlds that i considers possible (i.e. all worlds in X), and since all worlds are closed under deduction, if $\varphi \rightarrow \psi$ holds in all $w \in X$, then ψ will also hold in all $w \in X$, and therefore $K_i\psi$ will hold as well. So it is clear that logical omniscience stems from Kripke’s possible world semantics.

The problem of logical omniscience has been thoroughly studied, and various formalisms have been proposed to solve or circumvent it (Levesque, 1984; Fagin and Halpern, 1987; Baltag, Renne, and Smets, 2014; Alechina and Logan, 2002; Ågotnes and Alechina, 2007; Solaki, 2017). In fact, in the beginning the goal of this PhD project was to deal with logical omniscience directly. As those proposed solutions show, however, there seems to be an inescapable trade-off between omniscience and logical power operating in epistemic logics, in the sense that the less “omniscient” the agents are, the more trivial their doxastic states become. In view of this obstacle, we decided to slightly change the direction of our research from directly solving logical omniscience to studying logics of evidence and belief, such as van Benthem and Pacuit (2011b); Baltag, Bezhanishvili, Özgün, and Smets (2016a); Shi, Smets, and Velázquez-Quesada (2018b).

1.3 Many-valued Logics and Evidence

Our departure point, however, was not the articles just mentioned, but Belnap’s four-valued logic (Belnap, 1977). In that logic, propositions can assume four values: *true*, *false*, *both* true and false or *neither*. Belnap explains that these values need not be viewed as entailing propositions that are true and false at the same time or having no truth-value, but instead they can be seen as regarding *information*. Then, value *both* is understood as there being information both for and against a proposition, and value *neither* as the absence of any information. From this base, we could model agents that are exposed to incomplete and inconsistent information, and investigate how those agents would form beliefs in the most rational way possible.

A logic for which *ex falsum quodlibet* ($\{\varphi, \neg\varphi\} \models \psi$, for all φ and ψ) does not hold is called a *paraconsistent logic* (Jaśkowski, 1948; Asenjo, 1966; Smiley, 1959; da Costa, 1974; Anderson and Belnap, 1975; Dunn, 1976). In other words, in that type of logic, contradictions do not lead to trivialisation. All the logics developed in this thesis are paraconsistent.

A first technical obstacle faced in this project was to completely adapt and formalise Belnap’s logic in a multi-agent modal setting. As a bonus, our first contribution is actually to give new insights on the intuitions and applicability of many-valued modal logics, showing that those logics can be more useful than it was previously thought.⁶

Given that initial setting, we come back to our opening topic about having the right “epistemic machinery” and tackle one of the main issues discussed in this thesis: the problem of *consolidation*. The latter is the name we give to operations that go from evidence states to belief states, in other words, *the process of forming evidence-based beliefs*. By studying this problem, we also make steps in the realm of resource-bounded (epistemic) agents, which is usually one of the common ways of preventing logical omniscience.

1.4 Overview

In this section we summarise the content of each chapter of this thesis. Notice that, as all chapters are based on papers, they can be understood

⁶ As remarked in the conclusion of Fitting (1991), very little has been said about intuitions underlying many-valued modal logics, a situation which, to the best of my knowledge, still persists.

even if read separately.

Chapter 2. A Multi-Agent Four-Valued Dynamic Epistemic Logic In this opening chapter we introduce the *four-valued epistemic logic* (FVEL), the main logic studied in this thesis. With FVEL one can represent multiple agents, the public evidence that exists (or is available) and the knowledge the agents have about the status of this evidence. All the later chapters will build on the formal apparatus developed in this chapter.

The syntax and semantics of FVEL are specified, some of its fragments are identified and some important properties are proven. A tableau proof system is offered, which is proven to be sound and complete. Finally, public announcements are added to the framework.

Chapter 3. Consolidations: Turning Evidence into Belief In the previous chapter, FVEL is introduced. That logic model agents and their evidence, but does not model any beliefs. In this chapter we introduce the concept of *consolidations*, the operations which turn evidence into beliefs.

First, preliminary aspects of evidence and the rationale for consolidations are discussed. Next, formal consolidation operations are defined, in the form of *model transformations* from evidence models to epistemic/doxastic models, and some of their properties are proven. Then this new formalism is placed in context with the literature, by comparing FVEL consolidations with consolidations defined by van Benthem and Pacuit (2011b).

Chapter 4. Social Consolidations: Evidence and Peerhood In Chapters 2 and 3, FVEL and consolidations are introduced, but those consolidations are quite self-centered: the agents do not take their peers' beliefs into account. In this chapter we draw on the literature from social epistemology on "peer disagreement" and reinterpret FVEL to model situations where the agents look to their peers in order to form more reliable beliefs – after all, the beliefs of others are also evidence, and it is always recommended that we form our beliefs based on the *total evidence*.

Some postulates inspired by Social Choice Theory are discussed and formalised for our setting. A dynamic operator is added, which enables us to formulate some additional but essential postulates, and serves as a basis for future developments on consolidations that take amounts of evidence

into account. The main technical result is a characterisation of a class of consolidations satisfying most of the rationality postulates.

Chapter 5. Iterative Social Consolidations: Private Evidence

In this final chapter, we build on the previous one by changing one of its assumptions: that the evidence of others is public. Here, the agents have private evidence, and consolidate their beliefs based on their own evidence plus their peers' *beliefs* – which in turn are based on their own evidence and their peers' beliefs, and so on. This leads to an iterative process of consolidation, and hence one of the main problems studied concerns the stabilisation of such processes.

Chapter 6. Conclusions In this chapter we wrap up by discussing the main achievements of the project and some of the desiderata left for future work.

1.5 Prerequisite Knowledge

To facilitate the reading of this thesis, the reader will benefit from knowing a few things in advance (that will not be covered here). The first recommendation is to have a basic understanding of propositional logic (a basic textbook is Barwise and Etchemendy (2000)). The second essential topic is modal logic. The latter is thoroughly explained in the textbook by Blackburn, de Rijke, and Venema (2002), but the basics of logic can also be found there (in the Appendix). It might also be useful to have a basic understanding of epistemic logic and public announcement logic. For those topics, some well-known options are the books by van Ditmarsch, van der Hoek, and Kooi (2007) and Meyer and van der Hoek (1995). In this thesis we will also build upon many-valued (modal) logics. The reader might want to take a look at Belnap (1977) and (Priest, 2008, Chapters 7-9 and 11a), but it should be possible to follow the present work without any previous knowledge on many-valued logics.

All these topics are widely studied, so the readers have plenty of other options at their disposal.

1.6 Publications

All the chapters have been based on papers submitted to international logic-related conferences and/or journals. Chapter 2 is based on Santos

(2018, 2020a):

Y. D. Santos. A dynamic informational-epistemic logic. In: A. Madeira and M. Benevides, editors, *Dynamic Logic. New Trends and Applications (DaLi Workshop)*. Volume 10669 of *Lecture Notes in Computer Science*, pages 64-81. Springer, 2018.

Y. D. Santos. A four-valued dynamic epistemic logic. *Journal of Logic, Language and Information*, 2020.

Chapter 3 is based on Santos (2019):

Y. D. Santos. Consolidation of belief in two logics of evidence. In: P. Blackburn, E. Lorini and M. Guo, editors, *International Conference on Logic, Rationality and Interaction (LORI)*, 2019. Volume 11813 of *Lecture Notes in Computer Science*, pages 57-70. Springer, 2019.

Chapter 4 is based on Santos (2020c):

Y. D. Santos. Social consolidations: Rational belief in a many-valued logic of evidence and peerhood. In A. Herzig and J. Kontinen, editors, *Foundations of Information and Knowledge Systems*, pages 58-78, 2020.

Chapter 5 is based on Santos (2020b):

Y. D. Santos. Iterative Social Consolidations: Forming Beliefs from Many-Valued Evidence and Peers' Opinions. In: G. Primiero, M. Slavkovik, S. Smets, chairs, *ECAI2020 Workshop NETREASON - Reasoning About Social Networks*.

Although not part of this thesis, the following papers (Santos, Matos, Ribeiro, and Wassermann, 2018; Matos, Guimarães, Santos, and Wassermann, 2019) have also been published during this PhD project, and are slightly related to this project in the sense that they also deal with belief dynamics with a resource-bounded perspective in mind.

Y. D. Santos, V. B. Matos, M. M. Ribeiro, and R. Wassermann. Partial meet pseudo-contractions. *International Journal of Approximate Reasoning*, 103: 11 – 27, 2018.

V. B. Matos, R. Guimarães, Y. D. Santos, and R. Wassermann. Pseudo-contractions as gentle repairs. In *Description Logic, Theory Combination, and All That. Essays Dedicated to Franz Baader on the Occasion of His 60th Birthday*, pages 385-403. Springer, 2019. 103: 11 – 27, 2018.

Chapter 2

A Multi-Agent Four-Valued Dynamic Epistemic Logic

2.1 Introduction

Epistemic logic usually features a set of propositions about the world, and models a group of agents and their knowledge (or beliefs) about these propositions. Despite being very useful, this simple model leaves out of the discussion an important factor in the formation of beliefs: evidence.

Belnap (1977) gave an interpretation to *first degree entailment* (FDE) (Dunn, 1976; Priest, 2008), a four-valued logic, centered on the idea of evidence. In that logic, a proposition p can be, besides true or false, *both* (true and false) or *neither* (true nor false). He interpreted these truth-values as the status of information possibly coming from several sources. For example, if *both* is the value assigned to p , then this means that some source supports the truth and another the falsity of p . The value *none* means that no information is available about p . In this way, the valuation already has an epistemic (not ontic) character.

Later, modal extensions of FDE have been developed, such as K_{FDE} (Priest, 2008) and BK (Odintsov and Wansing, 2010). As remarked by Fitting in the conclusion of Fitting (1991), very little has been said about intuitions underlying many-valued modal logics, a situation which seems to persist in the current literature. One of our main objectives in this chapter is to extend the Belnapian epistemic interpretation of FDE to a modal setting. By doing this, we simultaneously achieve two goals: (i) we design a four-valued modal logic suited to model situations where there is a publicly available body of potentially conflicting or incomplete evidence,

and a group of agents that might be uncertain about what evidence is actually there and about what others know about this evidence; and (ii) we provide an epistemic intuition to many-valued logics, contributing to their practical applicability.

K_{FDE} , in fact, is exactly a modal extension of FDE. The logic studied here, however, can express much more within the class of situations described in (i) due to the addition of the connective \sim and (to a lesser extent) public announcements. Nevertheless, much of the intuition presented here (item (ii)) can be transferred to K_{FDE} and BK.

The difficulty in extending the Belnapian interpretation lies in the fact that the valuation already has an epistemic character. The addition of a modal operator – which also has an epistemic nature – to this logic will, then, create two separate epistemic “layers”. Look at the classical epistemic model of Figure 2.1 (left). It represents a situation wherein an agent cannot distinguish between the truth and falsity of proposition p , or, equivalently, wherein the agent does not know whether p .

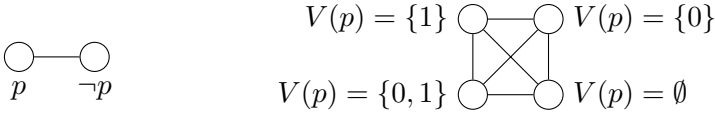


Figure 2.1: An epistemic model (left) and a four-valued epistemic model (right).

Now, compare this situation with the four-valued model of Figure 2.1 (right), where $\{1\}$, $\{0\}$, $\{0, 1\}$ and \emptyset mean, respectively, *true*, *false*, *both* and *none*. What is a plausible interpretation for this model? Here, the agent not only cannot distinguish between worlds where p is *true* or *false*, but also between worlds where it is *neither* true nor false, or *both*. If we adopt an epistemic interpretation of the valuations, what kind of interpretation is left for the operator \Box ? As mentioned before, we should think of the (four-valued) valuation function as representing evidence or information, while the accessibility relations account for the uncertainty of the agents about which evidential state is the correct one.

For example, we can regard the valuation as representing the information about some propositions stored in a database. The database only registers the information it receives, so it is well possible that at first it receives the information that p is *true*, but subsequently it receives (possibly from another source) the information that p is *false*. In this case the database contains contradictory information about p . The accessibility relations may symbolise, in this case, the knowledge of a user of this database.

The user may be in a state like the one in Figure 2.1 (right), where she considers it possible that the database is in any of the four possible states regarding p . That is the natural extension of the Belnapian interpretation to a modal setting: models depicting agents that are uncertain about evidential states. Notice that those agents do not possess knowledge about facts, but only a superficial knowledge about evidence itself.¹

Another example not involving databases can be given. Suppose that Anne lives in Groningen, and that she usually informs herself of the weather by watching the local television’s newscast. Let proposition G mean that *It will rain in Groningen tonight*. Now, imagine the situation in which Anne heard that G in the newscast of Channel 1, but $\neg G$ in the newscast of Channel 2. The status of G for Anne is now contradictory. In this logic, however, we are not going to talk about ontic literals such as G and $\neg G$, but only about epistemic literals g and $\neg g$, meaning *there is evidence for G* and *there is evidence against G* , respectively. The confusion of Anne about G is denoted by assigning value *both* to g . Moreover, assuming that Anne is always up to date with the weather news from Channels 1 and 2, she will always know what is the four-valued status of g . So, in a state where G was announced to be both true and false, Anne is aware of that. She does not consider a world to be possible where only $\neg G$ was announced, for she already knows this is not the case. Bart, who lives in Rotterdam, on the other hand, does not have access to Groningen weather information in his local newscast, so he considers all of the four values to be possible for g . Now we can have a formula like $\Box_a(g \wedge \neg g)$, meaning that *Anne knows that there is information supporting both the truth and the falsity of g* .

The rest of this chapter will explore in detail this logic of evidence, which we will simply call *four-valued epistemic logic* (FVEL, in short). In Section 5.2 we define the syntax and semantics of the logic, and present some of its basic properties. The formalism is a variant of the logic BK (Odintsov and Wansing, 2010). In Section 2.3 we present a sound and complete tableau system. In Section 2.4 we show some correspondence results concerning classical epistemic logic axioms. As a fundamental part of modern dynamic epistemic logics, public announcements are added to FVEL in Section 2.5, and are shown not to increase expressivity. We also extend the tableau system with rules for public announcements, and prove completeness. In the interpretation proposed here, public announcements

¹ Usually there is a link between evidence and reality, but in this formalism, once evidence is allowed to be misleading, these concepts have to be fully independent.

will only have the effect of changing agents' knowledge about evidence, but not the evidence itself.² This clearly leaves open the possibility for other kinds of dynamics, but they are not explored here. In Section 2.6 we comment on related work. This chapter is placed among a rapidly growing body of literature on the topic of logics of evidence, some of which are discussed in Section 2.6.1. Conclusions and possibilities for future work are found in Section 2.7.

2.2 Four-Valued Epistemic Logic

In this section, we will define the syntax and the semantics of the logical language being examined.

Definition 2.1 (Syntax) *Let At be a countable set of atomic propositions and A a finite set of agents.³ A well-formed formula φ in our language \mathcal{L} is inductively defined as follows:*

$$\varphi ::= p \mid \sim \varphi \mid \neg \varphi \mid (\varphi \wedge \varphi) \mid \Box_i \varphi$$

with $p \in At$ and $i \in A$. The following abbreviations will be employed throughout the text: $(\varphi \vee \psi) \stackrel{\text{def}}{=} \neg(\neg\varphi \wedge \neg\psi)$; $(\varphi \rightarrow \psi) \stackrel{\text{def}}{=} (\neg\varphi \vee \psi)$; $(\varphi \leftrightarrow \psi) \stackrel{\text{def}}{=} ((\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi))$; $(\varphi \tilde{\vee} \psi) \stackrel{\text{def}}{=} \sim(\sim\varphi \wedge \sim\psi)$; $(\varphi \tilde{\rightarrow} \psi) \stackrel{\text{def}}{=} (\sim\varphi \tilde{\vee} \psi)$; $\varphi \tilde{\leftrightarrow} \psi \stackrel{\text{def}}{=} (\varphi \tilde{\rightarrow} \psi) \wedge (\psi \tilde{\rightarrow} \varphi)$; $\Diamond_i \varphi \stackrel{\text{def}}{=} \neg \Box_i \neg \varphi$. Parentheses will be omitted when there is no room for ambiguity.

Later, we will refer to several fragments of the language \mathcal{L} :

Definition 2.2 (FVEL Fragments) *Consider the following fragments of \mathcal{L} :*

1. Propositional Fragment: *the subset of \mathcal{L} not containing formulas with symbols \Box_i , for any $i \in A$.*
2. FDE Fragment: *the subset of the Propositional Fragment above not containing formulas with the symbol \sim .*
3. KFDE Fragment: *the subset of \mathcal{L} not containing formulas with \sim .*

² An evidence-changing operation would entail modification in the valuation function. We do that in Chapter 4 (Section 4.5), but under a different interpretation of FVEL.

³ In AI and computer science the set of agents is usually taken to be finite (cf. Meyer and van der Hoek (1995); Fagin, Halpern, Moses, and Vardi (1995)).

4. K Fragment: the subset of \mathcal{L} not containing formulas with the symbol \neg .

Definition 2.3 (Semantics) Given the non-empty finite set $A = \{1, 2, \dots, n\}$ of agents, an interpretation is a tuple $\mathcal{M} = (S, R, \mathcal{V})$, where S is a non-empty set of states, $R = (R_1, R_2, \dots, R_n)$ is an n -tuple of binary relations on S and $\mathcal{V} : At \times S \rightarrow 2^{\{0,1\}}$ is a valuation function that assigns to each proposition one of four truth values: $\{0\}$ is false (f), $\{1\}$ is true (t), \emptyset is none (n) and $\{0, 1\}$ is both (b).⁴ ⁵ With $p \in At$, $s \in S$, $i \in A$ and $\varphi, \psi \in \mathcal{L}$, the satisfaction relation \models is inductively defined as follows:

$\mathcal{M}, s \models p$	iff $1 \in \mathcal{V}(p, s)$
$\mathcal{M}, s \models \neg p$	iff $0 \in \mathcal{V}(p, s)$
$\mathcal{M}, s \models (\varphi \wedge \psi)$	iff $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, s \models \psi$
$\mathcal{M}, s \models \neg(\varphi \wedge \psi)$	iff $\mathcal{M}, s \models \neg\varphi$ or $\mathcal{M}, s \models \neg\psi$
$\mathcal{M}, s \models \Box_i \varphi$	iff for all $t \in S$ s.t. sR_it , it holds that $\mathcal{M}, t \models \varphi$
$\mathcal{M}, s \models \neg\Box_i \varphi$	iff there is a $t \in S$ such that sR_it and $\mathcal{M}, t \models \neg\varphi$
$\mathcal{M}, s \models \sim \varphi$	iff $\mathcal{M}, s \not\models \varphi$
$\mathcal{M}, s \models \neg\sim \varphi$	iff $\mathcal{M}, s \models \varphi$
$\mathcal{M}, s \models \neg\neg\varphi$	iff $\mathcal{M}, s \models \varphi$

Below we derive the truth conditions for $\vee, \rightarrow, \leftrightarrow$ and \Diamond .⁶

$\mathcal{M}, s \models (\varphi \vee \psi)$	iff $\mathcal{M}, s \models \varphi$ or $\mathcal{M}, s \models \psi$
$\mathcal{M}, s \models (\varphi \rightarrow \psi)$	iff $\mathcal{M}, s \models \neg\varphi$ or $\mathcal{M}, s \models \psi$
$\mathcal{M}, s \models (\varphi \leftrightarrow \psi)$	iff $(\mathcal{M}, s \models \neg\varphi \text{ or } \mathcal{M}, s \models \psi)$ and $(\mathcal{M}, s \models \varphi \text{ or } \mathcal{M}, s \models \neg\psi)$

⁴ We use the symbols \mathcal{V} and \mathcal{M} for four-valued valuations and FVEL models, respectively, whereas V and M will be reserved for other (usually bivalent) valuations and models in some parts of this and other chapters.

⁵ Although we work with arbitrary accessibility relations throughout most of this chapter for generality, Section 2.4 presents some results that illustrate the effects of restricting R .

⁶ Note that $\mathcal{M}, s \models \sim\Box_i\sim\varphi$ iff $\mathcal{M}, s \models \neg\Box_i\neg\varphi$, but $\mathcal{M}, s \models \neg\sim\Box_i\sim\varphi$ iff for all t such that sR_it , $\mathcal{M}, t \models \sim\varphi$, whereas $\mathcal{M}, s \models \neg\neg\Box_i\neg\varphi$ iff for all t such that sR_it , $\mathcal{M}, t \models \neg\varphi$.

$$\begin{aligned}
\mathcal{M}, s &\models \Diamond_i \varphi && \text{iff there is a } t \in S \text{ such that } sR_i t \text{ and } \mathcal{M}, t \models \varphi \\
\mathcal{M}, s &\models \neg \Diamond_i \varphi && \text{iff for all } t \in S \text{ s.t. } sR_i t, \text{ it holds that } \mathcal{M}, t \models \neg \varphi
\end{aligned}$$

Now, we can talk not only about 4-valued atoms, but also about 4-valued formulas in general.

Definition 2.4 (Extended Valuation Function) *We define the extended valuation function $\overline{\mathcal{V}} : \mathcal{L} \times S \rightarrow 2^{\{0,1\}}$ as follows:*

$$\begin{aligned}
1 \in \overline{\mathcal{V}}(\varphi, s) &\text{ iff } \mathcal{M}, s \models \varphi \\
0 \in \overline{\mathcal{V}}(\varphi, s) &\text{ iff } \mathcal{M}, s \models \neg \varphi
\end{aligned}$$

Using the above definition, we say that a formula φ has value *both* at s , for example, if and only if $\overline{\mathcal{V}}(\varphi, s) = \{0, 1\}$, which is the case when both $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, s \models \neg \varphi$. Truth and falsity of formulas are evaluated independently, and for that reason we define semantic conditions for each negated formula separately. Even though the semantics of \neg above is defined case by case,⁷ the connective is still truth-functional, as we will see in Section 2.2.3.

2.2.1 Intended Readings of Formulas

Since the semantics of FVEL is non-compositional, the readings of its formulas will be non-compositional as well. The four values combined with a modality plus an additional negation also create further complications, which are clarified below.

The intended readings of purely propositional formulas follow the Belnapian view of FDE as talking about evidence: non-modal formulas φ and $\neg \varphi$ are read as *there is evidence for φ* and *there is evidence against φ* , respectively. The second negation (\sim) is classical: $\sim \varphi$ means that *it is not the case that φ* . We can see the propositional fragment of FVEL (Definition 2.2.1) as a logic that *preserves evidence*, a concept mentioned in a recent paper by Carnielli and Rodrigues (2019).

⁷ Odintsov and Wansing (2010) uses two support relations \models^+ and \models^- , dispensing with the case by case semantics. While their formalism and ours have the same expressivity, ours has a larger number of formulas (see more on this comparison in Section 2.6).

The \Box operator inherits its natural reading from epistemic logic, but we have to remember that propositional formulas are read as statements about evidence. So, for example, $\Box_i\varphi$ and $\Box_i\neg\varphi$ will have the intended meaning of *agent i knows that there is evidence for φ* and *agent i knows that there is evidence against φ* , respectively. Nested \Box formulas are read in the expected way: $\Box_i\Box_j\varphi$ just means that *agent i knows that j knows that φ* (again, the same remark about the reading of a propositional φ applies here).

It is worth noting that all formulas of FVEL are four-valued according to our semantics, not only the propositional ones. What does it mean, then, to say that $\Box_i\varphi$ has value *both* or *none*? With the intuitions provided so far (for instance, the database example), it certainly does not make sense to say that an agent knows and doesn't know φ at the same time. Breaking down the semantics, however, we can see that a statement such as $\overline{\mathcal{V}}(\Box_i\varphi, s) = \{0, 1\}$ means, in fact, that $\mathcal{M}, s \models \Box_i\varphi$ and $\mathcal{M}, s \models \neg\Box_i\varphi$, which is equivalent to $\mathcal{M}, s \models (\Box_i\varphi) \wedge (\Diamond_i\neg\varphi)$ (recall the definition of \Diamond_i), that is, *agent i knows that there is evidence for φ but considers possible that there is evidence against it as well*. Likewise, we find that $\overline{\mathcal{V}}(\Box_i\varphi, s) = \emptyset$ means that *agent i considers it possible that there is no evidence for φ and she knows that there is no evidence against it*.

Finally, we have to remark that any formula with a \neg in front of a \Box , such as $\neg\Box_ip$, does not have a straightforward reading. Nevertheless, we claim (without proof) that one can always convert such formulas into equivalent ones where \neg is restricted to propositional subformulas,⁸ allowing for an intuitive reading. For example, $\neg\Box_ip$ is equivalent to $\Diamond_i\neg p$, so it is simply read as *agent i considers possible that there is evidence against p* .

2.2.2 Validity and Entailment

We say that $\mathcal{M} \models \varphi$ if and only if $\mathcal{M}, s \models \varphi$ for all $s \in S$, where $\mathcal{M} = (S, R, \mathcal{V})$. A formula φ is *valid* ($\models \varphi$) if and only if $\mathcal{M} \models \varphi$ for all models \mathcal{M} . A *frame* is a pair $\mathcal{F} = (S, R)$. We say a formula φ is *valid in a frame* $\mathcal{F} = (S, R)$, that is, $\mathcal{F} \models \varphi$, if and only if, for all valuations \mathcal{V} , it holds that $\mathcal{M} \models \varphi$, where $\mathcal{M} = (S, R, \mathcal{V})$ (and we say \mathcal{M} is based

⁸ One can devise a recursive translation φ^N that maps any formula φ into an equivalent one, in the language of \mathcal{L} plus the abbreviations \vee and \Diamond , where all occurrences of \neg are in front of atoms, in the same vein as Fitting (2017, Definition 7.3). A proof that φ^N is equivalent to φ can be done by induction in the complexity of φ (in each translation rule the formula under \neg is mapped to subformulas of it). That φ^N is equivalent to φ can be shown by proving that each translation rule generates an equivalent formula. (Equivalence is defined in Section 2.2.3.)

on frame \mathcal{F}). If for all models \mathcal{M} and all states s it is the case that $\mathcal{M}, s \models \psi$ for all $\psi \in \Sigma$ implies $\mathcal{M}, s \models \varphi$, we say that $\Sigma \models \varphi$ (φ is a *logical consequence* of Σ). If $\Sigma \models \varphi$ holds, we say it is a *valid entailment* or a *valid inference*.

Logical consequence in classical logics preserves *truth*. Many-valued logics generalise this idea, with their logical consequence preserving *designated values*. Following Priest (2008) and others, we define $\{1\}$ and $\{0, 1\}$ as designated values, and \emptyset and $\{0\}$ as *non-designated values*. Notice that statements of the form $\mathcal{M}, s \models \varphi$, which can be translated to $1 \in \overline{\mathcal{V}}(\varphi, s)$, really are just saying that φ is designated. A formula is called designated (non-designated) with respect to a model \mathcal{M} and state s if it has a designated (non-designated) value at \mathcal{M}, s , i.e. $\overline{\mathcal{V}}(\varphi, s)$ is designated (non-designated). If one wants to check whether a formula φ has precisely the value *true* (or whatever other value) one just has to check two satisfaction statements: $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, s \models \neg\varphi$.

2.2.3 Basic Properties of FVEL

Connectives and Notable Fragments

Now we build the truth tables for the truth-functional connectives according to the truth definitions given above. The ones for \wedge , \neg and \vee turn out to be identical to the ones in Priest (2008, p.146).

φ	n	f	t	b
$\neg\varphi$	n	t	f	b

Table 2.1: $\neg\varphi$.

φ	n	f	t	b
$\sim\varphi$	t	t	f	f

Table 2.2: $\sim\varphi$.

$\varphi \backslash \psi$	n	f	t	b
n	n	f	n	f
f	f	f	f	f
t	n	f	t	b
b	f	f	b	b

Table 2.3: $\varphi \wedge \psi$.

$\varphi \backslash \psi$	n	f	t	b
n	n	n	t	t
f	n	f	t	b
t	t	t	t	t
b	t	b	t	b

Table 2.4: $\varphi \vee \psi$.

$\varphi \backslash \psi$	n	f	t	b
n	n	n	t	t
f	t	t	t	t
t	n	f	t	b
b	t	b	t	b

Table 2.5: $\varphi \rightarrow \psi$.

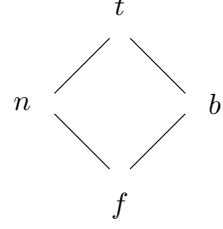


Figure 2.2: Lattice L4.

Example for Table 2.1 ($\neg \mathbf{b} = \mathbf{b}$): $\overline{\mathcal{V}}(\varphi, s) = \{0, 1\}$ iff $0 \in \overline{\mathcal{V}}(\varphi, s)$ and $1 \in \overline{\mathcal{V}}(\varphi, s)$ iff $\mathcal{M}, s \models \neg\varphi$ and $\mathcal{M}, s \models \varphi$ iff $\mathcal{M}, s \models \neg\varphi$ and $\mathcal{M}, s \models \neg\neg\varphi$ iff $1 \in \overline{\mathcal{V}}(\neg\varphi, s)$ and $0 \in \overline{\mathcal{V}}(\neg\varphi, s)$ iff $\overline{\mathcal{V}}(\neg\varphi, s) = \{0, 1\}$.

Example for Table 2.4 ($\mathbf{n} \vee \mathbf{b} = \mathbf{t}$):⁹ Recall that disjunction is defined in terms of conjunction and negation. $\mathcal{M}, s \models \neg(\neg\varphi \wedge \neg\psi)$ iff $\mathcal{M}, s \models \neg\neg\varphi$ or $\mathcal{M}, s \models \neg\neg\psi$ iff $\mathcal{M}, s \models \varphi$ or $\mathcal{M}, s \models \psi$ iff $1 \in \overline{\mathcal{V}}(\varphi, s)$ or $1 \in \overline{\mathcal{V}}(\psi, s)$, which is true, for $\overline{\mathcal{V}}(\psi, s) = \{0, 1\}$. $\mathcal{M}, s \models \neg\neg(\neg\varphi \wedge \neg\psi)$ iff $\mathcal{M}, s \models \neg\varphi \wedge \neg\psi$ iff $\mathcal{M}, s \models \neg\varphi$ and $\mathcal{M}, s \models \neg\psi$ iff $0 \in \overline{\mathcal{V}}(\varphi, s)$ and $0 \in \overline{\mathcal{V}}(\psi, s)$, which is false, for $\overline{\mathcal{V}}(\varphi, s) = \emptyset$. Therefore $\mathcal{M}, s \models \varphi \vee \psi$ holds, but $\mathcal{M}, s \models \neg(\varphi \vee \psi)$ does not, thus $1 \in \overline{\mathcal{V}}(\varphi \vee \psi)$ and $0 \notin \overline{\mathcal{V}}(\varphi \vee \psi)$, hence $\overline{\mathcal{V}}(\varphi \vee \psi) = \{1\}$.

The next observation follows from the truth tables and the semantics of \Box :

Observation 2.5 *For all models \mathcal{M} , if for all states s and $p \in \text{At}$ it holds that $\mathcal{V}(p, s) \notin \{\emptyset, \{0, 1\}\}$, then for all states t and $\varphi \in \mathcal{L}$ it holds that $\overline{\mathcal{V}}(\varphi, t) \notin \{\emptyset, \{0, 1\}\}$.*

If we leave \neg out (fragment of Definition 2.2.4), we are left with (the validities of) classical modal logic, with designated values behaving as *true*, and non-designated values behaving as *false*.

Moreover, observing these truth tables, we notice that the fragment resulting from leaving \sim and \Box out (Definition 2.2.2) behaves exactly as FDE (Dunn, 1976; Priest, 2008).¹⁰ Conjunction and disjunction are given

⁹ What $n \vee b = t$ is saying is that if we have propositions φ and ψ such that we have no evidence about the first but we do have evidence both for and against the second, then we have only evidence *for* (and not against) $\varphi \vee \psi$. Does this make sense? Indeed, evidence against this disjunction would require evidence against both its disjuncts, but we have no evidence whatsoever about φ .

¹⁰ Sometimes we omit the agent index from the \Box operator, for simplicity. Everything that is said here is generalisable to a language with multiple boxes.

$\varphi \backslash \psi$	n	f	t	b
n	f	f	t	t
f	f	f	t	t
t	t	t	t	t
b	t	t	t	t

Table 2.6: $\varphi \tilde{\vee} \psi$.

$\varphi \backslash \psi$	n	f	t	b
n	t	t	t	t
f	t	t	t	t
t	f	f	t	t
b	f	f	t	t

Table 2.7: $\varphi \tilde{\rightarrow} \psi$.

by the meet and join, respectively, of the values in the lattice depicted in Figure 2.2, called $L4$ in Belnap (1977). Now, adding the modal operator \Box to FDE we obtain K_{FDE} (our fragment of Definition 2.2.3), a logic which Priest (2008) has studied. He provides a complete tableau system for this logic. Moreover, he shows that this logic contains no validities, as is the case for FDE itself. In the class of four-valued Kripke models, FVEL is strictly more expressive than K_{FDE} : formulas such as $\sim p$ can only be expressed in the former.

We can also build the truth tables for $\tilde{\vee}$ and $\tilde{\rightarrow}$. Despite these connectives being binary functions accepting two four-valued parameters, they behave analogously to their classical (Boolean) counterparts. They can be viewed as a composition of a function that compresses designated values into *true* and non-designated values into *false* (just like a double application of the operator \sim) with the corresponding Boolean function. In other words, if *or* is classical disjunction and *imp* is classical implication, $x \tilde{\vee} y = or(\sim \sim x, \sim \sim y)$ and $x \tilde{\rightarrow} y = imp(\sim \sim x, \sim \sim y)$. It is also relevant to remark that when the operands take on only classical values, both pairs of operators (\vee, \rightarrow and $\tilde{\vee}, \tilde{\rightarrow}$) behave exactly alike.

The propositional part of FVEL (Definition 2.2.1) can be considered a fragment of the bilattice logic in Arieli and Avron (1996), and the later is strictly more expressive than the former. Moreover, our modal and public announcement extensions have many similarities with BPAL (Rivieccio, 2014a) (more on these comparisons in Section 2.6.2).

Validities

We can define \top , a validity, as $\top \stackrel{\text{def}}{=} (p \vee \sim p)$. While FDE has no validities, FVEL has an infinity of them, including \top . Moreover, all propositional tautologies (built with \sim) are still validities in FVEL, as expected, but there are other valid formulas with both \sim and \neg , such as $\sim p \vee \neg \sim p$. All validities in FVEL have the connective \sim or one of its derivative connectives ($\tilde{\vee}$ and $\tilde{\rightarrow}$). Some standard modal validities are also valid in FVEL when

built using \leftrightarrow , e.g. $\Box(\varphi \wedge \psi) \leftrightarrow (\Box\varphi \wedge \Box\psi)$ and $\Diamond(\varphi \vee \psi) \leftrightarrow (\Diamond\varphi \vee \Diamond\psi)$.

A logic L' is a *conservative extension* (C.E.) of a logic L iff the language of L' contains the language of L and all validities of L are also validities of L' . Building on the observations of Section 2.2.3, we can establish the following. K_{FDE} is a C.E. of FDE, and FVEL is a C.E. of K_{FDE} . FVEL is also a C.E. of classical modal logic (taking \sim as classical negation), which is the fragment of Definition 2.2.4. Bilattice logic (Arieli and Avron, 1996) is a C.E. of the propositional fragment of FVEL (Definition 2.2.1).

Equivalence

Logical equivalence (sameness in truth value) cannot be expressed by $\varphi \leftrightarrow \psi$ in FVEL. Look at Table 2.8. The diagonal should be *designated*, and the rest *non-designated*. In fact, in this case even the biconditional connective (\leftrightarrow) derived using \sim instead of \neg does not give a truth table which is designated in the diagonal and non-designated everywhere else, for it treats $\{1\}$ and $\{0, 1\}$ as equals (and the same goes for \emptyset and $\{0\}$), resulting in a weaker type of equivalence (see Table 2.9).

$\varphi \backslash \psi$	n	f	t	b
n	n	n	n	t
f	n	t	f	b
t	n	f	t	b
b	t	b	b	b

Table 2.8: $\varphi \leftrightarrow \psi$.

$\varphi \backslash \psi$	n	f	t	b
n	t	t	f	f
f	t	t	f	f
t	f	f	t	t
b	f	f	t	t

Table 2.9: $\varphi \leftrightarrow \psi$.

φ	φ^n	φ^f	φ^t	φ^b
n	t	f	f	f
f	f	t	f	f
t	f	f	t	f
b	f	f	f	t

Table 2.10: φ^n , φ^f , φ^t and φ^b .

$\varphi \backslash \psi$	n	f	t	b
n	t	f	f	f
f	f	t	f	f
t	f	f	t	f
b	f	f	f	t

Table 2.11: $\varphi \leftrightarrow \psi$.

The reason for adding the classical negation (\sim) to a language which already has a negation operator (\neg) is that this increases the expressivity of the language.¹¹ For instance, we can now define formulas discriminating

¹¹ Interestingly, Girard and Tanaka (2016) show that the standard definition of $p \rightarrow q$ as $\neg p \vee q$ does not suffice to prove reduction axioms for public announcements when

which of the four truth values a formula φ has: $\varphi^n \stackrel{\text{def}}{=} (\sim\varphi \wedge \sim\neg\varphi)$; $\varphi^f \stackrel{\text{def}}{=} \sim\sim(\sim\varphi \wedge \neg\varphi)$; $\varphi^t \stackrel{\text{def}}{=} \sim\sim(\varphi \wedge \sim\neg\varphi)$; $\varphi^b \stackrel{\text{def}}{=} \sim\sim(\varphi \wedge \neg\varphi)$. As can be seen in Table 2.10, φ^i is true if and only if φ has truth value i , for $i \in \{n, f, t, b\}$, and false otherwise. Now we can read $\Box_a \varphi^i$ as *Agent a knows that the status of evidence for φ is i* (where $i \in \{t, f, b, n\}$). Using these connectives, it is easy to see that a stronger notion of logical equivalence can be expressed in FVEL:

$$\varphi \leftrightarrow \psi \stackrel{\text{def}}{=} (\varphi^n \wedge \psi^n) \vee (\varphi^f \wedge \psi^f) \vee (\varphi^t \wedge \psi^t) \vee (\varphi^b \wedge \psi^b)$$

Since this formula is complex and difficult to evaluate, we will often favor the use of metalanguage as follows:

$$\varphi \equiv \psi \stackrel{\text{def}}{=} (\mathcal{M}, s \models \varphi \text{ iff } \mathcal{M}, s \models \psi) \text{ and } (\mathcal{M}, s \models \neg\varphi \text{ iff } \mathcal{M}, s \models \neg\psi),$$

for all models \mathcal{M} and all states s .

The formula $\varphi \leftrightarrow \psi$ is *true* if φ and ψ have the same truth value and *false* otherwise (as shown in Table 2.11).

Proposition 2.6 $\varphi \equiv \psi$ iff $\models \varphi \leftrightarrow \psi$.

Proof Suppose $\varphi \equiv \psi$. Then, by the definition of \equiv , for all models \mathcal{M} and states s , $\mathcal{M}, s \models \varphi$ iff $\mathcal{M}, s \models \psi$ and $\mathcal{M}, s \models \neg\varphi$ iff $\mathcal{M}, s \models \neg\psi$, and therefore φ and ψ have the same truth value (in all states of all models), i.e. for all models \mathcal{M} and states s : $\overline{\mathcal{V}}(\varphi, s) = \overline{\mathcal{V}}(\psi, s)$. This implies (by Table 2.10) that for any model \mathcal{M} and state s , either $\mathcal{M}, s \models \varphi^n \wedge \psi^n$ or $\mathcal{M}, s \models \varphi^f \wedge \psi^f$ or $\mathcal{M}, s \models \varphi^t \wedge \psi^t$ or $\mathcal{M}, s \models \varphi^b \wedge \psi^b$. By the definition of $\varphi \leftrightarrow \psi$, it follows that, for all models \mathcal{M} and states s , $\mathcal{M}, s \models \varphi \leftrightarrow \psi$.

Now for the other direction. Suppose $\models \varphi \leftrightarrow \psi$. Then for every \mathcal{M} and s , $\mathcal{M}, s \models \varphi \leftrightarrow \psi$, which by definition means that either $\mathcal{M}, s \models \varphi^n \wedge \psi^n$ or $\mathcal{M}, s \models \varphi^f \wedge \psi^f$ or $\mathcal{M}, s \models \varphi^t \wedge \psi^t$ or $\mathcal{M}, s \models \varphi^b \wedge \psi^b$. But, by Table 2.10, $\mathcal{M}, s \models \chi^i$ iff $\overline{\mathcal{V}}(\chi, s) = i$, for $i \in \{n, f, t, b\}$. Therefore, if any of the statements of the form $\mathcal{M}, s \models \varphi^i \wedge \psi^i$ mentioned before hold, then $\overline{\mathcal{V}}(\varphi, s) = \overline{\mathcal{V}}(\psi, s)$ (for any model \mathcal{M} and state s), and thus, by the definition of $\overline{\mathcal{V}}$, $\varphi \equiv \psi$ holds. \square

working with an epistemic extension of Priest's three-valued Logic of Paradox. To circumvent that, they introduce an alternative implication. Our classical negation has a similar role with respect to our reduction axioms of Section 2.5.

Equi-satisfiability and equivalence coincide in classical logics, but here they differ. One has to keep this in mind when analysing statements such as $\mathcal{M}, s \models \neg \sim \psi$ (which appears, for example, when checking Riveccio (2014a)’s axiom $\langle \alpha \rangle \mathbf{t} \leftrightarrow \sim \alpha$ and our validity for public announcements (An \neg An)). One might be tempted to replace $\sim \sim \psi$ by ψ (as they are equi-satisfiable) and obtain $\mathcal{M}, s \models \neg \psi$. Although for all models \mathcal{M} and states s we have that $\mathcal{M}, s \models \sim \sim \psi$ iff $\mathcal{M}, s \models \psi$, it holds that $\sim \sim \psi \not\equiv \psi$. So the only simplification possible there is by using the semantic clause for $\neg \sim \varphi$, which gives us $\mathcal{M}, s \models \sim \psi$. We prove later (Proposition 2.24) that substitution of equivalent formulas in φ yields a formula which is equivalent to φ .

Example 2.7 (The coffee example) *Now we describe the situation depicted in Figure 2.3. John (j) knows that there are studies regarding health benefits of coffee consumption, for he often sees headlines about the subject. However, he never cared enough to read those articles, so he is sure that there is evidence for or against (or even both for and against) coffee being beneficial for health (p), but he does not know exactly what is the status of the evidence about p , he only knows that there is some information. Looking at Figure 2.3, one can easily see that $\Box_j((p \wedge \sim p) \vee (\neg p \wedge \sim p) \vee (p \wedge \neg p))$, which is equivalent to $\Box_j(p \vee \neg p)$, holds in the actual world (s_3).*

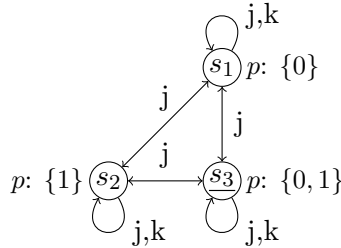


Figure 2.3: Some evidence for p

Kate (k), on the other hand, is a researcher on the effects of coffee on health, and for this reason she knows exactly what evidence is available (notice that her relation R_k has only reflexive arrows). We can see that $\mathcal{M}, s_3 \models \Box_k(p \wedge \neg p)$, that is, in this state, Kate actually knows that there is evidence both for and against the benefits of coffee. Moreover, John knows Kate and her job, so he also knows that she knows about p , whatever its status is (using abbreviations defined in Section 2.2.3: $\Box_j(\Box_k p^f \vee \Box_k p^t \vee \Box_k p^b)$). Likewise, Kate knows that John simply knows that there is some information about p ($\Box_k(\Box_j(p \vee \neg p) \wedge \sim \Box_j(p \wedge \neg p))$).

2.3 Tableaux

A tableau is a structure used to check derivability and theoremhood. In this section we will show how to build a tableau to verify whether $\Sigma \vdash \varphi$ (φ is derivable from Σ in FVEL), where $\Sigma \cup \{\varphi\} \in \mathcal{L}$ and Σ is finite.¹²

Our tableau system is inspired by the one for K_{FDE} given in Priest (2008, p. 248). A tableau is a tree, that is, an irreflexive partially-ordered set (N, E) , where N is a (possibly infinite) set of nodes and $E \subseteq N \times N$, with a unique maximum element $r \in N$, the *root*. A minimum element of N w.r.t. E is called a *leaf*. A (possibly infinite) sequence of nodes where each element is related to the next by E is called a *path*, and a maximal path is called a *branch*. All nodes are of the form $(\psi, +i)$, $(\psi, -i)$ or (ir_mj) , where $\psi \in \mathcal{L}$, $m \in A$ and $i, j \in \mathbb{N}$. A branch is *closed* if it contains nodes $(\psi, +i)$ and $(\psi, -i)$, for some $\psi \in \mathcal{L}$ and $i \in \mathbb{N}$. Otherwise, the branch is *open*.

Let $\Sigma = \{\sigma_1, \dots, \sigma_n\}$, $n \in \mathbb{N}$. A *tableau for $\Sigma \vdash \varphi$* starts with the so-called *initial list for $\Sigma \vdash \varphi$* , defined as follows:

$$\begin{array}{c} (\sigma_1, +0) \\ \vdots \\ (\sigma_n, +0) \\ (\varphi, -0) \end{array}$$

Notice that the initial list is a sequence of nodes forming a single branch (edges omitted above).

The construction of the tableau for $\Sigma \vdash \varphi$ proceeds by way of applying *rules* of the tableau calculus for FVEL (R1-R14 below). On the top of a rule we find the rule's *pre-conditions*: a set of schematic nodes. If we can find a set of nodes in a branch of the tableau - the *target nodes* - that are instances of the pre-conditions, we say that that instance of the rule is *applicable* to those nodes of that branch. The process of *applying a rule*, thus, consists in verifying that it is applicable to a set of target nodes of a branch, and then appending nodes to the leaf of the target branch according to the rule: for rules R5-R7 and R10-R14 one node is appended to the leaf; for rules R1, R4 and R8-R9 two nodes are appended in sequence; and for rules R2-R3 two nodes are appended forming separate

¹² Compare Odintsov and Wansing (2010), which provides a tableau method for BK (discussed in Section 2.6).

branches. If all rules that are applicable to any set of target nodes in a branch have been applied, the branch is *complete*. If all open branches are complete, we say the tableau is *complete*.

We say that $\Sigma \vdash \varphi$ iff there is a tableau for $\Sigma \vdash \varphi$ where all branches are closed. Otherwise we write $\Sigma \not\vdash \varphi$. If $\emptyset \vdash \varphi$ (short: $\vdash \varphi$), we say φ is an *FVEL theorem*.

Whenever $\Sigma \not\vdash \varphi$, we can read off a *countermodel* for $\Sigma \vdash \varphi$ from any complete open branch of a tableau for $\Sigma \vdash \varphi$, which we dub a model *induced* by that branch (following Priest (2008)):

Definition 2.8 *Let b be a complete open branch of a tableau. We say an FVEL model $\mathcal{M} = (S, R, \mathcal{V})$ is induced by b iff \mathcal{M} is such that:*

- $s_i \in S$ iff $(\psi, +i), (\psi, -i), (ir_m j)$ or $(jr_m i)$ appears in b ;
- $s_i R_m s_j$ iff $(ir_m j)$ appears in b ;
- $1 \in \mathcal{V}(p, s_i)$ if $(p, +i)$ appears in b ;
- $1 \notin \mathcal{V}(p, s_i)$ if $(p, -i)$ appears in b ;
- $0 \in \mathcal{V}(p, s_i)$ if $(\neg p, +i)$ appears in b ;
- $0 \notin \mathcal{V}(p, s_i)$ if $(\neg p, -i)$ appears in b .

For future proofs, we also need the following definition (adapted from Priest (2008)):

Definition 2.9 *An FVEL model $\mathcal{M} = (S, R, \mathcal{V})$ is faithful to a branch b of a tableau iff there are functions $f : \mathbb{N} \rightarrow W$ and $g : \mathbb{N} \rightarrow A$ such that for all $i \in \mathbb{N}$, $\varphi \in \mathcal{L}$ and $m \in A$:*

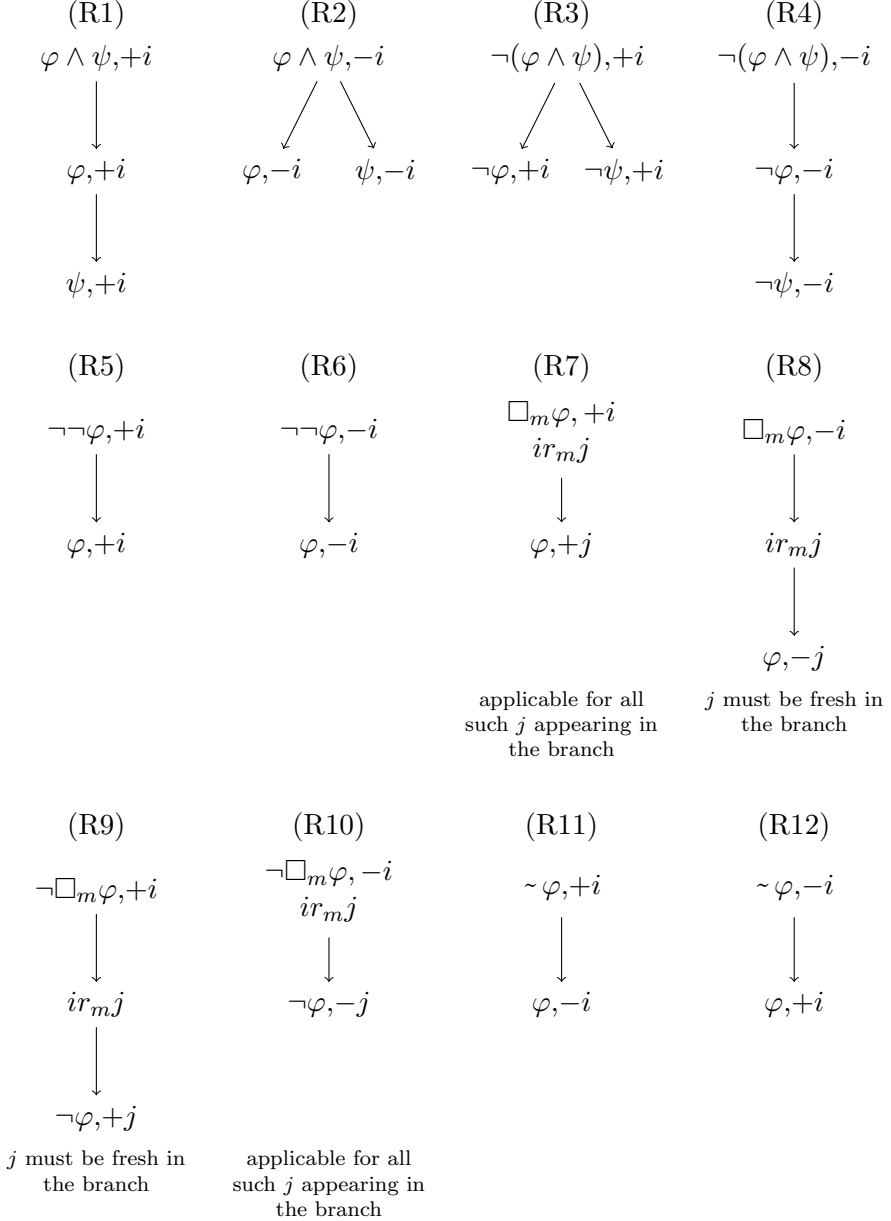
- if $(\varphi, +i)$ is on b , then $1 \in \overline{\mathcal{V}}(\varphi, f(i))$;
- if $(\varphi, -i)$ is on b , then $1 \notin \overline{\mathcal{V}}(\varphi, f(i))$;
- if $(ir_m j)$ is on b , then $f(i)R_{g(m)}f(j)$.

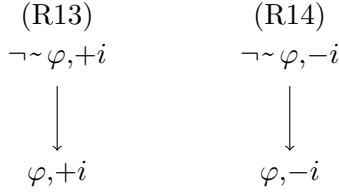
Clearly, a model induced by a branch is faithful to it.

The rules R1-R2 and R5-R8 below are directly taken from the tableau system for K_{FDE} (Priest, 2008, p. 248).¹³ We then modify the rules for negated conjunctions and boxes, rules R3-R4 and R9-R10, respectively,

¹³ The representation of tableau rules here is similar to that in Priest (2008).

since in our language \vee and \Diamond are only abbreviations. Then, we add four more rules for classical negation (R11-R14). This tableau system will be further augmented in Section 2.4 to prove some correspondence results between the tableau system and classes of frames and in Section 2.5 to cope with public announcements.





Figures 2.4 and 2.5 show two examples of proofs using the tableau system. In the first example, one of the branches closes, but no rule is applicable to any nodes in the other branch, which is left open, showing therefore that the derivation $\{\neg(p \wedge \neg q), p\} \vdash q$ (which is equivalent to $\{p \rightarrow q, p\} \vdash q$) does not hold. The second example proves the theorem $\sim(p \wedge \sim p) \wedge \sim(\neg p \wedge \sim \neg p)$.

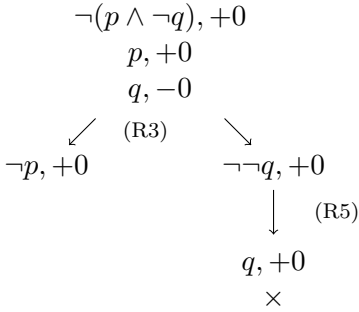


Figure 2.4: A tableau for $\{\neg(p \wedge \neg q), p\} \not\vdash q$, which remains open.

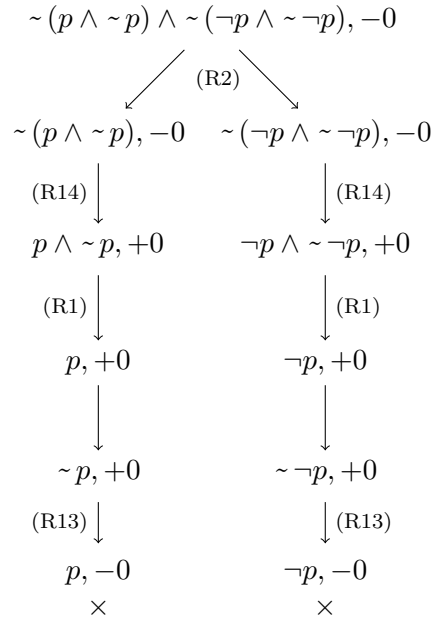


Figure 2.5: A closed tableau for $\vdash \sim(p \wedge \sim p) \wedge \sim(\neg p \wedge \sim \neg p)$: $(p, +0)$ contradicts $(p, -0)$, and $(\neg p, +0)$ contradicts $(\neg p, -0)$.

Before proving soundness, we need to show that the following Soundness Lemma (adapted from Priest (2008, Lemma 11a.9.3)) holds:

Lemma 2.10 *Let b be any branch of a tableau and \mathcal{M} an FVEL model. If \mathcal{M} is faithful to b , and a tableau rule is applied to b , then it produces at least one extension b' such that \mathcal{M} is faithful to b' .*

Proof This proof is just a modification of the proof for the Soundness Lemma for K_{FDE} , in Priest (2008, Lemma 11a.9.3).

Suppose as induction hypothesis that functions $f : \mathbb{N} \rightarrow W$ and $g : \mathbb{N} \rightarrow A$ show \mathcal{M} to be faithful to a branch b containing at least the initial list and the premises of the rule in question. The cases for rules R1-R2 and R5-R8 are already covered there, we need to prove the lemma for the remaining rules; We will use Definition 2.9 instead of the definition of faithful interpretation in Priest (2008, Definition 11a.9.2) (this adaptation only concerns notation).

For rule R3, since \mathcal{M} is faithful to b , we have $1 \in \overline{\mathcal{V}}(\neg(\varphi \wedge \psi), f(i))$. For \mathcal{M} to be faithful to one of the extensions of b (according to Definition 2.9), either $1 \in \overline{\mathcal{V}}(\neg\varphi, f(i))$ or $1 \in \overline{\mathcal{V}}(\neg\psi, f(i))$ has to hold. But, by our semantics, this is precisely what is implied by $1 \in \overline{\mathcal{V}}(\neg(\varphi \wedge \psi), f(i))$.

The case for rule R4 is similar. Since \mathcal{M} is faithful to b , $1 \notin \overline{\mathcal{V}}(\neg(\varphi \wedge \psi), f(i))$. This happens exactly when $1 \notin \overline{\mathcal{V}}(\neg\varphi, f(i))$ and $1 \notin \overline{\mathcal{V}}(\neg\psi, f(i))$, which are the new conditions in the extended branch.

For rule R9, since \mathcal{M} is faithful to b , we have $1 \in \overline{\mathcal{V}}(\neg\Box_m\varphi, f(i))$. By our semantics, there has to be an s such that $f(i)R_ms$ and $M, s \models \neg\varphi$. After applying R9, nodes ir_mj (with j fresh in b) and $\neg\varphi, +j$ are appended to b . Then, take f' identical to f except that $f'(j) = s$. Also, if there was no previous mention of r_m in b , take g' identical to g except that $g'(m) = m$; otherwise just take $g' = g$. Thus, $f'(i)R_{g'(m)}f'(j)$ and $1 \in \overline{\mathcal{V}}(\neg\varphi, f'(j))$, and therefore M is faithful to the extension of b .

For rule R10, we have $1 \notin \overline{\mathcal{V}}(\neg\Box_m\varphi, f(i))$ and $f(i)R_{g(m)}f(j)$. From our semantics, we get that for all t such that $f(i)R_{g(m)}t$, we have $1 \notin \overline{\mathcal{V}}(\neg\varphi, t)$. In particular, we have $1 \notin \overline{\mathcal{V}}(\neg\varphi, f(j))$, which is what we needed to show.

For rule R11, we want to prove that $1 \notin \overline{\mathcal{V}}(\varphi, f(i))$. Since M is faithful, $1 \in \overline{\mathcal{V}}(\sim\varphi, f(i))$. But then $\mathcal{M}, f(i) \models \sim\varphi$, which implies $\mathcal{M}, f(i) \not\models \varphi$ and we are done. The case for R12 is analogous.

For rule R13, we need to prove that $1 \in \overline{\mathcal{V}}(\varphi, f(i))$. But by hypothesis our interpretation \mathcal{M} is faithful to the branch, so $1 \in \overline{\mathcal{V}}(\neg\sim\varphi, f(i))$. The result follows: $\mathcal{M}, f(i) \models \neg\sim\varphi$, then $\mathcal{M}, f(i) \models \varphi$, and we are done. The case for rule R14 is analogous. \square

Similarly, we need the following Completeness Lemma (adapted from Priest (2008, Lemma 11a.9.6)):

Lemma 2.11 *Let b be a complete open branch of a tableau, and let $\mathcal{M} = (S, R, \mathcal{V})$ be an FVEL model induced by b . Then:*

- *If $(\varphi, +i)$ is on b , then $1 \in \overline{\mathcal{V}}(\varphi, s_i)$;*

- If $(\varphi, -i)$ is on b , then $1 \notin \overline{\mathcal{V}}(\varphi, s_i)$;
- If $(\neg\varphi, +i)$ is on b , then $0 \in \overline{\mathcal{V}}(\varphi, s_i)$;
- If $(\neg\varphi, -i)$ is on b , then $0 \notin \overline{\mathcal{V}}(\varphi, s_i)$.

Proof This is also an adaptation of the proof of Priest (2008, Lemma 11a.9.6). The proof is by induction on the length of φ . The base case where $\varphi = p$ is covered by Definition 2.8. I.H.: the lemma holds for all φ' with length smaller than n , where n is the length of φ . The step will be shown by cases. Consider a formula φ of length n . The case where $\varphi = \neg p$ for some atom p is also covered by Definition 2.8.

First, $\varphi = (\psi \wedge \chi)$. The cases when $(\psi \wedge \chi, \pm i)$ are on b were already covered by Priest (2008, Lemma 11a.9.6).

$(\neg(\psi \wedge \chi), +i)$ is on b : We need to show $0 \in \overline{\mathcal{V}}(\psi \wedge \chi, s_i)$. But since b is complete, rule R3 has been applied to the node, which means that $(\neg\psi, +i)$ or $(\neg\chi, +i)$ is on b . But then (by the I.H.) $0 \in \overline{\mathcal{V}}(\psi, s_i)$ or $0 \in \overline{\mathcal{V}}(\chi, s_i)$, which means that $0 \in \overline{\mathcal{V}}(\psi \wedge \chi, s_i)$ by our semantics.

$(\neg(\psi \wedge \chi), -i)$ is on b : We need to show $0 \notin \overline{\mathcal{V}}(\psi \wedge \chi, s_i)$. Since b is complete, R4 has been applied, so $(\neg\psi, -i)$ and $(\neg\chi, -i)$ are on b . By I.H. we have $0 \notin \overline{\mathcal{V}}(\psi, s_i)$ and $0 \notin \overline{\mathcal{V}}(\chi, s_i)$, which implies the desired result by the semantics.

Now, $\varphi = \Box_m \psi$. The non-negated cases were shown by Priest (2008, Lemma 11a.9.6). Let us cover the remaining ones:

$(\neg\Box_m \psi, +i)$ is on b : We have to show that $0 \in \overline{\mathcal{V}}(\Box_m \psi, s_i)$. Since b is complete, R9 has been applied, so $(ir_m j)$ and $(\neg\psi, +j)$ are on b , for some j . By Definition 2.8, $s_i R_m s_j$, and by I.H. $0 \in \overline{\mathcal{V}}(\psi, s_j)$. By our semantics, $0 \in \overline{\mathcal{V}}(\Box_m \psi, s_i)$.

$(\neg\Box_m \psi, -i)$ is on b : We have to show that $0 \notin \overline{\mathcal{V}}(\Box_m \psi, s_i)$. There are two cases. If there is no j such that $(ir_m j)$ is in b , then by Definition 2.8 there is no j such that $s_i R_m s_j$. Then by semantics $M, s \not\models \neg\Box_m \psi$, so $0 \notin \overline{\mathcal{V}}(\Box_m \psi, s_i)$. Now, if there are $(ir_m j)$ nodes on b , since b is complete, R10 has been applied for all such j to obtain nodes $(\neg\psi, s_j)$. By Definition 2.8, for all such j , $s_i R_m s_j$. By I.H. this implies that for all such j , $0 \notin \overline{\mathcal{V}}(\psi, s_j)$, so there is no t such that $s_i R_m t$ and $\mathcal{M}, t \models \neg\psi$, and therefore $\mathcal{M}, s_i \not\models \neg\Box_m \psi$, which gives us the desired result.

Now the cases where $\varphi = \sim\psi$.

$(\sim\psi, +i)$ is on b : Then, since the branch is complete, $(\psi, -i)$ is on b . By the I.H., $1 \notin \overline{\mathcal{V}}(\psi, s_i)$. Then $\mathcal{M}, s_i \not\models \psi$, so $\mathcal{M}, s_i \models \sim\psi$ and finally $1 \in \overline{\mathcal{V}}(\sim\psi, s_i)$. The case when $(\sim\psi, -i)$ is on b is analogous.

$(\neg\sim\psi, +i)$ is on b : then, since b is complete, rule R13 has been applied and $(\psi, +i)$ is on b . Then by I.H. $1 \in \overline{\mathcal{V}}(\psi, s_i)$, so $\mathcal{M}, s_i \models \psi$, which implies $\mathcal{M}, s_i \models \neg\neg\psi$, and therefore $1 \in \overline{\mathcal{V}}(\neg\neg\psi, s_i)$. Also, notice that since $\mathcal{M}, s_i \models \neg(\neg\psi)$, we have that $0 \in \overline{\mathcal{V}}(\neg\psi, s_i)$, as desired. \square

Now we can prove soundness and completeness of this enhanced tableau system with respect to FVEL.

Theorem 2.12 *For any finite set of formulas $\Sigma \cup \{\varphi\}$, $\Sigma \vdash \varphi$ iff $\Sigma \models \varphi$.*

Proof This proof is an extension of the proofs in Priest (2008), for K_{FDE} .

For soundness, we adapt Priest (2008, Definition 11a.9.2, Lemma 11a.9.3 and Theorem 11a.9.4, pp. 255-256). The proof for the theorem is found in Priest (2008, Theorem 2.9.4, p. 32). Our only addition here is to the Soundness Lemma (Priest, 2008, Lemma 11a.9.3), which we proved in Lemma 2.10.

For completeness, the situation is very similar. We adapt Priest (2008, Definition 11a.9.5, Completeness Lemma 11a.9.6 and Theorem 11a.9.7, p. 256-257). Again, the only additions (besides minor adaptations of notation) concern the lemma, which we proved in Lemma 2.11. The proof of Priest (2008, Completeness Theorem 11a.9.7) is identical to that of Priest (2008, Theorem 2.9.7, p. 33). \square

2.4 Correspondence Results

Now we will take a look at standard axioms and inference rules from modal logics. Consider the following inference/rule schemes:

$$\begin{array}{ll} \text{(MP)} & \text{from } \vdash \varphi \text{ and } \vdash \varphi \rightarrow \psi \text{ infer } \vdash \psi \\ \text{(NEC)} & \text{from } \vdash \varphi \text{ infer } \vdash \Box_m \varphi \end{array}$$

Proposition 2.13 *MP does not preserve validity in FVEL.*¹⁴

Proof Counterexample: $\varphi = \neg(p \wedge \neg p \wedge \sim(p^n))$, $\psi = \sim p \vee p^t$.

¹⁴ One might suspect that $\varphi, \varphi \rightarrow \psi \vdash \psi$ could be valid here, but this is not the case, as is also noted by Priest (2008, Sections 8.6.5 and 9.2.1) w.r.t. FDE. Also, see Figure 2.4.

p	φ	$\neg\varphi \vee \psi$	ψ
n	t	t	t
f	t	t	t
t	t	t	t
b	b	b	f

□

Proposition 2.14 *The rule NEC preserves validity in FVEL.*

Proof Suppose an arbitrary φ is provable in FVEL. So we have a closed tableau with root $(\varphi, -0)$. Now, we can build a tableau for $\Box\varphi$ with the following procedure. First, relabel each number in the tableau for φ with its successor (ignoring the sign). Notice that uniformly changing the labels does not affect the validity of the formula being tested. Then, append the two lines below immediately above the root, obtaining the following closed tableau for $\Box\varphi$:

$\Box\varphi, -0$

$0r1$

$\varphi, -1$

[rest of the relabelled tableau for φ]

□

Now consider the following typical modal logical axioms, and their versions built with $\tilde{\vee}$ and $\tilde{\rightarrow}$ (agent indices removed for readability):

$$\begin{array}{llll}
(K) & \Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi) & (\tilde{K}) & \Box(\varphi \tilde{\rightarrow} \psi) \tilde{\rightarrow} (\Box\varphi \tilde{\rightarrow} \Box\psi) \\
(T) & \Box\varphi \rightarrow \varphi & (\tilde{T}) & \Box\varphi \tilde{\rightarrow} \varphi \quad \sim (\Box\varphi \wedge \sim \varphi) \\
(4) & \Box\varphi \rightarrow \Box\Box\varphi & (\tilde{4}) & \Box\varphi \tilde{\rightarrow} \Box\Box\varphi \quad \sim (\Box\varphi \wedge \sim \Box\Box\varphi) \\
(5) & \neg\Box\varphi \rightarrow \Box\neg\Box\varphi & (\tilde{5}) & \neg\Box\varphi \tilde{\rightarrow} \Box\neg\Box\varphi \quad \sim (\neg\Box\varphi \wedge \sim \Box\neg\Box\varphi) \\
(B) & \varphi \rightarrow \Box\Diamond\varphi & (\tilde{B}) & \varphi \tilde{\rightarrow} \Box\Diamond\varphi \quad \sim (\varphi \wedge \sim \Box\Diamond\varphi) \\
(D) & \Box\varphi \rightarrow \Diamond\varphi & (\tilde{D}) & \Box\varphi \tilde{\rightarrow} \Diamond\varphi \quad \sim (\Box\varphi \wedge \sim \Diamond\varphi)
\end{array}$$

Axiom K is not a theorem of FVEL, but \tilde{K} is. None of the other axioms above are theorems; this is expected, for recall that we are dealing with arbitrary accessibility relations. Whether these or any other formulas are theorems can be easily checked using the tableau method.

Proposition 2.15 $\mathcal{F} \models \tilde{K}$, for all frames $\mathcal{F} = (S, R)$.

Proof The only part needing explanation regards the semantics of $\tilde{\rightarrow}$: for any formulas φ and ψ , $\mathcal{M}, s \models \varphi \tilde{\rightarrow} \psi$ iff $\mathcal{M}, s \models \sim \varphi \tilde{\vee} \psi$ iff $\mathcal{M}, s \models \sim(\sim \varphi \wedge \sim \psi)$ iff $\mathcal{M}, s \models \sim(\varphi \wedge \sim \psi)$ iff $\mathcal{M}, s \not\models \varphi \wedge \sim \psi$ iff $\mathcal{M}, s \not\models \varphi$ or $\mathcal{M}, s \not\models \sim \psi$ iff $\mathcal{M}, s \not\models \varphi$ or $\mathcal{M}, s \models \psi$ iff $\mathcal{M}, s \models \varphi$ implies $\mathcal{M}, s \models \psi$.

The rest amounts to proving that $\mathcal{M}, s \models \Box(\varphi \tilde{\rightarrow} \psi) \tilde{\rightarrow} (\Box \varphi \tilde{\rightarrow} \Box \psi)$, for arbitrary $\mathcal{M} = (S, R, \mathcal{V})$ and $s \in S$. That formula is satisfied in a state s of a model \mathcal{M} iff at least one of the following holds: (a) there is a t such that sRt and $\mathcal{M}, t \models \varphi$ and $\mathcal{M}, t \not\models \psi$, (b) there is a t such that sRt and $\mathcal{M}, t \not\models \varphi$, or (c) for all t such that sRt , it holds that $\mathcal{M}, t \models \psi$. If condition (c) holds, we are done, so let us assume that (c) does not hold. This implies that there is a state t such that $\mathcal{M}, t \not\models \psi$. If $\mathcal{M}, t \models \varphi$, then condition (a) holds, otherwise, condition (b) holds. Either way, \tilde{K} is satisfied at s . \square

Not surprisingly, the correspondences between some properties of frames and validity of formulas still hold, as shown by the propositions below. For the next proofs, let $\mathcal{F} = (S, R)$, and consider the following frame properties:

ρ	(reflexivity)	For all $s \in S, sRs$.
τ	(transitivity)	For all $s, t, r \in S, sRt$ and tRr implies sRr .
σ	(symmetry)	For all $s, t \in S, sRt$ implies tRs .
η	(seriality)	For all $s \in S$, there is a $t \in S$ such that sRt .
ϵ	(Euclideaness)	For all $s, t, r \in S, sRt$ and sRr implies tRr .

Proposition 2.16 $\mathcal{F} \models \tilde{T}$ iff \mathcal{F} is reflexive.

Proof \tilde{T} is given by $\sim(\Box \varphi \wedge \sim \varphi)$. By the semantics, \tilde{T} is satisfied in a state s iff at least one of the following holds: (a) there is a t such that sRt and $1 \notin \mathcal{V}(\varphi, t)$, or (b) $1 \in \mathcal{V}(\varphi, s)$.

\Leftarrow : Consider a reflexive \mathcal{F} . For all valuations, for each state s either $1 \in \mathcal{V}(\varphi, s)$ or $1 \notin \mathcal{V}(\varphi, s)$. In the first case, the axiom is satisfied by condition (b). In the second case, since sRs and $1 \notin \mathcal{V}(\varphi, s)$, condition (a) holds.

\Rightarrow : Let us suppose that the frame \mathcal{F} is not reflexive. Consider a state s for which sRs does not hold. We need to show that there is a valuation for which \tilde{T} does not hold in s . If we take a valuation \mathcal{V} where $1 \notin \mathcal{V}(p, s)$, but $1 \in \mathcal{V}(p, t)$ for all t such that sRt , we will have $\sim p \wedge \Box p$ at s . \square

Proposition 2.17 $\mathcal{F} \models \tilde{4}$ iff \mathcal{F} is transitive.

Proof Axiom $\tilde{4}$ is given by $\sim(\Box\varphi \wedge \sim\Box\Box\varphi)$. $\tilde{4}$ is satisfied in a state s iff this state satisfies at least one of the following conditions: (a) there is a t such that sRt and $1 \notin \overline{\mathcal{V}}(\varphi, t)$, or (b) for all t and r such that sRt and tRr it is the case that $1 \in \overline{\mathcal{V}}(\varphi, r)$.

\Leftarrow : Consider a transitive frame. If condition (b) does not hold for some formula φ , then there are states s, t, r such that sRt and tRr and $1 \notin \overline{\mathcal{V}}(\varphi, r)$. But since the frame is transitive, we have sRr and thus condition (a) is satisfied.

\Rightarrow : Consider a non-transitive frame. Hence, there are states s, t, r such that sRt and tRr but not sRr . Consider a valuation \mathcal{V} where $1 \in \mathcal{V}(p, x)$ for all x such that sRx , and $1 \notin \mathcal{V}(p, r)$. In that case, $(S, R, \mathcal{V}), s \models \Box p \wedge \sim\Box\Box p$. This concludes the proof. \square

Proposition 2.18 $\mathcal{F} \models \tilde{B}$ iff \mathcal{F} is symmetric.

Proof Axiom \tilde{B} ($\sim(\varphi \wedge \sim\Box\Diamond\varphi)$) is satisfied in a state s iff: (a) $1 \notin \overline{\mathcal{V}}(\varphi, s)$ or (b) for all t there is an r such that sRt , tRr and $1 \in \overline{\mathcal{V}}(\varphi, r)$.

\Leftarrow : Either $1 \notin \overline{\mathcal{V}}(\varphi, s)$ or $1 \in \overline{\mathcal{V}}(\varphi, s)$. In the first case, (a) is satisfied. In the second case, if \mathcal{F} is symmetric, then for any t such that sRt it is also the case that tRs , therefore condition (b) is satisfied.

\Rightarrow : Suppose \mathcal{F} is not symmetric, that is, there are states s, t such that sRt but not tRs . Consider the instance $\sim(p \wedge \sim\Box\Diamond p)$ and a valuation where $1 \in \mathcal{V}(p, s)$ – which violates condition (a) – and $1 \notin \mathcal{V}(p, r)$ for all $r \neq s$. Since $(t, s) \notin R$, there is no state r such that tRr and $1 \in \mathcal{V}(p, r)$, and therefore (b) is violated in s . \square

Proposition 2.19 $\mathcal{F} \models \tilde{D}$ iff \mathcal{F} is serial.

Proof Axiom \tilde{D} ($\sim(\Box\varphi \wedge \sim\Diamond\varphi)$) is satisfied in a state s iff: (a) there is a t such that sRt and $1 \notin \overline{\mathcal{V}}(\varphi, t)$, or (b) there is a t such that sRt and $1 \in \overline{\mathcal{V}}(\varphi, t)$. Condition (a) or (b) is satisfied iff: (c) there is a t such that sRt .

\Leftarrow : Suppose \mathcal{F} is serial. Then for any state s there is a t such that sRt , and therefore condition (c) is satisfied in s .

\Rightarrow : Suppose \mathcal{F} is not serial. Then there is a state s such that there is no t with sRt . So s violates condition (c), and therefore s does not satisfy \tilde{D} . \square

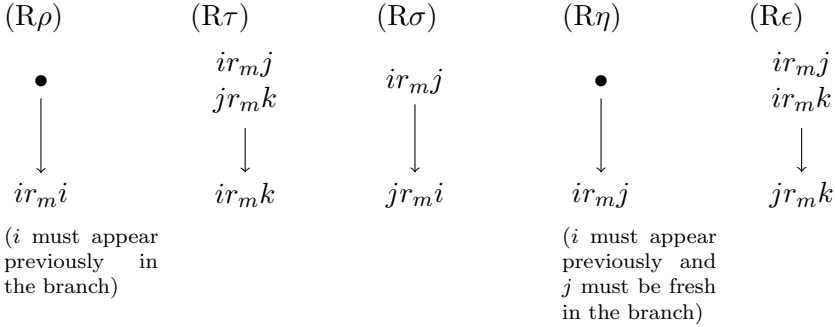
Proposition 2.20 $\mathcal{F} \models \tilde{5}$ iff \mathcal{F} is Euclidean

Proof Axiom $\tilde{5}$ ($\sim(\neg\Box\varphi \wedge \sim\Box\neg\Box\varphi)$) is satisfied in a state s iff: (a) for all t such that sRt it is the case that $0 \notin \overline{\mathcal{V}}(\varphi, t)$, or (b) for all t such that sRt there is an r with tRr and $0 \in \overline{\mathcal{V}}(\varphi, r)$.

\Leftarrow : Suppose \mathcal{F} is Euclidean. For any state s , condition (a) is either satisfied or not. If it is, $\tilde{5}$ is satisfied. Now suppose (a) is violated at s , that is, there is a state r such that sRr and $0 \in \overline{\mathcal{V}}(\varphi, r)$. Since \mathcal{F} is Euclidean, sRt implies tRr for any t , and then condition (b) is satisfied for s . Therefore $\tilde{5}$ is satisfied in either case.

\Rightarrow : Suppose \mathcal{F} is not Euclidean, that is, there are states s, t, r such that sRt , sRr but tRr does not hold. Let \mathcal{V} be the valuation such that $0 \in \mathcal{V}(p, r)$ – which violates condition (a) for formula $\varphi = p$ – but $0 \notin \mathcal{V}(p, w)$ for all $w \neq r$, which violates condition (b) at s (because sRt but there is no z such that tRz and $0 \in \mathcal{V}(p, z)$). Therefore, $(S, R, \mathcal{V}), s \not\models \tilde{5}$. \square

Now, we can augment the tableau system with any combination of the rules below and show that it is complete with respect to the corresponding class of models.¹⁵



Let $\star \subseteq \{\rho, \tau, \sigma, \eta, \epsilon\}$. We use the symbol \vdash_\star for the provability relation of the tableau system augmented with rules R \circ , for each $\circ \in \star$, and \models_\star to represent satisfiability restricted only to models satisfying properties in \star .

Theorem 2.21 For all finite sets of formulas $\Sigma \cup \{\varphi\}$, and $\star \subseteq \{\rho, \tau, \sigma, \eta, \epsilon\}$, the following statement holds: $\Sigma \vdash_\star \varphi$ iff $\Sigma \models_\star \varphi$.

¹⁵ The notation used for these rules follows the frame property names defined previously, which in turn are taken from Priest (2008, Section 3.2.3).

Proof This proof is similar to the proof of Theorem 2.12, the only difference is that now we consider particular classes of models, and augment the tableau system with its corresponding rule(s).

We will again build upon Priest (2008)’s proofs. Again, the main modifications are in the soundness and completeness lemmas (Lemmas 11a.9.3 and 11a.9.6 of Priest (2008), respectively). The soundness and completeness theorems remain unchanged (Theorems 11a.9.4 and 11a.9.7 of Priest (2008), respectively, whose actual proofs are found in Theorems 1.11.3 and 2.9.7, respectively).

Reflexive models: Now we will prove soundness and completeness of the tableau system augmented with the rule $R\rho$ w.r.t. reflexive models. Let us first analyse soundness. In this case, the model \mathcal{M} mentioned in the soundness lemma (Lemma 2.10) should be restricted to be a reflexive model. We only have to check the new rule $R\rho$, because for all the other rules it was already shown that there will be at least one faithful extension (our restriction of \mathcal{M} to reflexive models is still covered by the lemma, which says “any model”). Suppose b is faithful to \mathcal{M} , which is reflexive, and that b' is generated from b by the application of $R\rho$. If the added node is (ir_mi) , then i has occurred in b , but since \mathcal{M} is reflexive, $f(i)R_m f(i)$ is in \mathcal{M} .

For completeness, since the new rule $R\rho$ does not involve any formula, the only thing we need to show is that the induced model will always be reflexive. Suppose the label i occurs on the branch. Then, since the branch is complete, at some point the rule $R\rho$ should be applied, generating the node ir_mi . By the definition of induced model, we conclude it is indeed reflexive. This finishes the proof of the first statement.

Transitive models: The second statement concerns transitive models and the tableau with $R\tau$. For soundness, the lemma should be rephrased again to consider only transitive models. Now we have to check whether the application of the rule $R\tau$ to b will produce a faithful extension b' . Suppose b is faithful to \mathcal{M} , and contains nodes (ir_mj) and (jr_mk) . By applying $R\tau$ we get (ir_mk) . But since b is faithful to \mathcal{M} , $f(i)R_m f(j)$ and $f(j)R_m f(k)$ are in \mathcal{M} , and since \mathcal{M} is transitive, $f(i)R_m f(k)$ is also in \mathcal{M} , and therefore b' is faithful to \mathcal{M} .

For completeness we need to ensure that the induced model is transitive. Suppose the nodes (ir_mj) and (jr_mk) occur on the complete branch. Then at some point the rule $R\tau$ had to be applied, with (ir_mk) as outcome. By the definition of induced model, we conclude it is transitive.

Symmetric and Euclidean models: The proofs for symmetric and Eu-

clidean models are analogous to the previous ones.

Serial models: Soundness: we need to check if the application of $R\eta$ to b will generate at least one faithful branch b' . Suppose that label i occurs in b and that we apply $R\eta$, generating only one new node: (ir_mj) . Since b is faithful to \mathcal{M} , which is serial, $f(i)R_m f(j)$ is in \mathcal{M} , for some f , and thus b' is faithful to \mathcal{M} .

Completeness: We will show that the induced model \mathcal{M} is serial. Suppose i occurs on the complete open branch. Then at some point $R\eta$ must be applied and thus (ir_mj) will be on the branch as well. Now the same happens with the new label j , and so on, *ad infinitum*. So this infinitely long branch will contain, for all labels i that occur on it, some node (ir_mj) , and therefore, by the definition of induced models, \mathcal{M} is serial. (Notice that the issue of infinitely long branches does not prevent completeness – it might affect decidability, but we are not concerned with it in this proof.) \square

Decidability of tableau provability for these systems can be shown using the proofs for standard modal logics (like those in Halpern and Moses (1992, Section 6.3) and Fitting (1983, Chapter 8, Section 7)), and making minor adaptations. The four-valuedness of FVEL does not change anything with respect to decidability, since what may cause infinite branches is always the modal part, in particular when transitivity is involved. This problem is usually solved in the literature by detecting and preventing creation of new labels (worlds) if the formulas associated with them are identical to the ones associated with some previous label in the branch.

2.5 Public Announcements

In this section, we extend the language with public announcements, provide a set of reduction validities¹⁶ and prove completeness for this extended language. The first time an axiomatisation was given to a four-valued modal logic with public announcements can be credited to Riviaccio (2014b,a), with Bilattice Public Announcement Logic (BPAL). The reduction axioms for BPAL are all valid in FVEL if our language is extended with the missing connectives (more on the comparison between FVEL and BPAL in Section 2.6.2).

¹⁶ The term “reduction axioms” is commonly used in the literature, but here we are working with validities, not axioms.

The semantics for the new operator are defined as shown below. Differently from Plaza (1989, 2007); van Ditmarsch, van der Hoek, and Kooi (2007), we define a separate clause for the negated announcement, in line with the rest of our semantics:

$$\begin{array}{ll} \mathcal{M}, s \models [\varphi]\psi & \text{iff } \mathcal{M}, s \models \varphi \text{ implies } \mathcal{M}|_{\varphi}, s \models \psi \\ \mathcal{M}, s \models \neg[\varphi]\psi & \text{iff } \mathcal{M}, s \models \varphi \text{ and } \mathcal{M}|_{\varphi}, s \models \neg\psi \end{array}$$

where $\mathcal{M} = (S, R, \mathcal{V})$ and $\mathcal{M}|_{\varphi} = (S', R', \mathcal{V}')$, with $S' = \{s \in S \mid \mathcal{M}, s \models \varphi\}$, $R' = R \cap (S' \times S')$ and $\mathcal{V}' = \mathcal{V}|_{At \times S'}$.

The model of Figure 2.1 (right), upon the public announcement of $\neg p$, would be transformed according to Figure 2.6.

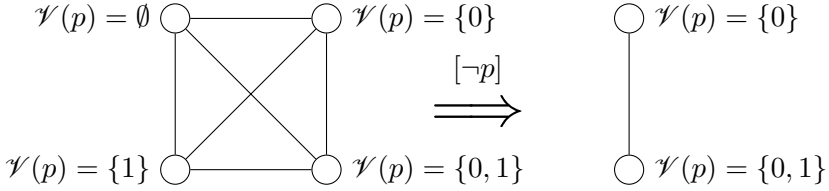


Figure 2.6: The announcement of $\neg p$.

Notice that, for propositional atoms, the announcement of p does not delete worlds where $\neg p$ holds, but only worlds where p does not hold, that is, worlds where $\sim p$ holds. To delete worlds where $\neg p$ holds we would have to announce $\sim \neg p$, so that only worlds s with $\mathcal{M}, s \models \sim \neg p$ (which is equivalent to $\mathcal{M}, s \not\models \neg p$) would survive.

As explained earlier, public announcements in FVEL do not change the evidence itself (that would require a valuation-changing operation), just what agents know about it. This is not to say that only the accessibility relations are altered: what actually happens is that evidential states not conforming to the announcement, which is a truthful description of the actual evidential situation, are removed. So, for example, if the formula φ^b is announced, any state where φ does not have value *both* (that is, where there is not evidence both for and against φ) will be removed. Note also that this kind of announcement which specifies one out of the four truth values would not be possible in a logic such as K_{FDE} , which lacks \sim .

Example 2.22 Consider again Example 2.7. Now suppose the actual world is s_2 , and so p (coffee is beneficial for health) is true, i.e., there

is only positive evidence for p (and Kate knows that). Suppose also that Kate announces that a paper was published in a very respectable journal reassessing all the main studies that concluded that coffee was not beneficial for health, and that the new paper concluded that those studies were not reliable due to sloppy methodology. Now this is equivalent to an announcement of $\Box_k \sim \neg p$ (Kate knows that there is no evidence for the falsity of p). This announcement results in the removal of worlds where evidence for the falsity of p is present, namely s_1 and s_3 . The resulting model is the one in Figure 2.7, where John knows the status of p too. The formula $\sim \Box_j(p \wedge \sim \neg p) \wedge [\Box_k \sim \neg p] \Box_j(p \wedge \sim \neg p)$, which is satisfied in s_2 before the announcement, reflects the fact that John does not know the status of p , but after Kate's announcement he learns that p is true.

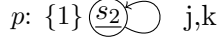


Figure 2.7: No false evidence for p .

These examples show the dynamics of the agents' knowledge about available information/evidence. It might be puzzling, however, to notice that these models actually do not say much about factual knowledge. Nevertheless, it is based on information and evidence that one can form knowledge and beliefs. This observation calls for an extension of FVEL in which knowledge about evidence could be converted into factual knowledge or belief. This endeavor is left for the next chapters.

2.5.1 Reduction Validities

As is the case for Public Announcement Logic (Plaza, 1989, 2007; Gerbrandy and Groeneveld, 1997), public announcements in FVEL do not increase expressivity. Any formula with public announcements can be rewritten as a standard FVEL formula, through the use of the following reduction validities.

$$\begin{array}{ll}
 [\varphi]p \leftrightarrow \sim \varphi \vee p & (\text{AnAt}) \\
 [\varphi]\neg p \leftrightarrow \sim \varphi \vee \neg p & (\text{An}\neg) \\
 [\varphi](\psi \wedge \chi) \leftrightarrow [\varphi]\psi \wedge [\varphi]\chi & (\text{An}\wedge) \\
 [\varphi]\neg(\psi \wedge \chi) \leftrightarrow [\varphi]\neg\psi \vee [\varphi]\neg\chi & (\text{An}\neg\wedge)
 \end{array}$$

$$\begin{array}{ll}
[\varphi]\Box_m\psi \leftrightarrow \sim\varphi \vee \Box_m[\varphi]\psi & (\text{An}\Box) \\
[\varphi]\neg\Box_m\psi \leftrightarrow \sim\varphi \vee \neg\Box_m[\varphi]\psi & (\text{An}\neg\Box) \\
[\varphi]\sim\psi \leftrightarrow \sim\varphi \vee \sim[\varphi]\psi & (\text{An}\sim) \\
[\varphi]\neg\sim\psi \leftrightarrow \sim\varphi \vee \sim\sim[\varphi]\psi & (\text{An}\neg\sim) \\
[\varphi][\psi]\chi \leftrightarrow [\varphi \wedge [\varphi]\psi]\chi & (\text{AnAn}) \\
[\varphi]\neg[\psi]\chi \leftrightarrow [\varphi](\sim\sim\psi \wedge [\psi]\neg\chi) & (\text{An}\neg\text{An})
\end{array}$$

Proposition 2.23 *All formulas above for public announcements in FVEL are valid.*

Proof For each validity $\varphi \leftrightarrow \psi$, we will prove the equivalent $\varphi \equiv \psi$ instead of $\models \varphi \leftrightarrow \psi$ (relying on Proposition 2.6). To verify that a validity of the form $\varphi \equiv \psi$ is correct, one just has to check that the semantic conditions for $\mathcal{M}, s \models \varphi$ are equivalent to the conditions for $\mathcal{M}, s \models \psi$, and do the same for $\mathcal{M}, s \models \neg\varphi$ and $\mathcal{M}, s \models \neg\psi$ (with arbitrary \mathcal{M} and s).

(AnAt): $[\varphi]p \equiv \sim\varphi \vee p$:

We will check the truth conditions for each side of the validity and make sure they are equivalent.

$$\begin{array}{l}
\mathcal{M}, s \models [\varphi]p \\
\text{iff } \mathcal{M}, s \not\models \varphi \text{ or } \mathcal{M}|_\varphi, s \models p \\
\text{iff } \mathcal{M}, s \models \sim\varphi \text{ or } \mathcal{M}|_\varphi, s \models p
\end{array}$$

But since $\mathcal{M}|_\varphi = (S', R', \mathcal{V}')$ with $\mathcal{V}'(p, s) = \mathcal{V}(p, s)$ for all atoms p and $s \in S'$, we have $\mathcal{M}, s \models p$ iff $\mathcal{M}|_\varphi, s \models p$, and therefore $\mathcal{M}, s \models [\varphi]p$ iff $\mathcal{M}, s \models \sim\varphi \vee p$. Now we have to check the truth conditions for the negations of each side of the validity, and see if they are equivalent. This will guarantee that the formulas on each side of the equivalence always take on the same truth value.

$$\begin{array}{l}
\mathcal{M}, s \models \neg[\varphi]p \\
\text{iff } \mathcal{M}, s \models \varphi \text{ and } \mathcal{M}|_\varphi, s \models \neg p
\end{array}$$

And:

$$\begin{array}{l}
\mathcal{M}, s \models \neg(\sim\varphi \vee p) \\
\text{iff [by the definition of } \vee] \\
\mathcal{M}, s \models \neg\neg(\neg\sim\varphi \wedge \neg p) \\
\text{iff } \mathcal{M}, s \models \neg\sim\varphi \wedge \neg p \\
\text{iff } \mathcal{M}, s \models \varphi \wedge \neg p
\end{array}$$

iff $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, s \models \neg p$

It turns out that those conditions are exactly the same, for (again) $\mathcal{M} = (S, R, \mathcal{V})$ and $\mathcal{M}|_\varphi = (S', R', \mathcal{V}')$ with $\mathcal{V}'(p, s) = \mathcal{V}(p, s)$ for all atoms p and $s \in S'$. Therefore validity (AnAt) holds.

(An \neg): $[\varphi]\neg p \equiv \sim \varphi \vee \neg p$:

$\mathcal{M}, s \models [\varphi]\neg p$
iff $\mathcal{M}, s \not\models \varphi$ or $\mathcal{M}|_\varphi, s \models \neg p$
iff $\mathcal{M}, s \not\models \varphi$ or $0 \in \mathcal{V}(p, s)$
iff $\mathcal{M}, s \not\models \varphi$ or $\mathcal{M}, s \models \neg p$
iff $\mathcal{M}, s \models \sim \varphi \vee \neg p$

And for the negated formulas:

$\mathcal{M}, s \models \neg[\varphi]\neg p$
iff $\mathcal{M}, s \models \varphi$ and $\mathcal{M}|_\varphi, s \models \neg\neg p$
iff $\mathcal{M}, s \models \varphi$ and $\mathcal{M}|_\varphi, s \models p$

And:

$\mathcal{M}, s \models \neg(\sim \varphi \vee \neg p)$
iff $\mathcal{M}, s \models \neg\neg(\neg \sim \varphi \wedge \neg\neg p)$
iff $\mathcal{M}, s \models \varphi \wedge p$
iff $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, s \models p$

Since $\mathcal{M}|_\varphi$'s valuation is only a restriction of \mathcal{M} 's valuation, both truth conditions are the same.

(An \wedge): $[\varphi](\psi \wedge \chi) \equiv [\varphi]\psi \wedge [\varphi]\chi$:

The proof for this and the following validities will follow the same structure as the previous one. We will just show the truth conditions.

$\mathcal{M}, s \models [\varphi](\psi \wedge \chi)$
iff $\mathcal{M}, s \not\models \varphi$ or $\mathcal{M}|_\varphi, s \models \psi \wedge \chi$
iff $\mathcal{M}, s \not\models \varphi$ or ($\mathcal{M}|_\varphi, s \models \psi$ and $\mathcal{M}|_\varphi, s \models \chi$)

And:

$\mathcal{M}, s \models [\varphi]\psi \wedge [\varphi]\chi$
iff $\mathcal{M}, s \models [\varphi]\psi$ and $\mathcal{M}, s \models [\varphi]\chi$
iff ($\mathcal{M}, s \not\models \varphi$ or $\mathcal{M}|_\varphi, s \models \psi$) and ($\mathcal{M}, s \not\models \varphi$ or $\mathcal{M}|_\varphi, s \models \chi$)
iff $\mathcal{M}, s \not\models \varphi$ or ($\mathcal{M}|_\varphi, s \models \psi$ and $\mathcal{M}|_\varphi, s \models \chi$)

Now for the conditions for the negated formulas.

$\mathcal{M}, s \models \neg[\varphi](\psi \wedge \chi)$

$\text{iff } \mathcal{M}, s \models \varphi \text{ and } \mathcal{M}|_{\varphi}, s \models \neg(\psi \wedge \chi)$
 $\text{iff } \mathcal{M}, s \models \varphi \text{ and } (\mathcal{M}|_{\varphi}, s \models \neg\psi \text{ or } \mathcal{M}|_{\varphi}, s \models \neg\chi)$

And:

$\mathcal{M}, s \models \neg([\varphi]\psi \wedge [\varphi]\chi)$
 $\text{iff } \mathcal{M}, s \models \neg[\varphi]\psi \text{ or } \mathcal{M}, s \models \neg[\varphi]\chi$
 $\text{iff } (\mathcal{M}, s \models \varphi \text{ and } \mathcal{M}|_{\varphi}, s \models \neg\psi) \text{ or } (\mathcal{M}, s \models \varphi \text{ and } \mathcal{M}|_{\varphi}, s \models \neg\chi)$
 $\text{iff } \mathcal{M}, s \models \varphi \text{ and } (\mathcal{M}|_{\varphi}, s \models \neg\psi \text{ or } \mathcal{M}|_{\varphi}, s \models \neg\chi)$

(An \neg): $[\varphi]\neg(\psi \wedge \chi) \equiv [\varphi]\neg\psi \vee [\varphi]\neg\chi$:

$\mathcal{M}, s \models [\varphi]\neg(\psi \wedge \chi)$
 $\text{iff } \mathcal{M}, s \not\models \varphi \text{ or } \mathcal{M}|_{\varphi}, s \models \neg(\psi \wedge \chi)$
 $\text{iff } \mathcal{M}, s \not\models \varphi \text{ or } \mathcal{M}|_{\varphi}, s \models \neg\psi \text{ or } \mathcal{M}|_{\varphi}, s \models \neg\chi$
 $\text{iff } \mathcal{M}, s \models [\varphi]\neg\psi \text{ or } \mathcal{M}, s \models [\varphi]\neg\chi$
 $\text{iff } \mathcal{M}, s \models [\varphi]\neg\psi \vee [\varphi]\neg\chi$

For the negated formulas:

$\mathcal{M}, s \models \neg[\varphi]\neg(\psi \wedge \chi)$
 $\text{iff } \mathcal{M}, s \models \varphi \text{ and } \mathcal{M}|_{\varphi}, s \models \psi \wedge \chi$

And:

$\mathcal{M}, s \models \neg([\varphi]\neg\psi \vee [\varphi]\neg\chi)$
 $\text{iff } \mathcal{M}, s \models \neg[\varphi]\neg\psi \text{ and } \mathcal{M}, s \models \neg[\varphi]\neg\chi$
 $\text{iff } \mathcal{M}, s \models \varphi \text{ and } \mathcal{M}|_{\varphi}, s \models \psi \text{ and } \mathcal{M}|_{\varphi}, s \models \chi$
 $\text{iff } \mathcal{M}, s \models \varphi \text{ and } \mathcal{M}|_{\varphi}, s \models \psi \wedge \chi$

(An \Box): $[\varphi]\Box\psi \equiv \sim\varphi \vee \Box[\varphi]\psi$:

$\mathcal{M}, s \models [\varphi]\Box\psi$
 $\text{iff } \mathcal{M}, s \not\models \varphi \text{ or } \mathcal{M}|_{\varphi}, s \models \Box\psi$
 $\text{iff } \mathcal{M}, s \not\models \varphi \text{ or for all } t \in S' \text{ s.t. } sR't(\mathcal{M}|_{\varphi}, t \models \psi)$

And:

$\mathcal{M}, s \models \sim\varphi \vee \Box[\varphi]\psi$
 $\text{iff } \mathcal{M}, s \not\models \varphi \text{ or for all } t \in S \text{ s.t. } sRt(\mathcal{M}, t \models [\varphi]\psi)$
 $\text{iff } \mathcal{M}, s \not\models \varphi \text{ or for all } t \in S \text{ s.t. } sRt(\mathcal{M}, t \not\models \varphi \text{ or } \mathcal{M}|_{\varphi}, t \models \psi)$
 $\text{iff [by classical logic]}$
 $\mathcal{M}, s \not\models \varphi \text{ or } (\mathcal{M}, s \models \varphi \text{ and for all } t \in S \text{ s.t. } sRt(\mathcal{M}, t \not\models \varphi \text{ or } (\mathcal{M}, t \models \varphi \text{ and } \mathcal{M}|_{\varphi}, t \models \psi)))$
 $\text{iff [by set theory: } S = (S \setminus S') \cup S']$
 $\mathcal{M}, s \not\models \varphi \text{ or } (\mathcal{M}, s \models \varphi \text{ and }$

for all $t \in S \setminus S'$ s.t. $sRt(\mathcal{M}, t \not\models \varphi$ or $(\mathcal{M}, t \models \varphi$ and $\mathcal{M}|_\varphi, t \models \psi)$)
and

for all $t \in S'$ s.t. $sRt(\mathcal{M}, t \not\models \varphi$ or $(\mathcal{M}, t \models \varphi$ and $\mathcal{M}|_\varphi, t \models \psi)$)
iff [by: $\mathcal{M}, s \models \varphi$ for all $s \in S'$, and $\mathcal{M}, s \not\models \varphi$ for all $s \in S \setminus S'$]
 $\mathcal{M}, s \not\models \varphi$ or $(\mathcal{M}, s \models \varphi$ and
for all $t \in S \setminus S'$ s.t. sRt ($true$ or $(\mathcal{M}, t \models \varphi$ and $\mathcal{M}|_\varphi, t \models \psi)$) and
for all $t \in S'$ s.t. sRt ($false$ or $(true$ and $\mathcal{M}|_\varphi, t \models \psi)$))
iff [by classical logic]
 $\mathcal{M}, s \not\models \varphi$ or $(\mathcal{M}, s \models \varphi$ and for all $t \in S'$ s.t. $sRt(\mathcal{M}|_\varphi, t \models \psi)$)
iff [by: $s \in S$ and $\mathcal{M}, s \models \varphi$ iff $s \in S'$; sRt and $s, t \in S'$ iff $sR't$]
 $\mathcal{M}, s \not\models \varphi$ or $(\mathcal{M}, s \models \varphi$ and for all $t \in S'$ s.t. $sR't(\mathcal{M}|_\varphi, t \models \psi)$)
iff [by classical logic]
 $\mathcal{M}, s \not\models \varphi$ or for all $t \in S'$ s.t. $sR't(\mathcal{M}|_\varphi, t \models \psi)$

For the negated formulas:

$\mathcal{M}, s \models \neg[\varphi]\Box\psi$
iff $\mathcal{M}, s \models \varphi$ and $\mathcal{M}|_\varphi, s \models \neg\Box\psi$
iff $\mathcal{M}, s \models \varphi$ and there is a $t \in S'$ ($sR't$ and $\mathcal{M}|_\varphi, t \models \neg\psi$)

And:

$\mathcal{M}, s \models \neg(\sim\varphi \vee \Box[\varphi]\psi)$
iff $\mathcal{M}, s \models \neg\neg(\neg\sim\varphi \wedge \neg\Box[\varphi]\psi)$
iff $\mathcal{M}, s \models \varphi \wedge \neg\Box[\varphi]\psi$
iff $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, s \models \neg\Box[\varphi]\psi$
iff $\mathcal{M}, s \models \varphi$ and there is a $t \in S$ (sRt and $\mathcal{M}, t \models \neg[\varphi]\psi$)
iff $\mathcal{M}, s \models \varphi$ and there is a $t \in S$ (sRt and $(\mathcal{M}, t \models \varphi$ and $\mathcal{M}|_\varphi, t \models \neg\psi)$)

But, since $\mathcal{M}, t \models \varphi$ in the latest condition, $t \in S'$.

(An- \Box): $[\varphi]\neg\Box\psi \equiv \sim\varphi \vee \neg\Box[\varphi]\psi$:

$\mathcal{M}, s \models [\varphi]\neg\Box\psi$
iff $\mathcal{M}, s \not\models \varphi$ or $\mathcal{M}|_\varphi, s \models \neg\Box\psi$
iff $\mathcal{M}, s \not\models \varphi$ or there is a $t \in S'$ ($sR't$ and $\mathcal{M}|_\varphi, t \models \neg\psi$)

And:

$\mathcal{M}, s \models \sim\varphi \vee \neg\Box[\varphi]\psi$
iff $\mathcal{M}, s \not\models \varphi$ or $\mathcal{M}, s \models \neg\Box[\varphi]\psi$
iff $\mathcal{M}, s \not\models \varphi$ or there is a $t \in S$ (sRt and $\mathcal{M}, t \models \neg[\varphi]\psi$)
iff $\mathcal{M}, s \not\models \varphi$ or there is a $t \in S$ (sRt and $\mathcal{M}, t \models \varphi$ and $\mathcal{M}|_\varphi, t \models \neg\psi$)

Notice that the final conditions for the satisfaction of both formulas are equivalent, for the states $t \in S$ such that $\mathcal{M}, t \models \varphi$ are exactly the states $t \in S'$; additionally, notice that in the existential condition we can consider that $\mathcal{M}, s \models \varphi$. Now for the negated formulas:

$$\begin{aligned} \mathcal{M}, s &\models \neg[\varphi]\neg\Box\psi \\ \text{iff } \mathcal{M}, s &\models \varphi \text{ and } \mathcal{M}|_{\varphi}, s \models \neg\neg\Box\psi \\ \text{iff } \mathcal{M}, s &\models \varphi \text{ and } \mathcal{M}|_{\varphi}, s \models \Box\psi \\ \text{iff } \mathcal{M}, s &\models \varphi \text{ and for all } t \in S' \text{ s.t. } sR't \text{ (} \mathcal{M}|_{\varphi}, t \models \psi \text{)} \end{aligned}$$

And:

$$\begin{aligned} \mathcal{M}, s &\models \neg(\sim\varphi \vee \neg\Box[\varphi]\psi) \\ \text{iff } \mathcal{M}, s &\models \neg\sim\varphi \text{ and } \mathcal{M}, s \models \neg\neg\Box[\varphi]\psi \\ \text{iff } \mathcal{M}, s &\models \varphi \text{ and } \mathcal{M}, s \models \Box[\varphi]\psi \\ \text{iff } \mathcal{M}, s &\models \varphi \text{ and for all } t \in S \text{ s.t. } sRt \text{ (} \mathcal{M}, t \models [\varphi]\psi \text{)} \\ \text{iff } \mathcal{M}, s &\models \varphi \text{ and for all } t \in S \text{ s.t. } sRt \text{ (} \mathcal{M}, t \not\models \varphi \text{ or } \mathcal{M}|_{\varphi}, t \models \psi \text{)} \end{aligned}$$

(An~): $[\varphi]\sim\psi \equiv \sim\varphi \vee \sim[\varphi]\psi$:

$$\begin{aligned} \mathcal{M}, s &\models [\varphi]\sim\psi \\ \text{iff } \mathcal{M}, s &\not\models \varphi \text{ or } \mathcal{M}|_{\varphi}, s \not\models \psi \end{aligned}$$

And:

$$\begin{aligned} \mathcal{M}, s &\models \sim\varphi \vee \sim[\varphi]\psi \\ \text{iff } \mathcal{M}, s &\not\models \varphi \text{ or } \mathcal{M}, s \not\models [\varphi]\psi \\ \text{iff } \mathcal{M}, s &\not\models \varphi \text{ or not (} \mathcal{M}, s \not\models \varphi \text{ or } \mathcal{M}|_{\varphi}, s \models \psi \text{)} \\ \text{iff } \mathcal{M}, s &\not\models \varphi \text{ or (} \mathcal{M}, s \models \varphi \text{ and } \mathcal{M}|_{\varphi}, s \not\models \psi \text{)} \end{aligned}$$

For the negated formulas:

$$\begin{aligned} \mathcal{M}, s &\models \neg[\varphi]\sim\psi \\ \text{iff } \mathcal{M}, s &\models \varphi \text{ and } \mathcal{M}|_{\varphi}, s \models \neg\sim\psi \\ \text{iff } \mathcal{M}, s &\models \varphi \text{ and } \mathcal{M}|_{\varphi}, s \models \psi \end{aligned}$$

And

$$\begin{aligned} \mathcal{M}, s &\models \neg(\sim\varphi \vee \sim[\varphi]\psi) \\ \text{iff } \mathcal{M}, s &\models \neg\neg(\neg\sim\varphi \wedge \neg\sim[\varphi]\psi) \\ \text{iff } \mathcal{M}, s &\models \varphi \wedge [\varphi]\psi \\ \text{iff } \mathcal{M}, s &\models \varphi \text{ and (} \mathcal{M}, s \not\models \varphi \text{ or } \mathcal{M}|_{\varphi}, s \models \psi \text{)} \\ \text{iff } \mathcal{M}, s &\models \varphi \text{ and } \mathcal{M}|_{\varphi}, s \models \psi \end{aligned}$$

(An¬~): $[\varphi]\neg\sim\psi \equiv \sim\varphi \vee \sim\sim[\varphi]\psi$:

$$\mathcal{M}, s \models [\varphi]\neg\sim\psi$$

iff $\mathcal{M}, s \not\models \varphi$ or $\mathcal{M}|_{\varphi}, s \models \neg \sim \psi$
 iff $\mathcal{M}, s \not\models \varphi$ or $\mathcal{M}|_{\varphi}, s \models \psi$

And:

$\mathcal{M}, s \models \sim \varphi \vee \sim \sim [\varphi] \psi$
 iff $\mathcal{M}, s \models \sim \varphi$ or $\mathcal{M}, s \models [\varphi] \psi$
 iff $\mathcal{M}, s \not\models \varphi$ or $\mathcal{M}|_{\varphi}, s \models \psi$

And for the negated formulas:

$\mathcal{M}, s \models \neg [\varphi] \neg \sim \psi$
 iff $\mathcal{M}, s \models \varphi$ and $\mathcal{M}|_{\varphi}, s \models \neg \neg \sim \psi$
 iff $\mathcal{M}, s \models \varphi$ and $\mathcal{M}|_{\varphi}, s \models \sim \psi$

And:

$\mathcal{M}, s \models \neg (\sim \varphi \vee \sim \sim [\varphi] \psi)$
 iff $\mathcal{M}, s \models \neg \sim \varphi$ and $\mathcal{M}, s \models \neg \sim \sim [\varphi] \psi$
 iff $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, s \models \sim [\varphi] \psi$
 iff $\mathcal{M}, s \models \varphi$ and not $(\mathcal{M}, s \not\models \varphi \text{ or } \mathcal{M}|_{\varphi}, s \models \psi)$
 iff $\mathcal{M}, s \models \varphi$ and $(\mathcal{M}, s \models \varphi \text{ and } \mathcal{M}|_{\varphi}, s \models \sim \psi)$
 iff $\mathcal{M}, s \models \varphi$ and $\mathcal{M}|_{\varphi}, s \models \sim \psi$

(AnAn): $[\varphi][\psi]\chi \equiv [\varphi \wedge [\varphi]\psi]\chi$:

$\mathcal{M}, s \models [\varphi][\psi]\chi$
 iff $\mathcal{M}, s \not\models \varphi$ or $\mathcal{M}|_{\varphi}, s \models [\psi]\chi$
 iff $\mathcal{M}, s \not\models \varphi$ or $\mathcal{M}|_{\varphi}, s \not\models \psi$ or $(\mathcal{M}|_{\varphi})|_{\psi}, s \models \chi$

And:

$\mathcal{M}, s \models [\varphi \wedge [\varphi]\psi]\chi$
 iff $\mathcal{M}, s \not\models \varphi \wedge [\varphi]\psi$ or $\mathcal{M}|_{\varphi \wedge [\varphi]\psi}, s \models \chi$
 iff $\mathcal{M}, s \not\models \varphi$ or $\mathcal{M}, s \not\models [\varphi]\psi$ or $\mathcal{M}|_{\varphi \wedge [\varphi]\psi}, s \models \chi$
 iff $\mathcal{M}, s \not\models \varphi$ or not $(\mathcal{M}, s \not\models \varphi \text{ or } \mathcal{M}|_{\varphi}, s \models \psi)$ or $\mathcal{M}|_{\varphi \wedge [\varphi]\psi}, s \models \chi$
 iff $\mathcal{M}, s \not\models \varphi$ or $(\mathcal{M}, s \models \varphi \text{ and } \mathcal{M}|_{\varphi}, s \not\models \psi)$ or $\mathcal{M}|_{\varphi \wedge [\varphi]\psi}, s \models \chi$
 iff $\mathcal{M}, s \not\models \varphi$ or $\mathcal{M}|_{\varphi}, s \not\models \psi$ or $\mathcal{M}|_{\varphi \wedge [\varphi]\psi}, s \models \chi$

Now the question is: are $(\mathcal{M}|_{\varphi})|_{\psi}, s \models \chi$ and $\mathcal{M}|_{\varphi \wedge [\varphi]\psi}, s \models \chi$ equivalent? We will just show that the domains $S_{\varphi|\psi}$ and $S_{\varphi \wedge [\varphi]\psi}$ (of $(\mathcal{M}|_{\varphi})|_{\psi}$ and $\mathcal{M}|_{\varphi \wedge [\varphi]\psi}$, respectively) are the same, which is a sufficient condition. We have $S_{\varphi|\psi} = \{s \in S_{\varphi} \mid \mathcal{M}|_{\varphi}, s \models \psi\} = \{s \in S \mid \mathcal{M}, s \models \varphi \text{ and } \mathcal{M}|_{\varphi}, s \models \psi\}$, where S and S_{φ} are the domains from \mathcal{M} and $\mathcal{M}|_{\varphi}$, respectively. Moreover, $S_{\varphi \wedge [\varphi]\psi} = \{s \in S \mid \mathcal{M}, s \models \varphi \wedge [\varphi]\psi\}$, but $\mathcal{M}, s \models \varphi \wedge [\varphi]\psi$

iff $\mathcal{M}, s \models \varphi$ and $\mathcal{M}|_{\varphi}, s \models \psi$. Therefore, $S_{\varphi \wedge [\varphi]\psi} = \{s \in S \mid \mathcal{M}, s \models \varphi \text{ and } \mathcal{M}|_{\varphi}, s \models \psi\}$.

Now for the negated formulas:

$$\begin{aligned} \mathcal{M}, s &\models \neg[\varphi][\psi]\chi \\ \text{iff } \mathcal{M}, s &\models \varphi \text{ and } \mathcal{M}|_{\varphi}, s \models \neg[\psi]\chi \\ \text{iff } \mathcal{M}, s &\models \varphi \text{ and } \mathcal{M}|_{\varphi}, s \models \psi \text{ and } (\mathcal{M}|_{\varphi})|_{\psi}, s \models \neg\chi \end{aligned}$$

And:

$$\begin{aligned} \mathcal{M}, s &\models \neg[\varphi \wedge [\varphi]\psi]\chi \\ \text{iff } \mathcal{M}, s &\models \varphi \wedge [\varphi]\psi \text{ and } \mathcal{M}|_{\varphi \wedge [\varphi]\psi}, s \models \neg\chi \\ \text{iff } \mathcal{M}, s &\models \varphi \text{ and } (\mathcal{M}, s \not\models \varphi \text{ or } \mathcal{M}|_{\varphi}, s \models \psi) \text{ and } \mathcal{M}|_{\varphi \wedge [\varphi]\psi}, s \models \neg\chi \\ \text{iff } \mathcal{M}, s &\models \varphi \text{ and } \mathcal{M}|_{\varphi}, s \models \psi \text{ and } \mathcal{M}|_{\varphi \wedge [\varphi]\psi}, s \models \neg\chi \end{aligned}$$

(An¬An): $[\varphi]\neg[\psi]\chi \equiv [\varphi](\sim\sim\psi \wedge [\psi]\neg\chi)$:

$$\begin{aligned} \mathcal{M}, s &\models [\varphi]\neg[\psi]\chi \\ \text{iff } \mathcal{M}, s &\not\models \varphi \text{ or } \mathcal{M}|_{\varphi}, s \models \neg[\psi]\chi \\ \text{iff } \mathcal{M}, s &\not\models \varphi \text{ or } (\mathcal{M}|_{\varphi}, s \models \psi \text{ and } (\mathcal{M}|_{\varphi})|_{\psi}, s \models \neg\chi) \\ \text{iff [by classical logic]} \\ \mathcal{M}, s &\not\models \varphi \text{ or } (\mathcal{M}|_{\varphi}, s \models \psi \text{ and } (\mathcal{M}|_{\varphi}, s \not\models \psi \text{ or } (\mathcal{M}|_{\varphi})|_{\psi}, s \models \neg\chi)) \end{aligned}$$

And:

$$\begin{aligned} \mathcal{M}, s &\models [\varphi](\sim\sim\psi \wedge [\psi]\neg\chi) \\ \text{iff } \mathcal{M}, s &\not\models \varphi \text{ or } \mathcal{M}|_{\varphi}, s \models \sim\sim\psi \wedge [\psi]\neg\chi \\ \text{iff } \mathcal{M}, s &\not\models \varphi \text{ or } (\mathcal{M}|_{\varphi}, s \models \psi \text{ and } \mathcal{M}|_{\varphi}, s \models [\psi]\neg\chi) \\ \text{iff } \mathcal{M}, s &\not\models \varphi \text{ or } (\mathcal{M}|_{\varphi}, s \models \psi \text{ and } (\mathcal{M}|_{\varphi}, s \not\models \psi \text{ or } (\mathcal{M}|_{\varphi})|_{\psi}, s \models \neg\chi)) \end{aligned}$$

And for the negated formulas:

$$\begin{aligned} \mathcal{M}, s &\models \neg[\varphi]\neg[\psi]\chi \\ \text{iff } \mathcal{M}, s &\models \varphi \text{ and } \mathcal{M}|_{\varphi}, s \models [\psi]\chi \\ \text{iff } \mathcal{M}, s &\models \varphi \text{ and } (\mathcal{M}|_{\varphi}, s \not\models \psi \text{ or } (\mathcal{M}|_{\varphi})|_{\psi}, s \models \chi) \\ \text{iff [by classical logic]} \\ \mathcal{M}, s &\models \varphi \text{ and } (\mathcal{M}|_{\varphi}, s \not\models \psi \text{ or } (\mathcal{M}|_{\varphi}, s \models \psi \text{ and } (\mathcal{M}|_{\varphi})|_{\psi}, s \models \chi)) \end{aligned}$$

And:

$$\begin{aligned} \mathcal{M}, s &\models \neg[\varphi](\sim\sim\psi \wedge [\psi]\neg\chi) \\ \text{iff } \mathcal{M}, s &\models \varphi \text{ and } \mathcal{M}|_{\varphi}, s \models \neg(\sim\sim\psi \wedge [\psi]\neg\chi) \\ \text{iff } \mathcal{M}, s &\models \varphi \text{ and } (\mathcal{M}|_{\varphi}, s \models \sim\sim\psi \text{ or } \mathcal{M}|_{\varphi}, s \models \neg[\psi]\neg\chi) \\ \text{iff } \mathcal{M}, s &\models \varphi \text{ and } (\mathcal{M}|_{\varphi}, s \not\models \psi \text{ or } (\mathcal{M}|_{\varphi}, s \models \psi \text{ and } (\mathcal{M}|_{\varphi})|_{\psi}, s \models \chi)) \end{aligned}$$

□

Before proving that any formula with public announcements can be rewritten as an equivalent formula of FVEL where the public announcement operator does not occur, we need the next proposition.

Proposition 2.24 (Substitution of Equivalents) *For all formulas φ, ψ, χ of FVEL with public announcements, $\varphi \equiv \psi$ implies $\chi \equiv \chi[\psi/\varphi]$. ($\chi[\psi/\varphi]$ is the formula that results from χ after substitution of all occurrences of φ by ψ .)*

Proof This is a quite straightforward proof by structural induction on χ .

I.H.: If χ' is a proper subformula of χ and $\psi \equiv \varphi$, then $\chi' \equiv \chi'[\psi/\varphi]$.

Base case: $\chi = p$. Assume $\varphi \equiv \psi$. Then, if $\varphi = p$, we have that $\chi \equiv p \equiv \varphi \equiv \psi \equiv \chi[\psi/\varphi]$.

Now, for all cases (including the base case) there are two trivial possibilities: φ is not a subformula of χ , and $\chi = \varphi$. In this cases the result is evident, so in the step we will cover the other cases. Now we look at each case based on the structure of χ ($\psi \equiv \varphi$ is always assumed):

$\chi = \delta \wedge \gamma$. For all models \mathcal{M} and states s , $\mathcal{M}, s \models \delta \wedge \gamma$ iff $\mathcal{M}, s \models \delta$ and $\mathcal{M}, s \models \gamma$ iff (by I.H., $\delta \equiv \delta[\psi/\varphi]$ and $\gamma \equiv \gamma[\psi/\varphi]$) $\mathcal{M}, s \models \delta[\psi/\varphi]$ and $\mathcal{M}, s \models \gamma[\psi/\varphi]$ iff $\mathcal{M}, s \models \delta[\psi/\varphi] \wedge \gamma[\psi/\varphi]$ iff $\mathcal{M}, s \models (\delta \wedge \gamma)[\psi/\varphi]$. So, $\mathcal{M}, s \models \chi$ iff $\mathcal{M}, s \models \chi[\psi/\varphi]$. By analogous reasoning we obtain $\mathcal{M}, s \models \neg\chi$ iff $\mathcal{M}, s \models \neg\chi[\psi/\varphi]$, and therefore $\chi \equiv \chi[\psi/\varphi]$.

The cases for $\chi = \neg\delta$, $\chi = \sim\delta$, $\chi = \Box_i\delta$ and $\chi = [\delta]\gamma$ are also quite straightforward, so we skip them. \square

The proposition above is not surprising for some logics, but note that some very basic properties such as Uniform Substitution do not hold for FVEL. For example, in the Moorean case: $[p]p$ is valid, but $q \wedge \sim\Box_i q$ is not. Now we can prove the following:

Proposition 2.25 *For any formula φ of FVEL with public announcements, a formula φ' of FVEL without public announcements can be found such that $\varphi \equiv \varphi'$.*

Proof This proposition is a direct consequence of Propositions 2.23 and 2.24. A proof can be obtained by induction on the number of announcements in φ . The induction step consists in an inside-out reduction in the number of announcements of φ , by using one of the reduction validities to eliminate one of the most nested announcements, making use of the substitution of equivalents result. We remark that, in the presence of

Proposition 2.24, validities (AnAn) and (An¬An) are redundant: we can make a complete reduction of any formula without using them.

Alternatively, we can use all of the reduction validities presented before, including (AnAn) and (An¬An), and obtain an outside-in reduction as in Plaza (1989, 2007), without making use of Proposition 2.24.¹⁷ \square

2.5.2 Tableaux

To account for public announcements, the tableau system can be extended with the following rule schemas (each of which actually represent eight

$$\begin{array}{cc}
 \text{(RPA1)} & \text{(RPA2)} \\
 \varphi, +i & \varphi, -i \\
 \downarrow & \downarrow \\
 \varphi[\chi/\psi], +i & \varphi[\chi/\psi], -i
 \end{array}$$

rules):

where $\psi \leftrightarrow \chi$ or $\chi \leftrightarrow \psi$ is one of the reduction validities above (except for (AnAn) and (An¬An)).¹⁸ Finally we can prove completeness of the extended tableau system with respect to FVEL with public announcements.

Theorem 2.26 *For any finite set of formulas $\Sigma \cup \{\varphi\}$ of FVEL with public announcements, $\Sigma \vdash \varphi$ iff $\Sigma \models \varphi$.*

Proof The proof system being considered here is the tableau calculus for FVEL (rules R1-R14) augmented with rules RPA1 and RPA2. Soundness is already proven (soundness for the tableau for FVEL is proven in Theorem 2.12, soundness of public announcements' reduction validities is proven in Proposition 2.23 and soundness of the substitution rules RPA1 and RPA2 follows from that and Proposition 2.24).

For completeness, suppose $\Sigma \models \varphi$. When building a tableau for $\Sigma \vdash \varphi$, right after the initial list we just need to apply rules RPA1 and RPA2 until we get equivalent versions without announcements for all formulas

¹⁷ We thank an anonymous reviewer for pointing this out. We show both possibilities, however, because the substitution result of Proposition 2.24 is an essential tool when dealing with FVEL.

¹⁸ See Balbiani, van Ditmarsch, Herzig, and de Lima (2010) for a different approach to tableaux for logics with public announcements, and Hansen (2010) for tableaux for logics with public announcements that use translations as rules in a similar fashion as we do.

in $\Sigma \cup \{\varphi\}$ (which are guaranteed to exist by Proposition 2.25). Let us denote the announcement-free version of Σ by Σ' , and of φ by φ' . First, $\Sigma \models \varphi$ implies $\Sigma' \models \varphi'$. Since the tableau without public announcements is complete, if $\Sigma' \models \varphi'$, then $\Sigma' \vdash \varphi'$. This means there is a closed tableau for $\Sigma' \vdash \varphi'$. But by applying the substitution rules we just obtained a tableau with a single branch that contains all the nodes in the initial list of the tableau for $\Sigma' \vdash \varphi'$. (Notice that adding nodes to the initial list of a tableau does not make it any harder for a tableau to close: these nodes can simply be ignored.) \square

2.6 Related Work

First we will discuss works that have a similar goal to ours, then we will comment on approaches that are comparable to ours from a technical viewpoint.

2.6.1 Logics of Evidence

The logic developed here can be compared to other epistemic logics in the literature that also deal with evidence (Renne, 2009; van Benthem and Pacuit, 2011b; Baltag, Renne, and Smets, 2012, 2014; Carnielli and Rodrigues, 2019; Fitting, 2017).

The closest works to ours have been developed roughly in parallel with it (see Santos (2018)), and come in two very recent papers (Carnielli and Rodrigues, 2019; Fitting, 2017). First, Carnielli and Rodrigues (2019) develop the *basic logic of evidence* (BLE), which is a propositional logic similar to FDE and whose philosophical motivations are closely related to FVEL's. As mentioned in Section 2.2.1, their logic is concerned with preservation of evidence, instead of truth preservation. Then, Fitting (2017) comes even closer to our work by developing a modal logic inspired by BLE. Fitting's logic KX4, however, is different from FVEL. It consists of a classical (two-valued) propositional base, extended with a modality which denotes existence of evidence. The behaviour of BLE, which is somewhat reflected in FVEL's propositional part, appears embedded in KX4 via the modal operator (with its propositional fragment being classical, i.e. representing ontic facts). One can claim that Fitting's approach is more intuitive than ours, but FVEL is, nevertheless, a natural modal extension of a popular many-valued logic (FDE).

Baltag, Renne, and Smets (2014) study a justification logic with an evidence function that resembles awareness functions (Fagin and Halpern, 1987): for each state, it gives a set of justification terms (“good”/correct evidence) that the agent possesses. Differently from Fagin and Halpern (1987), the evidence sets in Baltag et al. (2014) must abide by certain closure conditions. The first obvious difference between that paper and ours is that the only type of evidence being considered is “good” (true) evidence, whereas one of our main goals is to model agents having conflicting evidence.

Nevertheless, in a previous paper by the same authors (Baltag, Renne, and Smets, 2012), contradictory evidence is allowed. Both papers by Baltag et al., however, deal with evidence in a very different way than we do in this chapter. First, they use evidence as justifications for formulas, which are then used to grant explicit status to otherwise implicit beliefs/knowledge. That is, *the role of evidence is to make implicit beliefs explicit*, although implicit beliefs (and even implicit knowledge) are entirely independent from evidence. For this reason, what is called evidence in Baltag et al. (2012, 2014) is conceptually closer to the idea of awareness (as studied in Fagin and Halpern (1987)) than to that of evidence as we intuitively conceive it. We have not talked about factual beliefs in this chapter, but, in the next chapters, the concept of belief will be defined in a way such that it will be semantically dependent on the valuation, which here intuitively represents evidence.

In van Benthem and Pacuit (2011b), neighborhood semantics are employed to model evidence and its dynamics. In their logic, lacking and conflicting evidence is allowed, as well as contradictory beliefs, without implying a trivial epistemic state. One of the highlights of their paper is how they make use of the additional evidence structure to enable interesting dynamics: removal, addition, modification and combination of pieces of evidence. As we plan to do in future work with FVEL, the concept of belief depends entirely on evidence in their semantics (although neither does existence of evidence imply belief, nor vice-versa). That formalism largely differs from ours in a number of aspects. Van Benthem and Pacuit’s semantics for $\Box\varphi$ and $B\varphi$ (“the agent has evidence that implies φ ” and “the agent believes that φ ”, respectively) is rather involved. In comparison, FVEL comprises a simpler semantics, especially with respect to what constitutes possession of evidence. In the following chapters, we will devise a definition of belief for FVEL that depends on the agents’ knowledge of evidence, that is, on statements such as $\mathcal{M}, s \models \Box_i\varphi^t$ (*agent i knows that there is only*

positive evidence for φ). Thereby, realistic (and computationally feasible) rules for belief formation are obtained, such as *agent i believes φ iff she knows that there is only positive evidence for φ* , for example.¹⁹ Moreover, the “logic of combining evidence” in FVEL differs from that of van Benthem and Pacuit’s logic. For example, whereas in FVEL (i) $\mathcal{M}, s \models \Box_i \varphi$ and $\mathcal{M}, s \models \Box_i \psi$ together imply $\mathcal{M}, s \models \Box_i (\varphi \wedge \psi)$, and at the same time it is possible to have (ii) $\mathcal{M}, s \not\models \Box_i (p \vee \neg p)$, which makes agents in FVEL less (classically) logically omniscient (Hintikka, 1979); in their logic (i) does not hold, which allows for more fine-grained evidence, and (ii) is not possible. Whether these are good or bad properties is open for debate. We will dive deeper in this topic in the next chapter.

2.6.2 Other Many-Valued Modal Logics

Many authors have studied the subject of many-valued modal logics (Segerberg, 1967; Thomason, 1978; Ostermann, 1988; Morgan, 1979; Schotch, Jensen, Larsen, and MacLellan, 1978; Morikawa, 1988; Fitting, 1991; Odintsov and Wansing, 2010; Rievieccio, 2014b). Of these, the most closely related to ours are Odintsov and Wansing (2010) and Rievieccio (2014b). Both papers explore some kind of four-valued epistemic logics. We will now discuss the differences between these and our approach.

Logic BK. Odintsov and Wansing (2010) describe a logic called BK (a Belnapian variant of K), which is closely related to FVEL. They also provide a tableau system similar to ours, but their paper does not cover public announcements, nor the correspondence results presented here. There are other small differences between the two formalisms. The logic BK uses two entailment symbols, namely support for truth (\models^+) and support for falsity (\models^-), whereas we opted for an additional negation. While this small change still results in equi-expressive logics, FVEL can express statements like $\mathcal{M}, s \models \neg p \wedge \neg q$ directly, whereas BK always places the “negation” in front of the formula: $\mathcal{M}, s \models^- p \vee q$. The latter has a more natural equivalent in our logic: $\neg(p \vee q)$. Moreover, this choice allows us to announce a formula like $\neg p$, which in BK is only expressible w.r.t. a state of a model ($\mathcal{M}, s \models^- p$).

¹⁹ These rules for belief formation are in line with some discussions in epistemology. Feldman and Conee (1985), for example, argue that a rational belief is one that fits the evidence.

Bilattice Public Announcement Logic. Arieli and Avron (1996) present a four-valued propositional logic based on bilattices (recall Figure 2.2). This logic has been extended to a modal setting by Jung and Rievieccio (2013), and then augmented with public announcements by Rievieccio (2014b,a), who called it Bilattice Public Announcement Logic (BPAL).²⁰ Despite being a very different formalism, BPAL has many similarities with FVEL. First, for the propositional part, the connectives \wedge, \vee, \neg, \sim are identical in both logics. BPAL's \supset can be defined in FVEL by $x \supset y \stackrel{\text{def}}{=} \sim x \vee y$. BPAL defines constants for all truth values. While we can define $\mathbf{t} \stackrel{\text{def}}{=} (\sim p \vee \sim \sim p)$ and $\mathbf{f} \stackrel{\text{def}}{=} \sim \mathbf{t}$, the connectives \perp (always evaluated to *none*) and \top (always *both*) are not definable in FVEL, due to Observation 2.5 (which implies that for any frame, there is a valuation such that no formula has value *both* nor *none*).

Moreover, adding one of them is not enough to define the other, so in order to define all connectives of Arieli and Avron's logic in FVEL we need to add both. With these two new constants, we can also define Arieli and Avron's \oplus and \otimes in FVEL, and consequently all other connectives used in BPAL. In summary, adding \perp and \top to FVEL makes its propositional part equi-expressive with BPAL's.

Now, for the modal part, BPAL uses a four-valued relation. The definition of \Box comes from Jung and Rievieccio (2013), and is motivated by the definition of \Box in the standard translation of modal logic to first order logic: $v(\Box\varphi, w) = \bigwedge \{R(w, w') \rightarrow v(\varphi, w') : w' \in W\}$, where v is their valuation function, which maps pairs of formulas and worlds to one of the four truth values. The implication used in the definition is their connective \rightarrow , which gives a \Box that is strictly more expressive than the one defined with \supset . FVEL's \Box also aligns smoothly with the standard translation:

$$1 \in \overline{\mathcal{V}}(\Box\varphi, s) \text{ iff for all } s' \in S(sRs' \text{ implies } 1 \in \overline{\mathcal{V}}(\varphi, s'))$$

$$0 \in \overline{\mathcal{V}}(\Box\varphi, s) \text{ iff there is a } s' \in S(sRs' \text{ and } 0 \in \overline{\mathcal{V}}(\varphi, s'))$$

Even if we restrict BPAL's four-valued relation to a binary one (as in FVEL), boxed formulas will be evaluated differently. In BPAL, $\Box\varphi$ is *true* in state s whenever there is no accessible state; otherwise, we take the truth values of φ in all the accessible states, replacing *both* by *false*, and take their meet as the truth value of $\Box\varphi$. We can see immediately that in a model where there is only one state s with a reflexive arrow and with

²⁰ See also Bakhtiari, van Ditmarsch, and Rievieccio, which extends Modal Bilattice Logic with epistemic actions.

$\mathcal{V}(p, s) = \{0, 1\}$, $\Box p$ will be *both* in FVEL, but *false* in BPAL. In BPAL, $\Box \varphi$ can actually only assume truth values other than *both*. Moreover, as shown in Proposition 2.14, NEC preserves validity in FVEL, but it does not in BPAL. Despite these differences, the three additional axioms for the modal part given in Jung and Rievieccio (2013) are valid in FVEL (as long as we add the constant \perp). For the public announcements part, if we define $\langle \varphi \rangle \psi \stackrel{\text{def}}{=} \neg[\varphi] \neg \psi$ (plus \perp, \top) in FVEL, all the axioms for public announcements listed in Rievieccio (2014a, Section 4) are also valid here.

To summarise the differences: BPAL has two extra constants, which if added to FVEL make their propositional parts equi-expressive. The modal operator has different behaviour in each logic, NEC preserves validity only in FVEL, but the axioms for BPAL are all valid in FVEL. Another main difference between these works is that we present a tableau calculus, whereas Rievieccio (2014a) has a Hilbert-style axiomatic system.

Levesque’s Logic of Implicit and Explicit Beliefs. Also worth mentioning is Levesque (1984)’s Logic of Implicit and Explicit Belief. Although he is not concerned with the idea of evidence, his framework features a four-valued propositional part and two belief modalities: one implicit and one explicit. Validities are assessed according to standard possible worlds, so all classical tautologies are still valid, but beliefs take into account non-standard “situations”, allowing for non-omniscient agents (at least w.r.t. explicit beliefs).

2.7 Conclusions and Future Work

In this chapter, we presented a multi-agent four-valued logic that can model evidence and what a group of agents know about this evidence. In this way it is possible to model realistic scenarios where agents have access to an inconsistent or incomplete base of information. Some examples are the database scenario described in the introduction, or a robot that collects data through several sensors, which may result in inconsistent data due to sensors’ inaccuracy.

First degree entailment was used as the propositional basis for the logic, with its four-valued atoms playing the role of evidence, where a proposition could be both true and false or have no value at all. A modal layer was built on top of that. The accessibility relation, then, defines the knowledge of the agents about the possibly contradictory or incomplete evidence. Moreover, classical negation was added to the language, increasing its expressivity.

That addition allowed us to define an equivalence operator and reduction validities for public announcements, besides having a natural interpretation in the logic as well. A tableau calculus and some correspondence results were provided.

While on the technical side there are similarities among our approach and others, new results have been presented. Furthermore, the type of situations we model with this many-valued modal logic is different from the ones modelled by other logics of evidence. With this chapter, we aim to contribute to the study of many-valued modal logics by providing an intuitive reading (with potential practical applicability) to these formal tools.

There are several possible directions for further work. First, other update actions (along the lines of van Benthem, van Eijck, and Kooi (2006)) could be studied, such as the actions mentioned in the introduction, which change the informational layer instead of only changing the knowledge about it. These actions, instead of removing states, could just add or remove truth (or falsity) from the value of a proposition in all worlds. In Chapter 4, we explore these dynamic operations for a different interpretation of FVEL.

Our logic does not take into account the *amount* of evidence for and against propositions (as well as other aspects of evidence, such as reliability, source, etc.). FVEL could be modified to include this feature if we define the valuation to be a function $\mathcal{V} : At \times S \rightarrow \mathbb{N} \times \mathbb{N}$, where the first element of the pair, $\mathcal{V}(p, s)^+$, denotes the amount of evidence for p and the second element, $\mathcal{V}(p, s)^-$, denotes the amount of evidence against p . A belief modality could be introduced along the lines of $\mathcal{M}, s \models B_i p$ iff for all t such that sR_it , $\mathcal{V}(p, t)^+ > \mathcal{V}(p, t)^-$.²¹ As stated after Example 2.22, another obvious avenue for improvements is the study of methods for extracting factual knowledge/belief from these evidence models. Agents possessing an inconsistent or incomplete body of evidence could process this information to obtain a consistent epistemic state (along the lines of belief revision, in particular Tamminga (2001)). This is done in the next chapter.

Finally, we note that other interpretations for FVEL can be explored: for example, if we consider each state as the epistemic state of a particular agent, then these states would represent agents, and the accessibility relation would represent relations between them. This opens the possibility for new dynamics, where the knowledge of one agent is influenced by its

²¹ See Pacuit and Salame (2004) for a logic that deals with beliefs in a quantitative way, and Ghosh and de Jongh (2013) for one that compares strength of beliefs qualitatively. We also show how to count pieces of evidence in a different interpretation of FVEL in Chapter 4.

social network (see Baltag, Christoff, Rendsvig, and Smets (2019); Christoff and Hansen (2015)). Chapters 4 and 5 build on this idea.

Chapter 3

Consolidations: Turning Evidence into Belief

3.1 Introduction

Since its earliest formulations (Hintikka, 1962; von Wright, 1951), epistemic logic has been dealing with two out of the three components of the Platonic definition of knowledge (true justified belief),¹ namely truth and belief. With the advent of justification logic,² the missing element entered the picture. Justification logic enabled talking about *reasons* for believing, instead of only whether something is believed or not. We can see this idea of justifications and reasons as representing *evidence*: the agent believes φ if she has evidence for φ .

From the applied point of view, epistemic and doxastic logics have been used for decades to model the knowledge and beliefs of agents (see, for example, Fagin, Halpern, Moses, and Vardi (1995); Meyer and van der Hoek (1995); van Ditmarsch, van der Hoek, and Kooi (2007)). For intelligent agents, especially in real-world settings, however, that “missing element” is essential. These agents will often build up their beliefs from inputs that might be incomplete or even inconsistent. We can think of these inputs as evidence, broadening the concept of justification featured in justification logics. Real agents normally have access to raw, imperfect data, which

¹ This view, found in Plato’s dialogue *Theaetetus*, has been dominant in philosophy for centuries, but has been increasingly challenged since the publication of Gettier (1963).

² Justification logic goes back to Gödel’s work, but has in Artemov (1994, 1995, 2001) some of its earliest modern formulations. Its semantics has its origins in Mkrtychev (1997); Fitting (2005).

they process into a (preferably consistent) set of beliefs, which only then can be used to make sensible decisions and to act.

Like van Benthem and Pacuit (2011b,a); van Benthem, Fernández-Duque, Pacuit, et al. (2012); van Benthem, Fernandez-Duque, and Pacuit (2014); Fitting (2017); Carnielli and Rodrigues (2019); Özgün (2017), we presented in Chapter 2 a logic to model evidence (FVEL). But differently from those, FVEL does not feature a belief modality. Our initial goal here is to add beliefs to our framework. It is of little use to model evidence and not derive any beliefs from it. In the spirit of van Benthem et al. (2014), we assume that rational belief can be determined from evidence.³ However, we do not do that by extending FVEL models, similarly to the strategy in van Benthem et al. (2014). Instead, we extract a doxastic Kripke model representing the agents' beliefs from the FVEL model, which represents their evidence. With that, we not only accomplish the first goal of adding beliefs to the FVEL framework, but also introduce a dynamic perspective on forming beliefs from evidence. This new perspective, compared to the static one in van Benthem and Pacuit (2011b), where evidence and belief coexist, is akin to public announcement logic (Plaza, 2007, 1989; van Ditmarsch, van der Hoek, and Kooi, 2007) compared to epistemic logic: it adds a model-changing aspect. Rational beliefs, although pre-encoded in evidence, are not obtained for free, but require “computation”. This process of forming beliefs from evidence, which we call *consolidation*, is represented by transformations from evidence models to Kripke models. This idea generalises the static approach, because we can represent the “consolidation” of models where belief and evidence coexist as an automorphism from these models to themselves.

This chapter is structured as follows. In Section 3.2, we have a discussion about aspects of evidence and some rationality principles for consolidations. In Section 3.3, we present the main idea of this chapter, the so-called *cautious consolidation*, a transformation from FVEL evidence models to doxastic Kripke models. We also discuss some of its properties. The remainder of the chapter is concerned with comparing our work with another approach in the literature: the work started by van Benthem and Pacuit (2011b) and extended in van Benthem et al. (2012, 2014). Baltag et al. (2016a) also built upon those logics, offering more general topological semantics, but for the purpose of this chapter the models of van Benthem

³ Some philosophers, however, disagree with that. Joyce (2011), for example, writes: “(...) some Bayesians reject the idea that believers with the same objective evidence should end up in the same epistemic state”.

and Pacuit (2011b) will suffice. We cannot compare our consolidations with the ones from van Benthem et al. if we cannot compare those evidence models in the first place, so that is what is done in Section 3.4. Then in Section 3.5 we finally compare the consolidations per se. We lay out our conclusions and ideas left for future work in Section 3.6.

3.2 Preliminaries

We are going to work with FVEL as defined in the previous chapter (Definitions 2.1, 2.3 and 2.4). We will refer to the language of FVEL as specified in Definition 2.1 as $\mathcal{L}_{\square\sim}^n$ (where the exponent n stands for n agents).

FVEL expresses two types of facts: whether there is evidence for and/or against propositions (in a public sense); and first and higher-order knowledge of agents about these evidential facts. Our first goal is to add beliefs to this framework, and that will be done via operations called *consolidations*.

Before formally defining consolidations, a discussion of some preliminary concepts is in order.

3.2.1 Aspects of Evidence

Many different *aspects of evidence* are representable in formal systems. In what follows, we identify some of these aspects.

- **Existence:** the existence of evidence about a proposition. This is an aspect represented in FVEL and van Benthem and Pacuit (2011b).
- **Polarity:** indicates whether the evidence is for or against a certain proposition. This is another aspect of evidence modelled by FVEL.
- **Labelling:** pieces of evidence are named and distinguished. This is found, e.g., in justification logic, where such formulas as $t : \varphi$ indicate that t is a piece of evidence for φ . Notice that this, in principle, enables multiplicity of evidence for one and the same proposition, by the use of different names.
- **Source:** the specification of the sources of each piece of information.
- **Quantity:** the amount of evidence available about a certain proposition. This aspect could also be relative to sources (how many sources provide evidence about a proposition).

- **Reliability:** the degree of trust assigned to evidence (can be relative to sources as well).
- **Internal Structure:** the components that comprise the evidence. For example, if the evidence is a mathematical proof, this aspect would be represented by the structure of this proof, i.e., the lines that comprise it in a certain order. This aspect is also found in justification logic.⁴
- **External Structure:** the relationships between pieces of evidence, such as which pieces of evidence undermine which. These relationships are present in any logical representation of evidence (for example, p will usually undermine $\neg p$), but they can also be of extra-logical origin, in formalisms such as abstract argumentation frameworks (Dung, 1995).
- **Access:** the access each agent has to certain pieces of evidence. This is represented in FVEL as first and higher-order knowledge about evidence. It is also present in justification logic.

At first, we will do our analysis on FVEL, and as such we will limit ourselves to the aspects of evidence with which this logic is concerned, namely *existence*, *polarity*, *external structure* and *access*. In FVEL, the existence of evidence for or against propositions and what the agents know about it is represented, but there is no individuality of pieces of evidence, they are not named, they do not have internal structure, we cannot track the amount of evidence for a certain proposition, nor its reliability or sources. As a logical framework, FVEL inevitably brings some external structure to evidence: if there is evidence for both φ and ψ , then there is also evidence for $\varphi \wedge \psi$. This is a very simplistic picture of evidence, but more complexity would require richer languages.

Intuitively, we define *consolidation* simply as *the process of forming evidence-based beliefs*. More formally, given a certain *evidential landscape* (a setting describing certain aspects of evidence about a set of propositions), we want to be able to say what the agents (should) believe about the factual propositions. This process could be represented as a function from an evidential model to a set of belief sets (one for each agent). However, since we have higher-order knowledge about evidence, it will make sense to include beliefs about other agents' beliefs as well. Assuming that these

⁴ The term *internal structure* was used by Fitting (2009), with the same meaning.

sets of beliefs are deductively closed, our idea of consolidation formally boils down to a morphism from evidential models (FVEL models, at first) to epistemic models.⁵ This formalisation is opportune, for it enables us to exploit the mathematical richness of Kripke structures, instead of only working with unstructured sets of formulas.

Given the simplistic picture of evidence given by FVEL, one could criticise our enterprise by saying that if all we know is that there is some *soft*⁶ evidence for or against certain propositions, consolidating beliefs is very much an exercise in arbitrariness: we can believe anything and still be *consistent* with the evidential landscape. While this is true, there are many possible ways of implementing these consolidation functions between evidential and epistemic models, and some certainly seem more *rational* than others. For example, let us say that there is only evidence in favor of φ (and no evidence against it). How would we react if, despite knowing this, a certain person chooses to believe $\neg\varphi$? Undoubtedly, she would be vulnerable to criticism on grounds of irrationality. So, even in simple evidential landscapes, there are certain standards of rationality expected to be met when consolidating beliefs.

3.2.2 Principles of Rationality

In the previous section we listed some of the aspects of evidence that can be modelled. In this section we will discuss some principles of rationality based on models featuring only two aspects: *existence* and *polarity*. The goal is to make use of this simple setting to provide foundations for the analysis of rationality of more complex consolidations. The results of this section, albeit simple, will have a major import for this and the following chapters.

So, basically, we can have existence/absence of positive/negative evidence about a certain proposition, which gives us four possible evidential situations corresponding to the four truth values presented in Definition 2.3: *true* (only positive evidence), *false* (only negative), *none* and *both* (positive and negative). Consolidation, in this limited scenario, is simply a function

⁵ By epistemic models we mean Kripke models, specially KD45 and S5 models (see, for example, van Ditmarsch, van der Hoek, and Kooi (2007, Chapter 2) or Blackburn, de Rijke, and Venema (2002, Chapter 1)).

⁶ *Soft evidence* is *defeasible* evidence (Baltag, Renne, and Smets, 2012). Since it allows contradictory evidence, FVEL has only soft evidence. Whether the evidence is soft or hard can be seen as the *reliability* aspect. Hard evidence has maximum reliability, whereas soft evidence has less than that. FVEL does not represent this aspect: we cannot know how reliable a piece of evidence is.

from this set of truth values (let us call it $\mathcal{4} = \{t, f, n, b\}$) to the set $\mathcal{3} = \{-1, 0, 1\}$, where 1 represents belief in the formula in question ($B\varphi$), 0 represents belief in its negation ($B\neg\varphi$), and -1 represents absence of belief ($\neg B\varphi \wedge \neg B\neg\varphi$). Since belief in both the formula and its negation trivialises the belief state in traditional epistemic logics, we will ignore this possibility (one could say it is always irrational).⁷ Given this simple formalisation, the spectrum of all consolidations is the set of functions of signature $h : \mathcal{4} \rightarrow \mathcal{3}$, which amounts to $3^4 = 81$ functions, most of which will show to be unreasonable.

Now we will discuss some rationality principles. When formulating these principles, we are judging the adequacy of using these $\mathcal{4} \rightarrow \mathcal{3}$ functions as universal belief-forming processes. In other words, the principles should tell whether it would be rational to apply one such function as a belief-forming process regardless of context and propositions in question, only looking at the four-valued status of evidence.⁸ Of course this is not completely realistic, since we could apply different functions in different contexts. An alternative analysis would be to find which of those functions *could* be rational in some context. For instance, it seems that a function h with $h(n) = 1$ is somewhat inadequate to be taken as a universal belief-forming process (for it would entail trivialisation of the agent's doxastic state), but in a context where the agent is a person in a dangerous jungle and the proposition to be evaluated is *there is a predator nearby*, deciding for 1 even in the absence of evidence might be rational.⁹ Ultimately, entertaining true beliefs is not necessarily a direct goal of a rational agent, but more of a side effect of the attempt at maximising one's "utility". Anyhow, in this section we will take these functions as universal belief-forming processes.

A first principle was mentioned previously (in the end of Section 3.2.1): the agents' beliefs should not be flagrantly contradictory with evidence. If it is known that there is only positive (negative) evidence for φ , then it does

⁷ Although some authors (e.g. Priest and Routley (1989a,b)) support *dialetheism*, which states that there are sentences φ such that φ and $\neg\varphi$ are both true. On the other hand, Harman (1986) maintains that it might be rational to keep contradictory beliefs, once these beliefs are already present. The normative role of logic for beliefs (if any) is a hot (and fascinating) topic of debate (see, e.g., MacFarlane (2004)).

⁸ For advocates of *reliabilism* (in epistemology), a rational belief does not necessarily hinge on a justification, but is instead produced by a *reliable process* (see Goldman (1979); Armstrong (1973); Schmitt (1984); Feldman (1985); Ramsey (1931); Goldman (1975)).

⁹ This case, however, seems more in line with the concept of *acceptance* than with that of belief. See van Fraassen (1980), Stalnaker (1984, Chapter 5) or Harman (1986, Chapter 5) for some discussions on acceptance.

not make sense to believe in $\neg\varphi$ (φ). Let us call this requirement *respect for evidence* (RE). Another principle that seems to be a reasonable requirement for rationality of consolidations is that, if *only* positive (negative) evidence for φ is not enough to induce belief in φ ($\neg\varphi$), then other combinations of evidence are not so either. This principle will be named *unanimity dominance* (UD).

Principles RE and UD stem only from the simple assumptions listed previously. The next ones, however, depend on an additional underlying assumption: FVEL is the logic governing the propositions. In fact, any logic with the relevant properties of FVEL mentioned below is sufficient to justify the next principles.

The observation that $\varphi^n \equiv \neg\varphi^n$ and $\varphi^b \equiv \neg\varphi^b$ in FVEL leaves us with two possible courses of action in this analysis. The first is to recognise that any function $h : \mathcal{4} \rightarrow \mathcal{3}$ with $h(n) \neq -1$ or $h(b) \neq -1$ is *irrational*, for it would imply contradictory beliefs for propositions with those non-classical values. The second possibility is to understand those functions as applicable only to atomic propositions, then solving the contradictory belief problem. To avoid making further assumptions (such as that atomic propositions are special in some way), we will follow the former approach. Let us call this last principle *disregard for ambiguity* (DA).

Similarly, the fact that $\varphi^t \equiv \neg\varphi^f$ and $\varphi^f \equiv \neg\varphi^t$ in FVEL also has some implications. First, it shows that the polarity aspect is somewhat superfluous from FVEL's perspective, for evidence against φ is really just evidence for $\neg\varphi$ (note, however, that this is a peculiarity of FVEL, and not a general truth). Second, it prompts us to derive another postulate, for it implies that, if there is only positive (negative) evidence for some proposition φ , then there is only negative (positive) evidence for its negation. Therefore, an agent who decides to believe φ ($\neg\varphi$) based on this evidence will believe the negation of $\neg\varphi$ (φ), assuming that *double negation elimination* is present, which is the case for FVEL. From this we devise another postulate, *dependence of opposites* (DO). This postulate is not so much a rationality requirement, but more of an inevitability (in FVEL).

These principles can be formalised via rationality postulates.¹⁰ A function $h : \mathcal{4} \rightarrow \mathcal{3}$ is a *rational consolidation* iff:

$$\begin{array}{ll} \text{(RE+)} & h(t) \neq 0 \\ \text{(RE-)} & h(f) \neq 1 \\ \text{(UD+)} & h(t) \neq 1 \Rightarrow \forall v \in \mathcal{4} (h(v) \neq 1) \\ \text{(UD-)} & h(f) \neq 0 \Rightarrow \forall v \in \mathcal{4} (h(v) \neq 0) \end{array}$$

¹⁰ Our methodology here is partially inspired by AGM theory of belief revision (Alchourrón, Gärdenfors, and Makinson, 1985).

$$\begin{array}{ll}
(\text{DA}+) & h(b) = -1 \\
(\text{DO}) & h(t) = 1 \Leftrightarrow h(f) = 0
\end{array}
\qquad
\begin{array}{ll}
(\text{DA}-) & h(n) = -1
\end{array}$$

In the presence of postulates $\text{RE}\pm$, $\text{DA}\pm$ imply $\text{UD}\pm$, but they are listed anyway since they are independently justified (and one may still prefer the approach without $\text{DA}\pm$). Figure 3.1 has three examples of consolidations.

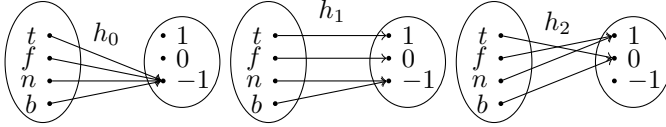


Figure 3.1: Consolidations h_0 and h_1 respect all postulates; h_2 only respects DO.

Excluding DO, these postulates leave us with only 4 rational consolidations (18 without $\text{DA}\pm$) out of the 81 possible ones: h_0 , h_1 , h_3 and h_4 ($h_3(n) = h_4(n) = h_3(b) = h_4(b) = -1$, $h_3(f) = 0$, $h_3(t) = -1$, $h_4(t) = 1$, $h_4(f) = -1$). With DO, however, only h_1 and the *absolutely sceptical* function h_0 remain. Since h_0 shows an unfruitful scepticism (despite being rational), h_1 stands out as the one interesting rational function.¹¹ In other words, if there is a $\mathcal{4} \rightarrow \mathcal{3}$ function that can be used as a useful and universal consolidation process, this function is h_1 . This does not mean, however, that other functions could not be rationally applied in specific cases. Therefore, for the following study of FVEL consolidations, we do not have a cogent argument forcing h_1 to be always respected. Nevertheless, it can and will be used as a sensible starting point and baseline.

3.3 A Consolidation Operation

Now that our preliminary concepts are in place, we want to be able to extract a Kripke model from an FVEL model, representing the beliefs obtained from the evidence in the latter, constituting a so-called *consolidation* operation.

3.3.1 Definitions

To define this operation we will need some essential notions:

¹¹ It turns out that our requirements for “respecting the evidence” are not as strict as elsewhere in the literature. For Feldman (2005), for instance, an agent respects her evidence when her beliefs correspond to what her evidence indicates (similarly to what h_1 does).

Definition 3.1 (Selection Function and Accepted Valuations) Let $Val = \{v : At \rightarrow \{0, 1\}\}$ be the set of all binary valuations. Given an FVEL model $\mathcal{M} = (S, R, \mathcal{V})$ and the set of agents $A = \{1, 2, \dots, n\}$, we define $\mathcal{V} = (\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_n)$, where $\mathcal{V}_i(s) \subseteq Val$ and $\mathcal{V}_i(s) \neq \emptyset$, for all $i \in A$ and $s \in S$. \mathcal{V} is called a (valuation) selection function for \mathcal{M} , and $\mathcal{V}_i(s)$ is the set of binary valuations that agent i accepts at s . $U_s = \bigcup_{i \in A} \mathcal{V}_i(s)$ is the set of valuations accepted by some agent at s .

Intuitively, the selection function \mathcal{V} gives the set of valuations that each agent finds plausible at each state. The idea is that these plausible valuations will bear a strong connection to the evidence possessed, by means of constraints imposed on \mathcal{V} . In principle, however, \mathcal{V} can be any function conforming to Definition 3.1.

We use s_v to denote the pair (s, v) , where $s \in S$ and $v \in Val$. Now we define *cluster consolidations* (Definition 3.2). Ideally, the consolidation would generate one state for each state in \mathcal{M} , with the same valuation. If FVEL were two-valued, that would be possible, but since it is four-valued, we generate a cluster of states for each state s , with one state s_v for each valuation v accepted at s according to \mathcal{V} .

Definition 3.2 (Cluster Consolidation) Let $\mathcal{M} = (S, R, \mathcal{V})$ be an FVEL model, \mathcal{V} be a selection function for \mathcal{M} . The cluster consolidation of \mathcal{M} (based on \mathcal{V}) is the Kripke model $\mathcal{M}! = (S', R', V)$, where:

- i. $S' = \{s_v \mid s \in S, v \in U_s\}$,¹²
- ii. if $s_v, t_u \in S'$ then: $s_v R'_i t_u$ iff $s R_i t$ and $u \in \mathcal{V}_i(t)$; and
- iii. $V(p, s_v) = v(p)$.

Definition 3.2 hopefully covers most reasonable consolidations, modulo some notion of equivalence. It covers a lot of unreasonable ones too. It does not reflect, however, any specific “consolidating policy”: it only defines a technically convenient class of consolidations, due to their modular nature (each state generating a cluster of states) and the way they link accepted valuations and evidence.

Now we define a type of cluster consolidation reflecting an actual policy: *cautious consolidation*. It is based on the following consolidating principle: *If there is only positive evidence for a proposition, then the agent believes it; if there is only negative evidence, then the agent believes its*

¹² Since the number of states in $\mathcal{M}!$ can be exponential in the number of elements of At , if At is countably infinite, S' may be uncountable (by Cantor’s Theorem).

negation; otherwise, the agent has no opinion about it. This principle can be codified in function h_1 , as discussed above. Consider the set of functions $H = \{h : \mathcal{Z} \rightarrow \mathcal{Z}\}$, mapping status of evidence to doxastic attitudes.

Definition 3.3 (Compatibility) ¹³ Let $h \in H$ and $Val_s^h = \{v \in Val \mid \text{for all } p \in At, \text{ if } h(\mathcal{V}(p, s)) \neq -1 \text{ then } v(p) = h(\mathcal{V}(p, s))\}$ be the set of binary valuations h -compatible with \mathcal{V} at s .

Definition 3.4 (Implementation) If $\mathcal{V}_i(s) = Val_s^h$ for all $s \in S$ and some $i \in A$, we say that \mathcal{V} implements h for agent i .

Definition 3.5 (h -consolidation) Let $h \in H$. $\mathcal{M}!$ is called an h -consolidation of \mathcal{M} for agent i iff $\mathcal{M}!$ is the cluster consolidation of \mathcal{M} based on \mathcal{V} , and \mathcal{V} implements h for agent i .

Let *cautious consolidation* be synonymous with h_1 -consolidation. A consolidation is characterised in Definition 3.5 relative to an agent. This allows consolidations to implement different belief formation policies for each agent.

Example 3.6 Figure 3.2 (left) shows a simple cautious consolidation, with one agent and one proposition with value true. The selection function is cautious, so the set of valuations accepted by the agent has to be h_1 -compatible with \mathcal{V} at s_1 . This is the case for a valuation v only if $v(p) = 1$. Then, according to Definition 3.2, there is only one state in the consolidated model (s'_1), which conforms to v (that is, p holds) and has a reflexive arrow, because the original state s_1 has one as well.

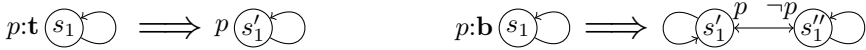


Figure 3.2: Cautious consolidations on positive (left) and conflicting evidence (right).

In Figure 3.2 (right), the value both for p admits two h_1 -compatible valuations: one in which p holds, and one in which p does not hold. Then, by Definition 3.2, two states must exist in the consolidation, and they should contain all possible arrows, because the original state has a reflexive arrow. The consolidation would be identical if p had value none: cautious consolidations do not distinguish between none and both (due to h_1).

¹³ For this and coming definitions, keep in mind that whenever \mathcal{V} , S or \mathcal{V} are mentioned, they are always relative to an underlying FVEL model $\mathcal{M} = (S, R, \mathcal{V})$.

Example 3.7 Figure 3.3 illustrates cautious consolidation applied to Example 2.7.

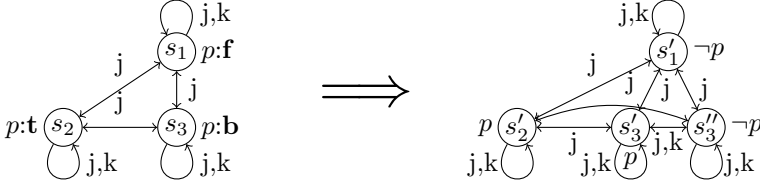


Figure 3.3: Cautious consolidation of Example 2.7.

The original model has three states and two agents. Each one of states s_1 and s_2 will have one corresponding state in the consolidated model (s'_1 and s'_2 , respectively), due to their valuations of p being false and true, respectively. On the other hand, s_3 will generate two states, since p has value both there, and accordingly has two h_1 -compatible valuations. Regarding accessibility, all states will have reflexive arrows, due to the original model being reflexive (this preservation property is shown later by Proposition 3.17). Connections between states of different clusters (for example, s'_1 and s'_3 , which were generated by s_1 and s_3 , respectively) will respect the connections between their matching states in the original model. Finally, s'_3 and s''_3 will be connected for both agents because both of them come from s_3 , and the FVEL model is reflexive.

3.3.2 Other Cluster Consolidations

Let us explore other forms of cluster consolidation. As can be anticipated by the name cautious consolidation, less cautious operations can be devised, in the sense that they might create more false beliefs.¹⁴ These strategies may be realised in two ways. One is to use a selection function \mathcal{V}_i implementing a function $h' : \mathcal{4} \rightarrow \mathcal{3}$ other than h_1 . For instance, if h' is like h_1 except that $h'(b) = 1$, then h' is a strategy that, in the face of conflicting evidence, always trusts positive evidence. We will not pursue this strategy because, as seen in Section 3.2.2, one can dismiss these functions as irrational. The other way is to use a \mathcal{V}_i that do not implement any $\mathcal{4} \rightarrow \mathcal{3}$ function. Then, differently from the previous one which just implements a general

¹⁴ Strictly speaking, even cautious consolidation already risks false beliefs. Despite all evidence pointing to the truth of a certain proposition, it can still be false (and vice versa). This evokes the whole *internalism* versus *externalism* debate in epistemology (Goldman, 2009).

strategy, additional decision making will be needed: agents with conflicting or missing evidence about a *certain* proposition p will be able to decide whether they believe p or $\neg p$.

This decision process for propositions with value *both* presupposes that the agents have some *additional information* besides the aspects of evidence represented by FVEL, for otherwise deciding between p or $\neg p$ would be completely arbitrary. For the value *none*, however, any decision would be arbitrary, since p having truth value *none* means that there is no evidence about p whatsoever.

One type of operation including these decision processes is *evaluative consolidation*. It is based on the principle that, if an agent has both positive and negative evidence about a proposition, she can ponder in which one she believes. In practice, this might reduce the number of states in the consolidated model. Let v^{-p} be the valuation identical to $v \in Val$ except that $v^{-p}(p) = |v(p) - 1|$.

Definition 3.8 (Evaluative Consolidation) *For all $s \in S$ and some $i \in A$, let $\mathcal{V}_i(s) \in E_s$, where $E_s = \{X \subseteq Val_s^{h_1} \mid X \neq \emptyset \text{ and } \forall p \in P, \forall v \in X: V(p, s) = \emptyset \Rightarrow \exists u \in X(u = v^{-p})\}$. We say \mathcal{V} is an evaluative selection function for agent i (or \mathcal{V}_i is an evaluative selection function). Evaluative consolidation is defined in the obvious way (analogously to Definition 3.5).*

Notice that \mathcal{V}_i carries within it not only the additional information mentioned earlier, but also some subjective judgement made by agent i , which could be different for another agent j , even in the same state.

Example 3.9 Figure 3.4 shows again Example 2.7, but this time agents k and j are performing (different) evaluative consolidations.

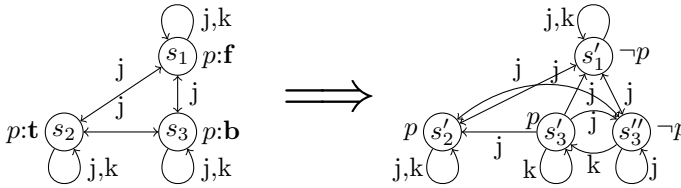


Figure 3.4: Evaluative consolidations applied to Example 2.7 (compare to Figure 3.3).

While agent k has a selection function such that there is no $v \in \mathcal{V}_k(s_3)$ such that $v(p) = 0$, for agent j there is no $v \in \mathcal{V}_j(s_3)$ such that $v(p) = 1$,

that is, in state s_3 , Kate would favour evidence for p while John would favour evidence against p . Notice how there are no k -arrows going to s_3'' , and no j -arrows going to s_3' . Notice also that if both agents had opted for p , state s_3'' would not exist in the consolidated model.

We can also define an operation which is dual to the last one, *opinionated consolidation*. This one is based on the principle that, in the absence of evidence, the agents may just come up with a truth value for a given proposition. Despite the mathematical similarity, opinionated consolidation cannot be considered rational, for it is based on arbitrary choices: if a proposition has value *none* then there is no evidence about it, therefore the judgement made by the agent is not grounded in evidence, but purely on subjective preferences.¹⁵

Definition 3.10 (Opinionated Consolidation) *For all $s \in S$ and some $i \in A$, let $\mathcal{V}_i(s) \in O_s$, where $O_s = \{X \subseteq \text{Val}_s^{h_1} \mid \forall p \in P, \forall v \in X: V(p, s) = \{0, 1\} \Rightarrow \exists u \in X(u = v^{-p})\}$. We say \mathcal{V} is an opinionated selection function for agent i . Opinionated consolidation is defined in the obvious way.*

This consolidation has more of a descriptive appeal than a normative one; although it does not look rational, it can at least be considered “natural”. A mixture of evaluative and opinionated consolidations can also be devised. In the case where all propositions with values *both* or *none* are disambiguated by the agent, the consolidation yields the maximum amount of beliefs.

Definition 3.11 (Mixed Consolidation) *For all $s \in S$ and some $i \in A$, $\mathcal{V}_i(s) \subseteq \text{Val}_s^{h_1}$. We say \mathcal{V} is a mixed selection function for agent i . Mixed consolidation is defined in the obvious way.*

It is not difficult to see that every cautious consolidation is an evaluative, opinionated and mixed consolidation as well, but a very *minimal* one at that. This idea of minimality refers to the number of decisions made by the agent: an evaluative (or opinionated, or mixed) consolidation is also cautious if the agent does not make any decision that she could possibly do, or, alternatively, accepts all possible valuations. In the same vein,

¹⁵ Despite our not clearly defining what is considered evidence, defining what falls within this category would have great import for the valuation V . Propositions valued *none* under certain definitions, for example, could have other values under others.

consolidations can be *maximal* if all decisions are made, that is, only a minimal number of valuations are accepted. Minimal consolidations in any of these classes (cautious, evaluative, opinionated and mixed) are also instances of consolidations of all the other classes. Maximal mixed consolidations have selection functions that map every state to a singleton set (a set with one valuation). Cautious consolidations are unique, so they are always maximal and minimal at the same time.

Although the context is different, it is worth noticing the parallel between cautious consolidations and maximal mixed consolidations on the one hand, and *full meet* and *maxichoice* operations in belief revision (Alchourrón, Gärdenfors, and Makinson, 1985) on the other. The former is the most cautious type of revision, making an intersection of all acceptable outcomes, whereas the latter is the most reckless, picking only one element of the set of acceptable outcomes. As with belief revision, here we are left with a dilemma between maximising the epistemic state and minimising the risk of incurring into false beliefs.

3.3.3 Properties

In this section we explore formal properties of consolidations. Proposition 3.12 represents a desideratum for cluster consolidations: that they “respect” the function h upon which they are based. In a cautious consolidation, for example, we want that if an agent a knows that the status of evidence for p is t in state s , that is, $\mathcal{M}, s \models \Box_a p^t$, then in the corresponding state of $\mathcal{M}!$ a will believe p . Now if $\Box_a p^f$ holds, a will believe $\neg p$, and otherwise a will believe neither p nor $\neg p$. Proposition 3.12 generalises this result for any function $h \in H$, for any number of “stacked boxes”, and for disjunctions of truth values of p . For example, with h_1 , if $\Box_a(p^b \vee p^n)$ holds, then the agent will not form beliefs about p . Let $h^{-1}(y)$ be the preimage of y by h : $h^{-1}(y) = \{x \in \mathcal{P}(\{0, 1\}) \mid h(x) = y\}$.

Proposition 3.12 *Given any FVEL model $\mathcal{M} = (S, R, \mathcal{V})$ and a function $h \in H$, consider an h -consolidation $\mathcal{M}! = (S', R', V)$ of \mathcal{M} for agent i_0 . For any such consolidation, for all $p \in \text{At}$ and $s \in S$: $\mathcal{M}, s \models \Box_{i_n} \dots \Box_{i_0}(p^{x_1} \vee \dots \vee p^{x_m}) \Rightarrow$*

$$\begin{cases} \mathcal{M}!, f(s) \models B_{i_n} \dots B_{i_0} p & \text{if } \{x_1, \dots, x_m\} \subseteq h^{-1}(1) \\ \mathcal{M}!, f(s) \models B_{i_n} \dots B_{i_0} \neg p & \text{if } \{x_1, \dots, x_m\} \subseteq h^{-1}(0) \\ \mathcal{M}!, f(s) \not\models B_{i_n} \dots B_{i_0} p & \text{if } \{x_1, \dots, x_m\} \cap h^{-1}(1) = \emptyset \\ \mathcal{M}!, f(s) \not\models B_{i_n} \dots B_{i_0} \neg p & \text{if } \{x_1, \dots, x_m\} \cap h^{-1}(0) = \emptyset \end{cases}$$

where for all $s \in S$, $f(s) = s_v$ for some $s_v \in S'$, and B_a is the belief modality associated with R'_a .

Before proving this proposition, we need the following lemma:

Lemma 3.13 *Let $\mathcal{M} = (S, R, \mathcal{V})$ be an FVEL model, $s \in S$, and $\mathcal{M}!$ be a cluster consolidation of \mathcal{M} implementing $h \in H$ for agent i . If $u, v \in U_s$, then $\mathcal{M}!, s_v \models B_i\varphi$ iff $\mathcal{M}!, s_u \models B_i\varphi$, for any formula φ .*

Proof We have $s_v R'_i t_w$ iff $(s R_i t$ and $w \in \mathcal{V}_i(t))$ iff $s_u R'_i t_w$. So $\{s' \in S' \mid s_v R'_i s'\} = \{s' \in S' \mid s_u R'_i s'\}$, and as such $\mathcal{M}!, s_v \models B_i\varphi$ iff $\mathcal{M}!, s_u \models B_i\varphi$. \square

Proof of Proposition 3.12 The proof will be by induction on n , but we will first prove separately the case when $n = 0$.

We want to show that if $\mathcal{M}, s \models \Box_i(p^{x_1} \vee \dots \vee p^{x_m})$ and:

- i. $\{x_1, \dots, x_m\} \subseteq h^{-1}(1)$, then $V(p, t_u) = 1$ in all states t_u such that $f(s)R'_i t_u$;
- ii. $\{x_1, \dots, x_m\} \cap h^{-1}(1) = \emptyset$, then $\exists t_u \in S'$ s.t. $f(s)R'_i t_u$, where $V(p, t_u) = 0$;
- iii. $\{x_1, \dots, x_m\} \subseteq h^{-1}(0)$, then $V(p, t_u) = 0$ in all states t_u such that $f(s)R'_i t_u$;
- iv. $\{x_1, \dots, x_m\} \cap h^{-1}(0) = \emptyset$, then $\exists t_u \in S'$ s.t. $f(s)R'_i t_u$, where $V(p, t_u) = 1$.

This entails the proposition. We will analyse each case:

$\{x_1, \dots, x_m\} \subseteq h^{-1}(1)$: Let $t_u \in S'$ be such that $f(s)R'_i t_u$. Since $\mathcal{M}, s \models \Box_i(p^{x_1} \vee \dots \vee p^{x_m})$, we have that $\mathcal{M}, t \models p^{x_1} \vee \dots \vee p^{x_m}$ for all t such that $s R_i t$. But this is true for t iff $M, t \models p^{x_1}$ or \dots or $\mathcal{M}, t \models p^{x_m}$. Since $f(s)R'_i t_u$, it holds that $s R_i t$ and $u \in \mathcal{V}_i(t)$. So u is h -compatible (with \mathcal{V} at t), and since $h(x_1) = \dots = h(x_m) = 1$, we have $V(p, t_u) = 1$. The case for $\{x_1, \dots, x_m\} \subseteq h^{-1}(0)$ is analogous.

$\{x_1, \dots, x_m\} \cap h^{-1}(1) = \emptyset$: Similar to the previous case, but now we have that $h(x_1) \neq 1, \dots, h(x_m) \neq 1$, so since $\mathcal{V}_i(t)$ is the set of h -compatible valuations (with \mathcal{V} at t), for any $u \in \mathcal{V}_i(t)$ either $u(p) = 0$ or $u^{-p} \in \mathcal{V}_i(t)$. In either case (ii) is satisfied. Case (iv) is analogous.

We now show that the proposition hold for the base case, where $n = 1$, and then we extend the result to all $n \geq 1$ by induction.

Assume $n = 1$ and $\mathcal{M}, s \models \Box_{i_1} \Box_{i_0}(p^{x_1} \vee \dots \vee p^{x_m})$. Let us prove by cases.

$\{x_1, \dots, x_m\} \subseteq h^{-1}(1)$: By $\mathcal{M}, s \models \Box_{i_1} \Box_{i_0} (p^{x_1} \vee \dots \vee p^{x_m})$ and the semantics of FVEL we conclude that (1) for all t, r such that $sR_{i_1} t R_{i_0} r$ we have $\mathcal{M}, s \models p^{x_1} \vee \dots \vee p^{x_m}$. From Definition 3.2 we have that $f(s)R'_{i_1} t_u R'_{i_0} r_v$ iff $sR_{i_1} t R_{i_0} r$ and $u \in \mathcal{V}_{i_1}(t)$ and $v \in \mathcal{V}_{i_0}(r)$. Fact (1) implies $\mathcal{M}, r \models p^{x_1}$ or ... or $\mathcal{M}, r \models p^{x_m}$. Formulas of type φ^y are satisfied in a state s iff $\overline{\mathcal{V}}(\varphi, s) = y$. This means that (1) implies $\overline{\mathcal{V}}(p, r) \in \{x_1, \dots, x_m\}$. But since $\{x_1, \dots, x_m\} \subseteq h^{-1}(1)$ and \mathcal{V}_{i_0} is h -compatible with \mathcal{V} at r , we have that $\mathcal{M}!, r_v \models p$ for all t_u and r_v such that $f(s)R'_{i_1} t_u R'_{i_0} r_v$. This concludes this case. The case for $\{x_1, \dots, x_m\} \subseteq h^{-1}(0)$ is analogous.

$\{x_1, \dots, x_m\} \cap h^{-1}(1) = \emptyset$: Similar to the previous case but now \mathcal{V}_{i_0} being h -compatible with \mathcal{V} at r implies that for all t_u there is some r_v s.t. $f(s)R'_{i_1} t_u R'_{i_0} r_v$ with $\mathcal{M}!, r_v \models \neg p$. This concludes this case. The case for $\{x_1, \dots, x_m\} \cap h^{-1}(0) = \emptyset$ is analogous.

Now we can use induction to finish the proof. As Induction Hypothesis (I.H.) we assume the proposition is valid for $n = k - 1$, and from this we prove that it is valid for $n = k$. Suppose that $\mathcal{M}, s \models \Box_{i_k} \dots \Box_{i_0} (p^{x_1} \vee \dots \vee p^{x_m})$. Again, let us go by cases.

$\{x_1, \dots, x_m\} \subseteq h^{-1}(1)$: we have to show that $\mathcal{M}!, f(s) \models B_{i_k} \dots B_{i_0} p$. By the semantics of FVEL we have that for all t s.t. $sR_{i_k} t$ we have $\mathcal{M}, t \models \Box_{i_{k-1}} \dots \Box_{i_0} (p^{x_1} \vee \dots \vee p^{x_m})$, but the I.H. this implies $\mathcal{M}!, f(t) \models B_{i_{k-1}} \dots B_{i_0} p$. But by Definition 3.2 we have that $s_v R'_{i_k} t_u$ iff $sR_{i_k} t$ and $u \in \mathcal{V}_{i_k}(t)$. Using Lemma 3.13 we have that for any such t_u it holds that $\mathcal{M}!, t_u \models B_{i_{k-1}} \dots B_{i_0} p$. This, of course, implies $\mathcal{M}!, f(s) \models B_{i_k} \dots B_{i_0} p$, which concludes this case. The other cases are identical, since the case condition is only relevant for the application of the I.H. \square

Function h is respected in a weak way, namely, only for atoms. Now consider the following translation function for formulas.

Definition 3.14 (Translation Function) *Let $\mathfrak{t} : \mathcal{L}_{\Box}^n \rightarrow \mathcal{L}_B^n$ be a function that translates FVEL formulas into a standard multimodal language with modal operators B_a for each $a \in A$ such that \sim is replaced by \neg , \Box_a is replaced by B_a , and the rest remains the same.*

The following result, as Proposition 3.12, establishes a correspondence between formulas in an FVEL model and in its consolidation. The result is limited to formulas with “classically-valued” atoms, but encompasses all formulas instead of only atoms.

Proposition 3.15 *Let $\mathcal{M} = (S, R, \mathcal{V})$ be an FVEL model and $\mathcal{M}! = (S', R', V)$ its cautious consolidation, and let φ be an FVEL formula such*

that for all atoms p occurring in φ , $\mathcal{V}(p, s) \in \{\{0\}, \{1\}\}$ for all $s \in S$. Then, for all $s \in S$, $\mathcal{M}, s \models \varphi$ iff $\mathcal{M}!, s_v \models \mathbf{t}(\varphi)$, for any $s_v \in S'$.

Proof This proposition can be proven by a simple induction on the structure of φ . The base case is the case for atoms, and the Induction Hypothesis is that the proposition holds for proper subformulas of φ . \square

Now let us check the preservation of frame properties under consolidations. Seriality, transitivity and Euclideanicity are preserved in general. Reflexivity and symmetry, however, are only preserved if there is a certain similarity among the selection functions \mathcal{V}_i . Notice that for all R'_i to be reflexive, all functions \mathcal{V}_i have to be equal. The following propositions are all relative to an FVEL model $\mathcal{M} = (S, R, \mathcal{V})$ and a cluster consolidation $\mathcal{M}! = (S', R', V)$ of \mathcal{M} , where $R = (R_1, \dots, R_n)$ and $R' = (R'_1, \dots, R'_n)$.

Proposition 3.16 *If R_i is serial (transitive, Euclidean), then R'_i is serial (transitive, Euclidean).*

Proposition 3.17 *If R_i is reflexive, then R'_i is reflexive iff for all $j \in A$ and all $s \in S$ it holds that $\mathcal{V}_j(s) \subseteq \mathcal{V}_i(s)$.*

Proposition 3.18 *If R_i is symmetric, then R'_i is symmetric iff for all $s, t \in S$ such that sR_itR_is it holds that $\mathcal{V}_j(s) \subseteq \mathcal{V}_i(s)$ for all $j \in A$.*

In the case where all the agents consolidate in the same manner (for example, through cautious consolidation), reflexivity, symmetry, transitivity, seriality and Euclideanicity are all preserved. Since we want the consolidated model to be a *doxastic* model, it is desirable that its relation be Euclidean, serial and transitive (KD45 models). These results provide sufficient conditions for that.

3.3.4 A Unified Language for Evidence and Beliefs

A detailed study of an extension of the language and logic of FVEL with beliefs is beyond the scope of this thesis, but we will suggest here how this can be done.

First, we have to recall that propositional formulas in FVEL are not about facts, but about evidence. For this reason, it is better to define belief over formulas of \mathcal{L}_B , the doxastic language of the consolidated model. We can define belief in FVEL model as follows:

$$\mathcal{M}, s \models B_a \mathbf{t}(\varphi) \quad \text{iff} \quad \mathcal{M}!, s_v \models B_a \mathbf{t}(\varphi)$$

where $\mathcal{M}! = (S', R', V)$ is the cautious consolidation of \mathcal{M} , and $s_v \in S'$.

In this language it is now possible to talk about formulas such as $\Box_a p^t \leftrightarrow B_a p$ or $\Box_a p^f \leftrightarrow B_a \neg p$, i.e., only positive (negative) evidence equals belief (disbelief), where $\varphi \leftrightarrow \psi \stackrel{\text{def}}{=} \sim(\varphi \wedge \sim \psi) \wedge \sim(\psi \wedge \sim \varphi)$. These formulas are valid, but if we employ another type of consolidation in the semantic definition above, they may not be.

Notice also that if $\mathcal{M}!$ is a KD45 model, for example, the behaviour of this new B_a operator in FVEL will be governed by that logic. But since the consolidation is completely determined by the original FVEL model, it should be possible to define semantics for B_a in FVEL without mentioning $\mathcal{M}!$.

3.4 Equivalence Between Evidence Models

Now we recall van Benthem and Pacuit (2011b)'s models (hereafter, B&P models). The goal is to compare, later, consolidations in B&P and FVEL models.

Definition 3.19 (*van Benthem and Pacuit, 2011b*) *A B&P model is a tuple $M = (S, E, V)$ with $S \neq \emptyset$ a set of states, $E \subseteq S \times \mathcal{P}(S)$ an evidence relation, and $V : At \rightarrow \mathcal{P}(S)$ a valuation function. We write $E(w)$ for the set $\{X \mid wEX\}$. We impose two constraints on E : for all $w \in S$, $\emptyset \notin E(w)$ and $S \in E(w)$.*

In B&P models, propositional formulas are about facts (not evidence), as usual.

Definition 3.20 (*van Benthem, Fernandez-Duque, and Pacuit, 2014*) *A w -scenario is a maximal $\mathcal{X} \subseteq E(w)$ such that for any finite $\mathcal{X}' \subseteq \mathcal{X}$, $\bigcap \mathcal{X}' \neq \emptyset$. Let $Sce_E(w)$ be the collection of w -scenarios of E .*

Definition 3.21 (*van Benthem and Pacuit, 2011b*) *A standard bimodal language $\mathcal{L}_{\Box B}$ (with \Box for evidence and B for belief) is interpreted over a B&P model $M = (S, E, V)$ in a standard way, except for B and \Box :*

$$M, w \models \Box \varphi \text{ iff } \exists X \text{ with } wEX \text{ and } \forall v \in X : M, v \models \varphi$$

$$M, w \models B \varphi \text{ iff } \forall \mathcal{X} \in Sce_E(w) \text{ and } \forall v \in \bigcap \mathcal{X}, M, v \models \varphi$$

Formulas such as $\Box\varphi$ mean that *the agent has evidence for φ* . Notice that an agent can have evidence for φ and $\neg\varphi$ at the same time, or have no evidence about φ whatsoever. This makes the status of evidence (in any given state) four-valued, just as in FVEL. Note also that the conditions for the satisfaction of $B\varphi$ tell us how the consolidation in B&P logic is done: *One believes what is supported by all pieces of evidence in all maximal consistent subsets of one's evidence (w -scenarios)*.

Now we want to be able to compare consolidations of B&P models to consolidations of FVEL models. For this, first, we need a way of establishing that an FVEL model and a B&P model are “equivalent” with respect to how evidence is represented. It only makes sense to compare consolidations if they depart from (roughly) the same evidential situation.

The “logics of evidence” in B&P logic and FVEL differ, the former being non-normal (so, for example, $\Box\varphi \wedge \Box\psi$ does not imply $\Box(\varphi \wedge \psi)$ in B&P logic, while in FVEL it does), and the latter being *First Degree Entailment* (FDE) (Dunn, 1976; Priest, 2008).¹⁶ Note, however, that this difference is more about how evidence is *manipulated* in these logics, than about how it is *represented*. For this reason, our equivalence in evidence is, fittingly, limited to literals.

Definition 3.22 (ev-equivalence) *Let $M = (S, E, V)$ be a B&P model and let $\mathcal{M} = (S', R, \mathcal{V})$ be an FVEL model. A relation $\dot{=}\subseteq S \times S'$ is an ev-equivalence between M and \mathcal{M} iff:*

1. $\dot{=}$ is a bijection;
2. If $s \dot{=} s'$, where $s \in S$ and $s' \in S'$, then, for all $p \in \text{At}$: $M, s \models \Box p$ iff $\mathcal{M}, s' \models \Box p$; and $M, s \models \Box\neg p$ iff $\mathcal{M}, s' \models \Box\neg p$.

We write $M \dot{=} \mathcal{M}$ if there exists an ev-equivalence between M and \mathcal{M} . $M \dot{=} M'$, $\mathcal{M} \dot{=} M$ and $\mathcal{M} \dot{=} \mathcal{M}'$ are defined analogously.

Now our job is to find, for each B&P or FVEL model, a model of the other type which is ev-equivalent to it, that is, that *represents* the same evidence.¹⁷ Since B&P models are single-agent, we assume from now

¹⁶ In other words: if there is evidence for Σ and $\Sigma \vdash_{\text{FDE}} \varphi$, then there is evidence for φ .

¹⁷ I opted for Definition 3.22 instead of an equivalence between $\Box p$ in B&P and p in FVEL models, because even though we do restrict FVEL models to the single-agent case, these models are still multi-agent in nature. So, while $\mathcal{M}, s \models p$ indicates that *there is evidence for p (at s)*, it is only when $\mathcal{M}, s \models \Box_a p$ holds that we should think that an agent *a* has (knowledge of) this evidence. On the other hand, in single-agent B&P models there is no semantic difference between *there is evidence for p* and *the agent has evidence for p* .

on that all models are single-agent. Much of the conversions between models that follow will be about removing aspects of evidence that are not represented in the other type of model.

3.4.1 From B&P to FVEL models

Consider the following *conversion* from B&P to FVEL models:

Definition 3.23 *Let $M = (S, E, V)$ be a B&P model. Define the FVEL model $FV(M) = (S, R, \mathcal{V})$, where $R = \{(s, s) \mid s \in S\}$ and for all $p \in At$ and states $s \in S$: $1 \in \mathcal{V}(p, s)$ iff $M, s \models \Box p$; and $0 \in \mathcal{V}(p, s)$ iff $M, s \models \Box \neg p$.*

We cannot expect a complete correspondence between M and $FV(M)$ in terms of satisfaction of formulas (in the vein of Proposition 3.37), for while propositional formulas in B&P models represent facts and \Box formulas represent the agent's evidence, in FVEL propositional formulas represent generally available evidence, while \Box formulas represent agents' knowledge of such evidence. This public/personal distinction for evidence in FVEL would be superfluous in B&P models, since they are not multi-agent. Nevertheless, we have the following correspondence:

Proposition 3.24 *For any B&P model $M = (S, E, V)$ and its FVEL counterpart $FV(M)$, for all states $s \in S$ and all literals $l \in \{p, \neg p\}$, with $p \in At$, we have:*

$$M, s \models \Box l \text{ iff } FV(M), s \models l \text{ iff } FV(M), s \models \Box l$$

Proof By the construction of $FV(M)$ we know that $FV(M), s \models p$ iff $M, s \models \Box p$ and the same for $\neg p$. But since R consists exactly of all reflexive arrows, $FV(M), s \models \Box p$ iff $FV(M), s \models p$ (again, the same for $\neg p$). \square

Corollary 3.25 *For any B&P model M , $M \doteq FV(M)$.*

3.4.2 From FVEL to B&P models

This direction is less straightforward than the conversion discussed above. Again we run into the problem of representing a four-valued model as a two-valued one.

Definition 3.26 Let $\mathcal{M} = (S, R, \mathcal{V})$ be an FVEL model. We build a B&P model $\text{BP}(\mathcal{M}) = (S', E, V)$ where $S' = \{s_v \mid s \in S \text{ and } v \in \text{Val}_s^{h_1}\}$ and $s_v \in V(p)$ iff $v(p) = 1$. Let $C(s) = \{t_v \in S' \mid sRt\}$. E is defined as follows:

$$E(s_v) = \{S'\} \cup \{X_p \subseteq C(s) \mid X_p \neq \emptyset, p \in \text{At}; t_u \in X_p \text{ iff } \mathcal{M}, s \models \Box p \text{ and } t_u \in V(p)\} \cup \{X_{\neg p} \subseteq C(s) \mid X_{\neg p} \neq \emptyset, p \in \text{At}; t_u \in X_{\neg p} \text{ iff } \mathcal{M}, s \models \Box \neg p \text{ and } t_u \notin V(p)\}$$

Definition 3.26 creates clusters of states for each original state in \mathcal{M} (similarly to the technique for cluster consolidations). Then, all clusters accessible from a state s_v are grouped together and “filtered” to form the “pieces of evidence” in $E(s_v)$, one for each literal that is known to be evidence in the corresponding state of the FVEL model. For example, if in a state s only evidence for the literal $\neg p$ is known (that is, $\mathcal{M}, s \models \Box \neg p$), then $E(s_v)$ will be $\{S', X_{\neg p}\}$, where $X_{\neg p}$ is a piece of evidence made up of all states accessible from s_v where $\neg p$ holds. See Figure 3.5.

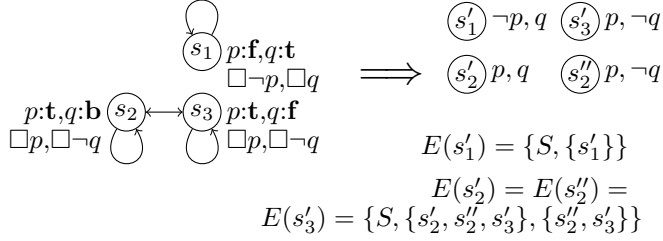


Figure 3.5: An example of BP being applied to an FVEL model.

Proposition 3.27 Let $\mathcal{M} = (S, R, \mathcal{V})$ be a serial FVEL model with $\text{BP}(\mathcal{M}) = (S', E, V)$. Then, for all $s \in S$, all v such that $s_v \in S'$ and all $l \in \{p, \neg p\}$, with $p \in \text{At}$: $\mathcal{M}, s \models \Box l$ iff $\text{BP}(\mathcal{M}), s_v \models \Box l$

Proof Let us first show that $\mathcal{M} \models \Box p$ entails $\text{BP}(\mathcal{M}), s_v \models \Box p$. Let us assume $\mathcal{M} \models \Box p$. We need to show that (i) $\exists X \in E(s_v)$ such that $\forall t \in X$ it holds that $\text{BP}(\mathcal{M}), t \models p$.

S' is not necessarily a piece of evidence matching the X of condition (i), so we have to check whether there is some X_p according to Definition 3.26 respecting those conditions. But X_p can only fail the condition if $\exists t_u \in X_p$ s.t. $\text{BP}(\mathcal{M}), t_u \models \neg p$, which means that $t_u \notin V(p)$ and thus $u(p) = 0$. If X_p is built according to Definition 3.26 this is not possible. So, if we can prove that a non-empty X_p according to Definition 3.26 exists, we are

done. $C(s)$ is empty iff $\nexists t$ s.t. sRt , but since the model is serial this is not possible. So $C(s)$ is non-empty and $\mathcal{M}, s \models \Box p$ is assumed, so we just need to guarantee that there is one $t_u \in C(s)$ s.t. $u(p) = 1$. But since $\mathcal{M}, s \models \Box p$, for all t s.t. sRt we have $\mathcal{M}, t \models p$, which by the definition of S' , V and $C(s)$ will guarantee that for all such t there is at least one u s.t. $t_u \in C(s)$ and $u(p) = 1$. This concludes this direction.

For the other direction, we will prove that $\mathcal{M} \not\models \Box p$ entails $\text{BP}(\mathcal{M}), s_v \not\models \Box p$, which gives us the desired result by *modus tollens*. We assume the former, which entails $\exists t$ s.t. sRt and $\mathcal{M}, t \not\models p$. Now for $\text{BP}(\mathcal{M}), s_v \not\models \Box p$ we just have to show that $\nexists X \in E(s_v)$ s.t. $\forall t_u \in X, \text{BP}(\mathcal{M}), t_u \models p$. We will show that this condition is indeed not satisfied by any $X \in E(s_v)$, for each case of Definition 3.26.

$X = S'$. If $\forall t_u \in S'$ it holds that $\text{BP}(\mathcal{M}), t_u \models p$, then there is no $t_u \in S'$ s.t. $u(p) = 0$. By the definition of S' , this means that in all states w , $\overline{\mathcal{V}}(p, w) = t$. But this contradicts our assumption that $\exists t$ s.t. sRt and $\mathcal{M}, t \not\models p$.

$X = X_p$, where $X_p \subseteq C(s)$ and $t_u \in X_p$ iff $\mathcal{M}, s \models \Box p$ and $u(p) = 1$. Since we are assuming $\mathcal{M}, s \not\models \Box p$, there is no non-empty X_p satisfying these conditions.

$X = X_q$, where $q \neq p$, $X_q \subseteq C(s)$ and $t_u \in X_q$ iff $\mathcal{M}, s \models \Box q$ and $u(q) = 1$. Since $\exists t$ s.t. sRt and $\mathcal{M}, t \not\models p$, then by the definitions of S' , V and $C(s)$ there is a $t_u \in C(s)$ s.t. $u(p) = 0$. Moreover, for any $u \in \text{Val}_t^{h_1}$ s.t. $u(q) = 1$ (as required by any $t_u \in X_q$) there is a $u' \in \text{Val}_t^{h_1}$ s.t. $u'(r) = u(r)$ for all $r \neq p$ and $u'(p) = 0$ – by the combinatorial nature of $\text{Val}_t^{h_1}$. So $\exists t_u \in X_q$ s.t. $\text{BP}(\mathcal{M}), t_u \not\models p$.

$X = X_{\neg p}$, where $X_{\neg p} \subseteq C(s)$ and $t_u \in X_{\neg p}$ iff $\mathcal{M}, s \models \Box \neg p$ and $u(p) = 0$. If $X_{\neg p}$ is non-empty, then by definition $\forall t_u \in X_{\neg p}$ has $\text{BP}(\mathcal{M}), t_u \not\models p$.

$X = X_{\neg q}$, where $q \neq p$, $X_{\neg q} \subseteq C(s)$ and $t_u \in X_{\neg q}$ iff $\mathcal{M}, s \models \Box \neg q$ and $u(q) = 0$. The argument is identical to the $X = X_q$ case.

The cases for $\neg p$ are completely analogous. □

Corollary 3.28 *For all serial FVEL models \mathcal{M} , $\text{BP}(\mathcal{M}) \stackrel{\circ}{=} \mathcal{M}$.*

3.4.3 Evaluating the conversions

Our conversions are satisfactory enough to produce ev-equivalent models, but unfortunately the following proposition can be easily verified:

Proposition 3.29 *Let M be a B&P model and \mathcal{M} be an FVEL model. Then, neither $\text{BP}(\text{FV}(M)) \cong M$ nor $\text{FV}(\text{BP}(\mathcal{M})) \cong \mathcal{M}$ are guaranteed to*

hold; where $M \cong M'$ denote that M is isomorphic to M' , and similarly for $\mathcal{M} \cong \mathcal{M}'$.

One reason why $\text{BP}(\text{FV}(M)) \cong M$ and $\text{FV}(\text{BP}(\mathcal{M})) \cong \mathcal{M}$ do not hold in general is simple: $\text{BP}(\mathcal{M})$ has more states than \mathcal{M} if the latter has any state where some atom has value b or n .

Definition 3.30 Let $M = (S, E, V)$ be a B&P model. We define the following conditions on M :

- **Consistent Evidence (CONS)** $\forall s \in S \forall X, Y \in E(s)$: if $\forall x \in X, M, x \models l$ then $\exists y \in Y, M, y \models l$, for all literals $l \in \{p, \neg p\}$, $p \in \text{At}$;
- **Complete Evidence (COMP)** $\forall s \in S \forall p \in \text{At} \exists X \in E(s)$ s.t. $\forall x \in X, M, x \models p$ or $\forall x \in X, M, x \models \neg p$;
- **Good Evidence (GOOD)** $s \in V(p)$ iff $\exists X \in E(s)$ s.t. $\forall x \in X, M, x \models p$
- **Simple Evidence (SIMP)** $\forall s \in S, E(s) = \{\{s\}, S\}$.

Proposition 3.31 *SIMP entails CONS, COMP and GOOD. CONS and COMP are sufficient and necessary for the preservation of S . CONS, COMP and GOOD are sufficient (but GOOD is not necessary) for preservation of V . SIMP is sufficient and necessary for preservation of E .*

Proof $\text{SIMP} \Rightarrow \text{GOOD}$: easily verifiable.

$\text{SIMP} \Rightarrow \text{COMP}$: easily verifiable (take $\{s\}$ as X).

$\text{SIMP} \Rightarrow \text{CONS}$: easily verifiable (if all states in S support p , then s supports p ; if s supports p , then there is a state in S which supports p : s itself).

For the following proofs we assume a B&P model $M = (S, E, V)$, $\text{FV}(M) = (S, R, \mathcal{V})$ and $\text{BP}(\text{FV}(M)) = (S', E', V')$. Preservation of S , more precisely, means $|S| = |S'|$. Preservation of V means that there is a bijection f from S to S' such that for all $s \in S$ and all $p \in \text{At}$: $s \in V(p)$ iff $f(s) \in V'(p)$. Preservation of E means that there is a bijection f from S to S' such that $\forall X \subseteq S$: $X \in E(s)$ iff $\{f(w) \mid w \in X\} \in E'(f(s))$.

Preservation of S : First, let us show that CONS and COMP imply $|S| = |S'|$, then the converse. By the definition of FV (Definition 3.23), we know that a proposition p in some state s of $\text{FV}(M)$ cannot have value *both* unless $M, s \models \Box p$ and $M, s \models \Box \neg p$. CONS prevents this. For *none*,

$M, s \not\models \Box p$ and $M, s \not\models \Box \neg p$ are needed. COMP prevents this. So CONS and COMP together imply that $\text{FV}(M)$ does not have any atom in any state with value b or n . So by the definitions of S' (in Definition 3.26) and of $\text{Val}_s^{h_1}$ we know that each state will have only one accepted valuation, and therefore $|S| = |S'|$.

Now let us show that $|S| = |S'|$ implies CONS and COMP. If CONS is violated, then for some p, s we have $M, s \models \Box p$ and $M, s \models \Box \neg p$. If COMP is violated, then for some p, s we have $M, s \not\models \Box p$ and $M, s \not\models \Box \neg p$. In either case, $\text{FV}(M)$ will have some proposition with value b or n , which again by Definition 3.26 will imply that $|S'| > |S|$.

Preservation of V : First, let us show that CONS, COMP and GOOD imply that V is preserved. We just showed that CONS and COMP imply $|S| = |S'|$. For all $s \in S$, let $f(s) = s_v$, where $s_v \in S'$. We have to show that for all s, p : $s \in V(p) \Rightarrow s_v \in V'(p)$ and $s_v \in V'(p) \Rightarrow s \in V(p)$.

By GOOD, $s \in V(p)$ implies that $M, s \models \Box p$. This implies that $1 \in \mathcal{V}(p, s)$. CONS implies $M, s \not\models \Box \neg p$, which makes $\mathcal{V}(p, s) = t$. Now there is only one v s.t. $s_v \in S'$, and by the definition of S' we have that $v(p) = 1$, and therefore $s_v \in V'(p)$.

For the other direction, we assume $s_v \in V'(p)$. This implies $v(p) = 1$, but by CONS and COMP we know this v is unique, which means that $\mathcal{V}(p, s) = t$, which is only the case if $M, s \models \Box p$ and $M, s \not\models \Box \neg p$. By GOOD, we derive that $s \in V(p)$.

Now we give a counterexample for why preservation of V does not imply GOOD. Let $S = \{s, t\}$, with $s \in V(p)$ and $t \notin V(p)$, $E(s) = \{\{t\}, S\}$ and $E(t) = \{\{s\}, S\}$ (notice that this violates GOOD). Now, $\text{BP}(\text{FV}(M))$ will have $S' = \{s_v, t_u\}$ for some v, u . If we make $f(s) = t_u$ and $f(t) = s_v$, V is preserved, but GOOD does not hold.

Preservation of E : First, let us show that SIMP implies the preservation of E . Since SIMP entails the other conditions, we know that it also preserves S and V . Let $f(s) = s_v$, where $s_v \in S'$ (this bijection was just shown to preserve V). Given this and $E(s) = \{\{s\}, S\}$ for all s (SIMP), we just need to show that $E'(s_v) = \{\{s_v\}, S'\}$. By CONS and COMP and the definition of S' we have that there is only one valuation compatible with each $s \in S$, and therefore (by the definition of $C(s)$) $E'(s_v) \subseteq \{\{s_v\}, S'\}$. Now $S' \in E'(s, v)$, so we only have to show that $\{s_v\} \in E'(s_v)$. First, note that if $v(p) = 1$ then $\mathcal{M}, s \models \Box p$ (and if $v(p) = 0$ then $\mathcal{M}, s \models \Box \neg p$), by the definition of $\text{Val}_s^{h_1}$ and R . So $\{s_v\}$ will be added either as X_p or $X_{\neg p}$ (definition of $E'(s_v)$ in Definition 3.26).

Now we show that the preservation of E entails SIMP. The preservation

of E entails the existence of a bijection between S and S' , which in turn entails CONS and COMP. The exact same reasoning as in the previous proof can be used to show that $E'(s_v) = \{\{(s_v)\}, S'\}$. Now assume $f(t) = (s_v)$, for some t (this t has to exist as f is a bijection). Then $E'(f(t)) = \{\{f(t)\}, S'\}$. But since E is preserved, $E(t) = \{\{t\}, S\}$, and since t is arbitrary, this just proves SIMP. \square

Corollary 3.32 $\text{BP}(\text{FV}(M)) \cong M$ iff SIMP holds.

Proof Proposition 3.31 just showed that SIMP implies CONS, COMP and GOOD, which in turn imply the preservation of V and S . So SIMP implies the preservation of S , V and E . Moreover, in the proof of Proposition 3.31 we saw that there is a bijection that preserves simultaneously V and E . This guarantees that $M \cong \text{BP}(\text{FV}(M))$. If $M \cong \text{BP}(\text{FV}(M))$ holds, then obviously E is preserved, which in turn implies, by Proposition 3.31, that SIMP holds. (Notice that the importance of this corollary is not just to show that the satisfaction of SIMP is equivalent to the preservation of S , R and V , but to show that this preservation occurs under one and the same bijection.) \square

Definition 3.33 Let $\mathcal{M} = (S, R, \mathcal{V})$ be an FVEL model. We define the following conditions on \mathcal{M} :

- **Classicality (CLAS)** $\forall p \in \text{At}, \forall s \in S : \mathcal{V}(p, s) \in \{t, f\}$;
- **Knowledge of Evidence (KNOW)** $\mathcal{M}, s \models p$ iff $\mathcal{M}, s \models \Box p$;
 $\mathcal{M}, s \models \neg p$ iff $\mathcal{M}, s \models \Box \neg p$;
- **Only-Reflexivity (REFL)** $R = \{(s, s) \mid s \in S\}$

Proposition 3.34 *REFL entails KNOW. CLAS is necessary and sufficient for preservation of S . CLAS and KNOW are sufficient (but KNOW is not necessary) for preservation of \mathcal{V} . CLAS and REFL are the necessary and sufficient conditions for preservation of R .*

Proof Let $\mathcal{M} = (S, R, \mathcal{V})$, $\text{BP}(\mathcal{M}) = (S', E, V)$ and $\text{FV}(\text{BP}(\mathcal{M})) = (S', R', \mathcal{V}')$. That REFL entails KNOW is easy to check.

Preservation of S : If CLAS holds, for each $s \in S$ there will be only one $v \in \text{Val}_s^{h_1}$, so $|S| = |S'|$. Now for the other direction we will assume that CLAS does not hold. Then there is some $p \in \text{At}$ and $s \in S$ s.t.

$\mathcal{V}(p, s) \in \{b, n\}$. In either case, by the definition of S' (Definition 3.26), there will be more than one $v \in Val_s^{h_1}$, and since for any t there is at least one $u \in Val_t^{h_1}$, we have $|S'| > |S|$.

Preservation of \mathcal{V} : By this we mean that there is a bijection f from S to S' s.t. for all $p \in At$ and all $s \in S$: $\mathcal{V}(p, s) = \mathcal{V}'(p, f(s))$.

First we show that CLAS and KNOW imply the preservation of \mathcal{V} . By CLAS we have $|S| = |S'|$, and by the def. of S' (Definition 3.26) there is a unique $v \in Val_s^{h_1}$, so take $f(s) = s_v$ where v is s.t. $s_v \in S'$. By Proposition 3.27 we know that $\mathcal{M}, s \models \Box p$ iff $BP(\mathcal{M}), s_v \models \Box p$. By Definition 3.23 we know that $1 \in \mathcal{V}'(p, s_v)$ iff $BP(\mathcal{M}), s_v \models \Box p$. Thus, $1 \in \mathcal{V}'(p, s_v)$ iff $\mathcal{M}, s \models \Box p$ and, by KNOW, $\mathcal{M}, s \models \Box p$ iff $\mathcal{M}, s \models p$, which boils down to $1 \in \mathcal{V}'(p, s_v)$ iff $1 \in \mathcal{V}(p, s)$. The reasoning for 0 and $\neg p$ is analogous. Since $f(s) = s_v$, \mathcal{V} is preserved.

Now we show that the preservation of \mathcal{V} implies CLAS, but does not imply KNOW. If \mathcal{V} is preserved then there is a bijection between S and S' , therefore S is preserved, which implies CLAS (as we shown above). Now a counterexample of \mathcal{M} where \mathcal{V} is preserved but KNOW does not hold. Let $S = \{s, t\}$, $R = \{(s, t), (t, s)\}$ and $\mathcal{V}(p, s) = t$ and $\mathcal{V}(p, t) = f$. Let $S' = \{s_v, t_u\}$. If we make $f(s) = t_u$ and $f(t) = s_v$, \mathcal{V} is preserved, but KNOW does not hold.

Preservation of R : By this we mean that there is a bijection f from S to S' s.t. sRt iff $f(s)R'f(t)$, for all $s, t \in S$.

First let us show that CLAS and REFL together imply the preservation of R . By CLAS we have $|S| = |S'|$, and by Definition 3.23 we have $R' = \{(s_v, s_v) \mid s_v \in S'\}$. Since REFL means $R = \{(s, s) \mid s \in S\}$ for all $s \in S$, just take $f(s) = s_v$, with v s.t. $s_v \in S'$, for all $s \in S$.

The other direction: since $R' = \{(s', s') \mid s' \in S'\}$, and we have a bijection f between S and S' , we conclude that $s'R't'$ iff $f^{-1}(s')Rf^{-1}(t')$, and therefore $R = \{(s, s) \mid s \in S\}$. \square

Corollary 3.35 $FV(BP(\mathcal{M})) \cong \mathcal{M}$ iff CLAS and REFL hold.

Proof The only thing worth noting here is that CLAS and REFL imply KNOW, and by CLAS and KNOW we have that \mathcal{V} is preserved with the bijection $f(s) = (s, v)$ for v s.t. $(s, v) \in S'$. The same bijection, as shown before, under CLAS and REFL, preserves R . Again, this is to guarantee that these properties not only preserve S , R and \mathcal{V} , but also do so under one and the same bijection. \square

The desired correspondences only hold under fairly strong conditions. These conditions are not arbitrary restrictions, but idealising conditions.¹⁸ This means that B&P and FVEL models have perfectly (ev-)equivalent counterparts under idealised scenarios, where evidence is factive, always present, complete and consistent, and where agents have perfect knowledge of what evidence is available. This correspondence breaks when we deviate from these assumptions to cover situations of imperfect evidence and imperfect knowledge. Now we can compare the two consolidations.

3.5 Comparing Consolidations

In van Benthem, Fernandez-Duque, and Pacuit (2014), a method for obtaining a relation from B&P models is provided:

Definition 3.36 (*van Benthem et al., 2014*) *Given a B&P model $M = (S, E, V)$, define $B_E \subseteq S \times S$ by $sB_E t$ if $t \in \bigcap \mathcal{X}$ for some $\mathcal{X} \in \text{Sce}_E(s)$.*

Consider a monomodal language \mathcal{L}_B with B as its modality.

Proposition 3.37 *Let $M = (S, E, V)$ be a B&P model and $M! = (S, B_E, V)$ its relational counterpart. Then, for all $\varphi \in \mathcal{L}_B$ and $s \in S$: $M, s \models \varphi$ iff $M!, s \models \varphi$.*

Proof The proof will be by induction on the structure of φ . Base: φ atomic; the proposition holds because V is the same for M and $M!$. I.H.: $M \models \varphi'$ iff $M! \models \varphi'$ for φ' subformula of φ . Step: $M, s \models \neg\varphi$ iff $M', s \not\models \varphi$ iff (by I.H.) $M', s \not\models \varphi$ iff $M', s \models \neg\varphi$. $M, s \models \varphi \wedge \psi$ iff ($M, s \models \varphi$ and $M, s \models \psi$) iff (by I.H.) ($M', s \models \varphi$ and $M', s \models \psi$) iff $M', s \models \varphi \wedge \psi$. $M, s \models B\varphi$ iff (for all s -scenarios χ , $\forall t \in \bigcap \chi$: $M, t \models \varphi$) iff (by I.H.) (for all s -scenarios χ , $\forall t \in \bigcap \chi$: $M', t \models \varphi$) iff ($\forall t$ s.t. $sB_E t$: $M', t \models \varphi$) iff $M', s \models B\varphi$. \square

This effectively proves that $M!$ is the consolidation for M found “implicitly” in van Benthem et al. (2014). Now given two models M (B&P) and \mathcal{M} (FVEL) such that $M \doteq \mathcal{M}$, how does $M!$ compare to $\mathcal{M}!$ (\mathcal{M} ’s cautious consolidation)?

¹⁸ S is added in SIMP and in the evidence sets generated by BP just to comply with the last condition of Definition 3.19. If we remove it from both places, Proposition 3.31 still holds.

Definition 3.38 Given $M \doteq \mathcal{M}$ under bijection f , we say that V matches \mathcal{V} iff: for all $p \in At$ and all $s' \in S'$, $\mathcal{V}(p, s') \in \{t, f\}$; and $s \in V(p)$ iff $\mathcal{V}(p, f(s)) = t$.

Proposition 3.39 Let $M \doteq \mathcal{M}$ under bijection f . $M! \cong \mathcal{M}!$ iff: V matches \mathcal{V} , and $f(s)Rf(t)$ iff $t \in \bigcap \mathcal{X}$ for some $\mathcal{X} \in Scc_E(s)$.

Proof Let $M = (S, E, V)$, $\mathcal{M} = (S', R, \mathcal{V})$, $M! = (S, B_E, V)$ and $\mathcal{M}! = (S'', R', V')$.

\Leftarrow : Since V and \mathcal{V} match, \mathcal{V} is classical (that is, it only assigns values t and f), which means that there will be a one-to-one correspondence between states of \mathcal{M} and $\mathcal{M}!$. M and $M!$ already have the same states, so through $M \doteq \mathcal{M}$ we have a correspondence between states of $M!$ and $\mathcal{M}!$. They will also have the same valuation, because the valuations of M and \mathcal{M} match, V is the same for M and $M!$, and by the definition of cautious consolidation $\mathcal{M}!$ will also have the same valuation as $M!$. Now, by assumption, $M!$ and \mathcal{M} have matching valuations, and since \mathcal{V} is classical, by the definition of cautious consolidation we have that R will be identical to R' under the bijection specified earlier, and by assumption R is isomorphic to B_E .

\Rightarrow : Since $M \doteq \mathcal{M}$, these models have the same number of states. The same goes for $M!$, and since $M! \cong \mathcal{M}!$, $\mathcal{M}!$ also has the same number of states. If \mathcal{V} were not classical, $\mathcal{M}!$ would have more states than \mathcal{M} , therefore \mathcal{V} is classical.

Since S is the same for M and $M!$, we can use f to map states of $M!$ into \mathcal{M} . By the definition of cluster consolidation and the fact that \mathcal{V} is classical we conclude that R and R' will be isomorphic, but since $M! \cong \mathcal{M}!$, this implies that R is isomorphic to B_E (in other words: the last condition of this proposition holds).

For each state of \mathcal{M} there is only one accepted valuation, and this valuation is compatible with \mathcal{V} . Since \mathcal{V} is classical, we will have that V' will match it. Now V and V' are isomorphic by assumption, so V and \mathcal{V} will match. \square

So the conditions for consolidations of ev-equivalent B&P and FVEL models to be isomorphic are rather strong: they must have matching valuations and \mathcal{M} 's relation has to mirror B_E .

3.6 Conclusion

We introduced *consolidation* as the process of forming beliefs from a given evidential state. This process can be formally represented by transformations from evidential (FVEL and B&P) models into doxastic Kripke models. We established the grounds for comparison between these different models, and then found the conditions under which their consolidations are isomorphic. Future work can use bisimilarity instead of isomorphism, and extend this methodology to other evidence logics. Would it be possible to define belief without resorting to two-valued Kripke models? Certainly, as all information used in the consolidation is already in the initial evidential models. The rationale here is that, since Kripke models are standard and widely-accepted formal representations of belief, we should be able to represent the beliefs that implicitly exist in evidential models using this tool. We also wanted to highlight the *process* of transforming evidence into beliefs.

The dynamic perspective on consolidations allows us to study, for example, the complexity of these operations, which is important if we are concerned with real agents forming beliefs from imperfect data. It is clear that consolidations of FVEL models tend to be much larger than those of B&P models, but, on the other hand, might be much easier to compute, given that B&P consolidations rely on the hard-to-compute concept of maximally consistent sets. FVEL models can also deal with multiple agents, and accept a function from status of evidence to doxastic attitude as a parameter (in this case, function $h_1 \in H$), allowing for some flexibility in consolidation policies. It would also be interesting to see if a consolidation like B&P's, where maximal consistent evidence sets are taken into account, would be possible in the context of FVEL. Is the converse possible: to apply the idea of H functions in B&P models?

A future extension of this work taking computational costs of consolidations into account would be in line with other work that tries to fight logical omniscience or to model realistic resource-bounded agents (Fagin and Halpern, 1987; Alechina, Logan, and Whitsey, 2004; Balbiani, Fernández-Duque, and Lorini, 2016; Alechina and Logan, 2002; Ågotnes and Alechina, 2007). As mentioned in Section 3.2.1, other *aspects* of evidence can also be considered, such as the *amount* of evidence for or against a certain proposition, the *reliability* of a source or a piece of evidence, etc.

When departing from ev-equivalent FVEL and B&P models, agents form different beliefs. Part of this is explained by the fact that these logics do not represent exactly the same class of evidence situations. But clearly

the consolidation policies also differ. Is one better than the other? At first glance, both seem to be reasonable, but more investigation could be done in this direction.

Moreover, how are changes in an FVEL (or other) evidence model reflected in its consolidation? *Evidence dynamics* for B&P logic are explored in van Benthem and Pacuit (2011b), in line with other dynamic logics of knowledge update and belief revision (van Ditmarsch, van der Hoek, and Kooi, 2007; van Benthem, 2011, 2007; Baltag and Smets, 2006; Plaza, 2007; Gerbrandy, 1999; Rott, 2009; Velázquez-Quesada, 2009).

In the next Chapter, we continue on the topic of consolidations, but within a different interpretation of FVEL.

Chapter 4

Social Consolidations: Evidence and Peerhood

4.1 Introduction

As seen in Chapter 2, *four-valued epistemic logic* (FVEL) was first designed to model scenarios where agents are uncertain about the evidence publicly available. Here we give another interpretation to this logic, where the binary relation represents peerhood connections. Therefore, each state will represent the evidential state of one agent. This puts this work in line with other network logics such as Baltag, Christoff, Rendsvig, and Smets (2019); Christoff and Hansen (2015).

In our setting, agents have four-valued evidence for propositions, embodied by a four-valued valuation function over atoms, which represents only evidence *for* that atom, only evidence *against* it, evidence *both for and against* it, or *no* evidence at all. Our main goal in this chapter is to find rational ways of forming beliefs for these agents, given their own evidence and their peers'. With that in mind, we establish some rationality postulates and check some definitions of belief that respect those postulates, and some that do not.

After that, we introduce a dynamic operator for addition/removal of evidence. This operator is used to axiomatise some of the postulates, but also to define two new ones, which serve to rule out some undesirable consolidations. We then prove that these axioms characterise a class of consolidations satisfying most of the main postulates. Finally, we show how this operator can be used to “count” peers, which in the future can be employed to define consolidations that form beliefs based on the amount

of evidence for or against something.

4.2 Syntax and Semantics

In this section we explore a variant of the *four-valued epistemic logic* (FVEL) of Chapter 2.

4.2.1 Syntax

Let At be a countable set of atoms. Below, $p \in At$; the classical part of the language is given by \mathcal{L}_0 ; the propositional part is given by \mathcal{L}_1 ; and the complete language is given by \mathcal{L} :

$$\begin{aligned}\mathcal{L}_0 \quad \psi &::= p \mid \sim \psi \mid (\psi \wedge \psi) \\ \mathcal{L}_1 \quad \chi &::= \psi \mid \sim \chi \mid (\chi \wedge \chi) \mid \neg \chi \\ \mathcal{L} \quad \varphi &::= \chi \mid \sim \varphi \mid (\varphi \wedge \varphi) \mid \Box \varphi \mid B\psi\end{aligned}$$

We abbreviate $\varphi \vee \psi \stackrel{\text{def}}{=} \sim(\sim\varphi \wedge \sim\psi)$ and $\Diamond\varphi \stackrel{\text{def}}{=} \sim\Box\sim\varphi$. We restrict belief to classical propositional formulas (\mathcal{L}_0) because formulas with \neg refer to evidence, and we do not want agents forming beliefs about evidence, only about facts.

Formulas such as p are read as *the agent has evidence for p* , whereas $\neg p$ is read as *the agent has evidence against p* , and $\sim\varphi$ as *it is not the case that φ* . We read $\Box\varphi$ as *φ holds for all peers* and $B\varphi$ as *the agent believes φ* .^{1 2}

4.2.2 Semantics

Models are tuples $\mathcal{M} = (S, R, \mathcal{V})$, where S is a finite set of agents, R is a binary relation on S representing “peerhood” and $\mathcal{V} : At \times S \rightarrow \mathcal{P}(\{0, 1\})$ is a four-valued valuation representing agents’ evidence: $\{1\}$ is *true* (t), $\{0\}$ is *false* (f), $\{0, 1\}$ is *both* (b) and \emptyset is *none* (n). A satisfaction relation

¹ Notice that our language is non-standard in that even though a formula in \mathcal{L}_1 has an evidential meaning (such as p meaning *the agent has evidence for p*), under the belief operator B these formulas are read as factual statements (e.g. Bp means that *the agent believes p* and not that *the agent believes that she has evidence for p*).

² We chose B (belief) instead of K (knowledge) because we are working with imperfect evidence, which can be misleading. Therefore, our agents can form false beliefs, which violate factivity, a standard requirement for knowledge.

is defined as follows:

$$\begin{aligned}
\mathcal{M}, s \models p &\text{ iff } 1 \in \mathcal{V}(p, s) & \mathcal{M}, s \models \neg p &\text{ iff } 0 \in \mathcal{V}(p, s) \\
\mathcal{M}, s \models \sim \varphi & & \text{ iff } \mathcal{M}, s \not\models \varphi \\
\mathcal{M}, s \models (\varphi \wedge \psi) & & \text{ iff } \mathcal{M}, s \models \varphi \text{ and } \mathcal{M}, s \models \psi \\
\mathcal{M}, s \models \neg(\varphi \wedge \psi) & & \text{ iff } \mathcal{M}, s \models \neg\varphi \text{ or } \mathcal{M}, s \models \neg\psi \\
\mathcal{M}, s \models \Box\varphi & & \text{ iff for all } t \in S \text{ s.t. } sRt, \text{ it holds that } \mathcal{M}, t \models \varphi \\
\mathcal{M}, s \models \neg\sim\varphi &\text{ iff } \mathcal{M}, s \models \varphi & \mathcal{M}, s \models \neg\neg\varphi &\text{ iff } \mathcal{M}, s \models \varphi
\end{aligned}$$

An extended valuation function $\overline{\mathcal{V}}$ can be defined differently for each type of formula. If $\varphi \in \mathcal{L}_1$, then: $1 \in \overline{\mathcal{V}}(\varphi, s)$ iff $\mathcal{M}, s \models \varphi$; $0 \in \overline{\mathcal{V}}(\varphi, s)$ iff $\mathcal{M}, s \models \neg\varphi$. Otherwise: $1 \in \overline{\mathcal{V}}(\varphi, s)$ iff $\mathcal{M}, s \models \varphi$ iff $0 \notin \overline{\mathcal{V}}(\varphi, s)$. As pointed out earlier, this logic can be seen as a modal extension of FDE (Belnap, 1977), with the addition of a classical negation. The logic FDE deals with evidence differently than other logics such as intuitionistic logic (Heyting, 1966; Troelstra and van Dalen, 1988). While both are weaker than classical logic, the concept of justification as existence of constructive proofs is much stronger than what we consider evidence in this thesis. In our case, evidence can be misleading, as mentioned before. FDE is more suitable for modelling situations with incomplete and inconsistent evidence, while FVEL extends this logic to a modal setting, enabling us to talk about multiple agents. Again, we are going to make use of the abbreviations φ^n , φ^f , φ^t and φ^b discriminating which of the four truth values a formula $\varphi \in \mathcal{L}_1$ has, as defined in Chapter 2 (page 19). We say that $\Sigma \models \varphi$ (Σ entails φ) when for all models \mathcal{M} and states s , if $\mathcal{M}, s \models \sigma$ for all $\sigma \in \Sigma$, then $\mathcal{M}, s \models \varphi$. We say that $\mathcal{M} \models \varphi$ if $\mathcal{M}, s \models \varphi$ for all states s of \mathcal{M} . And $\models \varphi$ (φ is *valid*) if $\mathcal{M} \models \varphi$ for all \mathcal{M} ; otherwise φ is *invalid*. If $\models \sim\varphi$, we say φ is *contradictory*, and if φ is neither contradictory nor valid, it is *contingent*. If a formula is valid or contingent, it is *satisfiable*. Call the *truth range* of φ the set $\{x \mid \text{there is a model } \mathcal{M} = (S, R, \mathcal{V}) \text{ and an } s \in S \text{ s.t. } \overline{\mathcal{V}}(\varphi, s) = x\}$. The following result will be useful for some of the proofs (and also applies to FVEL as defined in the previous chapters):

Proposition 4.1 *All formulas in \mathcal{L}_0 have one of the following four truth ranges: $\{\{1\}\}$, $\{\{0\}\}$, $\{\{0\}, \{1\}\}$, $\{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$. A formula in \mathcal{L}_1 can have any truth range in $\mathcal{P}(\mathcal{P}(\{0, 1\})) \setminus \emptyset$ except for $\{\emptyset\}$, $\{\{0, 1\}\}$, and $\{\emptyset, \{0, 1\}\}$.*

Proof (\mathcal{L}_0) This can be proved easily by induction, consulting the truth tables from Section 2.2.3. Base case: atoms. Clearly all atoms have truth

range $\{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$. I.H.: for any φ' that is a proper subformula of φ , the proposition holds. Step: $\varphi = \sim\psi$. Clearly any formula of this format can only have truth values $\{0\}$ or $\{1\}$, and therefore satisfy the proposition. $\varphi = \psi \wedge \chi$. This case is tedious but easy. We just have to check what are the possible truth values for $\psi \wedge \chi$ given each truth range for ψ and χ . By the I.H., ψ and χ have one of the truth ranges listed in the proposition. As an example, let us check the case for when ψ and χ have truth ranges $\{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$ and $\{\{0\}, \{1\}\}$, respectively. Well, in that case the possible truth values for $\psi \wedge \chi$ are the values in the truth table when we restrict one of the parameters to $\{0\}$ and $\{1\}$, which gives us the truth range $\{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$. If we do the same considering each of the other truth ranges listed in the proposition for ψ and χ , we conclude that all possible truth ranges for $\psi \wedge \chi$ are within the ones listed in the proposition statement.

(\mathcal{L}_1) If a valuation assigns no value \emptyset or $\{0, 1\}$ to any atom, then all formulas have “classical” truth values ($\{0\}$ or $\{1\}$), so it is not possible to have formulas with the truth ranges mentioned in the statement of the proposition. To show that the other truth ranges are possible, we give examples, followed by their truth ranges: $\sim(p \wedge \sim p)$: $\{\{1\}\}$. $\sim\sim(p \wedge \sim p)$: $\{\{0\}\}$. $\sim p$: $\{\{0\}, \{1\}\}$. p : $\{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$. $p \wedge \sim p$: $\{\emptyset, \{0\}\}$. $\neg(p \wedge \sim p)$: $\{\emptyset, \{1\}\}$. $p \wedge \neg p$: $\{\emptyset, \{0\}, \{0, 1\}\}$. $\neg(p \wedge \neg p)$: $\{\emptyset, \{1\}, \{0, 1\}\}$. $p \wedge \neg p \wedge \sim(p^n)$: $\{\{0\}, \{0, 1\}\}$. $\neg(p \wedge \neg p \wedge \sim(p^n))$: $\{\{1\}, \{0, 1\}\}$. $p \wedge \sim(p^n)$: $\{\{0\}, \{1\}, \{0, 1\}\}$. $p \wedge \sim(p^b)$: $\{\emptyset, \{0\}, \{1\}\}$. \square

The central question of this chapter is how to define the semantics for belief based on the evidence, a process we call *consolidation* (see Chapter 3). A key philosophical assumption of this project is that *rational belief is determined by evidence*.

4.3 Rationality Conditions for Consolidations

In this section we discuss the guiding principles and conditions that consolidations should respect.

4.3.1 Epistemic Autonomy versus Epistemic Authority

In social epistemology, there is currently a lot of debate around the topics of peer disagreement and higher-order evidence (Christensen, 2010; Lasonen-Aarnio, 2014; Fricker, 2006; Lehrer, 1977; Martini, Sprenger, and Colyvan, 2013; Hardwig, 1985; Foley, 2001). One important question in this debate

is: *What should a rational agent do when her peers – who she deems as rational as her – have different opinions on some proposition?* There are many different proposals in the literature as to what to do in this case. Nevertheless, we can roughly categorise them into two main groups: the *equal weight* views (Elga, 2007), and the *steadfast* views (Kelly, 2010). The former tend to consider the agent and her peers to be on equal footing, so if you and your peer disagree on something, your opinion should be something in the middle of both opinions. The latter claim that you are entitled to trust yourself more than you trust your peers – maybe because you have direct access to your evidence, as opposed to mere testimonial access to your peers’ evidence, or because of some other reason. In both views, the concept of *peerhood* is preeminent. It is assumed that, in what matters, you and your peers are of equal competence. Evidently, if one’s peer is far more competent than oneself in the topic at hand and one knows that, the rational thing to do is to defer to her judgement (but in that case she is *not* your peer). What enables peerhood is the lack of such higher-order knowledge: we usually do not know exactly how competent a peer is, so the reasonable (and modest) thing to do is to assume that the relevant people in the given case are (possibly) as competent as you, except if you have a “defeater” for that belief.³

4.3.2 Rationality Postulates

Now we propose and discuss a series of rationality postulates, mostly adapted from postulates from Social Choice Theory (SCT) (Arrow, 1951; Gibbard, 1973; Satterthwaite, 1975).⁴ SCT is concerned with determining outcomes of voting from certain voting profiles. The adaptation we make here is in the sense that a rational belief in propositions (atomic or otherwise) will be determined from the evidence possessed by the agent and her peers, so here “voting profiles” become evidence, and “election outcome” becomes belief attitude. Consolidations are not voting procedures, but involve the weighing of inputs to find a suitable outcome.

Condition 4.2 (Consistency) *For all models \mathcal{M} and $s \in S$: let $\Sigma = \{\varphi \in \mathcal{L}_0 \mid \mathcal{M}, s \models B\varphi\}$. Then $\Sigma \not\models p \wedge \sim p$.*

³ As a scientist investigating hypothesis H, you consider another scientist also investigating H to be your peer, but not if she committed fraud in the past.

⁴ Note, however, that we only make a loose connection to SCT here, not a formal one.

The condition above is the most important demand on our consolidations: rational belief has to be consistent.

Regardless of the semantics of B , which is not yet defined, the following function Att serves as a shorthand for the doxastic attitude of an agent s with respect to a formula φ (belief, disbelief or abstention):

Definition 4.3 (Attitude) *Let $\text{Att} : \mathcal{L}_0 \times S \rightarrow \{1, 0, -1\}$ be a function such that:*

- $\text{Att}(\varphi, s) = 1$ iff $\mathcal{M}, s \models B\varphi$;
- $\text{Att}(\varphi, s) = -1$ iff $\mathcal{M}, s \models B\sim\varphi$;
- otherwise, $\text{Att}(\varphi, s) = 0$.

(The function Att also depends on a model \mathcal{M} , but this will be left implicit. We will usually write Att' if we are referring to another model \mathcal{M}' , Att'' for \mathcal{M}'' , and so on. Notice also that this function is only well-defined in the context of consolidations that satisfy Consistency.)⁵

Condition 4.4 (Modesty) *For all models $\mathcal{M} = (S, R, \mathcal{V})$, all $s \in S$, and all contingent $\varphi \in \mathcal{L}_0$, there is a model $\mathcal{M}' = (S', R', \mathcal{V}')$ with $S \subseteq S'$ s.t. $\text{Att}(\varphi, s) \neq \text{Att}'(\varphi, s)$, where $\mathcal{V}|_s = \mathcal{V}'|_s$.*⁶

Condition 4.4 says that it is possible to change an agent's attitude toward a contingent formula just by changing her peerhood connections and the evidence of her peers. Modesty is adapted from the SCT postulate of *non-dictatorship*: the outcome of the election is not determined by one single agent. Condition 4.5 also comes from non-dictatorship, but for *Modesty* we think of the agent as her own dictator.

The plausibility of this postulate hinges on the plausibility of the claim that regardless of what evidence you have, it is not always rational to ignore others' evidence. This, in turn, depends on the outcome of the debate in epistemology discussed above. In any case, is the format of this postulate adequate? The restriction to contingent formulas seems justified: if we reject *Logical Omniscience*, it might be acceptable to abstain from judgement on tautologies and contradictions, but it seems irrational to expect one to be persuaded to abandon a belief in a tautology or adhere

⁵ Note that in the previous chapter we used 0 for disbelief and -1 for abstention, whereas here (and in the next chapter), just for technical convenience, we swap these values.

⁶ We denote by $\mathcal{V}|_s$ the restriction of a valuation \mathcal{V} to $At \times \{s\}$, with $s \in S$.

to a contradiction. Keeping $\mathcal{V}|_s$ untouched captures exactly the idea of not changing one's evidence, but possibly changing others'. The $S \subseteq S'$ part demands that the original agents be preserved. This is innocuous, for even if a change in belief demands the removal of a peer, that can be obtained by removing the connection (changing R); non-peers do not matter in our setting. A stronger variant of Modesty could be considered, *Strong Modesty*, where not only is it possible to change the attitude for any formula, but also any other attitude is possible. This could be plausible, but expecting a radical change in attitude (for example, from disbelief to belief) for any contingent proposition might require a huge *amount* of evidence, and we are not representing this *aspect* of evidence here; we do make a step in this direction in Section 4.5.

Condition 4.5 (No Gurus) *For all agents $s, t \in S$ (with $s \neq t$) and all contingent $\varphi \in \mathcal{L}_0$, there is a model $\mathcal{M} = (S, R, \mathcal{V})$ s.t. $\text{Att}(\varphi, s) \neq \text{Att}(\varphi, t)$.*

This condition says that for any formula there is a model such that the attitudes of two agents towards that formula differ, i.e., an agent's opinion is not determined by anyone else's. This postulate also stems from the postulate of non-dictatorship in SCT (in a more obvious way). We have the following:

Proposition 4.6 *A consolidation (see Definition 4.18) satisfying Modesty also satisfies No Gurus.*

Proof Take a contingent $\varphi \in \mathcal{L}_0$ and two agents $s, t \in S$, $s \neq t$, and a model $\mathcal{M} = (S, R, \mathcal{V})$. If $\text{Att}(\varphi, s) \neq \text{Att}(\varphi, t)$, then we are done, do let us assume that $\text{Att}(\varphi, s) = \text{Att}(\varphi, t)$. Also, assume that neither sRt nor tRs hold. By Modesty, there is a $\mathcal{M}' = (S', R', \mathcal{V}')$ with $S \subseteq S'$ such that $\text{Att}'(\varphi, s) \neq \text{Att}(\varphi, s)$. Now notice that we can build a \mathcal{M}'' by adding to \mathcal{M}' an isomorphic copy of \mathcal{M} (with fresh agent labels, say from s to s^*). Now we can exchange the agent label of t (which was already in \mathcal{M}') with the relabelled t^* (that came from the copy of \mathcal{M}). In this way, $(\mathcal{M}'', s) \simeq (\mathcal{M}', s)$ and $(\mathcal{M}'', t) \simeq (\mathcal{M}, t)$. By Proposition 4.16 (see later), $\text{Att}''(\varphi, s) = \text{Att}'(\varphi, s)$ and $\text{Att}''(\varphi, t) = \text{Att}(\varphi, t)$, and therefore $\text{Att}''(\varphi, s) \neq \text{Att}(\varphi, t)$. \square

So if Modesty is plausible, then this postulate has to plausible be as well. In principle, it might be odd to think that, for example, two biologists could rationally disagree on whether natural selection happens. This apparent

controversy is only superficial, though. If we stick to our key assumption that evidence determines rational belief, then that should be possible *given* they have access to different circles – with one of them possibly possessing misleading evidence.

Condition 4.7 (Equal Weight) *Consider any model $\mathcal{M} = (S, R, \mathcal{V})$, any two agents $s, t \in S$, and a valuation \mathcal{V}' such that $\mathcal{V}'(p, s) = \mathcal{V}(p, t)$, $\mathcal{V}'(p, t) = \mathcal{V}(p, s)$, and $\mathcal{V}'(p, u) = \mathcal{V}(p, u)$ for all $u \in S \setminus \{s, t\}$, for all $p \in At$. Then, if sRt it holds that, for all $\varphi \in \mathcal{L}_0$, $\text{Att}'(\varphi, s) = \text{Att}(\varphi, s)$.*

What this postulate says is that if you swap all your evidence with the evidence of one of your peers, your beliefs do not change: you treat your evidence and your peers' equally. It comes from the SCT postulate of *anonymity*: if we have the same voting profile but swap the voters, the outcome does not change. Again, the plausibility of this postulate depends on your position in the debate of Section 4.3.1.

Condition 4.8 (Atom Independence) *Consider any model $\mathcal{M} = (S, R, \mathcal{V})$. For any atom $p \in At$, if \mathcal{V}' is a valuation s.t. $\mathcal{V}'(p, s) = \mathcal{V}(p, s)$ for all $s \in S$, then $\text{Att}(p, s) = \text{Att}'(p, s)$ for all $s \in S$.*

The valuation of one atom should not interfere in the attitudes towards another. This postulate is adapted from *independence of irrelevant alternatives*: the outcome between x and y should only depend on voters opinions with respect to x and y ; changing the preferences between other candidates does not affect the outcome.

Let \preceq be the smallest reflexive and transitive relation $\preceq: \mathcal{P}(\{0, 1\}) \times \mathcal{P}(\{0, 1\})$ such that $\{0\} \preceq \emptyset$, $\{0\} \preceq \{0, 1\}$, $\emptyset \preceq \{1\}$ and $\{0, 1\} \preceq \{1\}$. Let \nsubseteq be the complement of \preceq , and define $x \prec y$ iff $x \preceq y$ and $y \nsubseteq x$.

Condition 4.9 (Monotonicity) *Consider a model $\mathcal{M} = (S, R, \mathcal{V})$ and a \mathcal{V}' which coincides with \mathcal{V} , except that $\mathcal{V}'(p, s) \neq \mathcal{V}(p, s)$ for one $s \in S$ and $p \in At$. If $\mathcal{V}(p, s) \prec \mathcal{V}'(p, s)$, then for all $t \in S$, $\text{Att}(p, t) \leq \text{Att}'(p, t)$. If $\mathcal{V}'(p, s) \prec \mathcal{V}(p, s)$, then for all $t \in S$, $\text{Att}'(p, t) \leq \text{Att}(p, t)$.*

Condition 4.9 states that if the valuation only changes positively/negatively for one atom and one agent, then the attitude towards this atom for any agent should either stay the same, or change according to the same trend (more positive/negative). Monotonicity was adapted from a homonymous SCT postulate: if a profile is altered only by promoting (demoting) one

candidate, the outcome should either change only by promoting (demoting) this candidate, or not change.

Now there is a question of adequacy of the format of this postulate. There is not always a unique way of changing a valuation to produce a certain change in the (extended) valuation of a complex formula, so we limited this postulate to atomic changes. The other question regarding format is why the postulate limits the valuation change to only one atom and one agent. Clearly changing one atom in one direction (according to \prec) for more agents, or changing several atoms in this fashion, should preserve monotonicity conditions. These “cumulative” effects are already covered by the postulate as it is.

Condition 4.10 (Doxastic Freedom) *Consider any set of agents S and any function $f : At \times S \rightarrow \{1, -1, 0\}$. Then there is a model $\mathcal{M} = (S, R, \mathcal{V})$ such that $\text{Att}(p, s) = f(p, s)$ for all $p \in At$ and $s \in S$.*

Doxastic Freedom says that any combination of attitudes towards atoms is possible for any agent. It is adapted from *non-imposition*: every outcome is achievable by some voting profile. This postulate seems somehow connected to Atom Independence. However:

Observation 4.11 *A consolidation satisfying Atom Independence does not necessarily satisfy Doxastic Freedom. A consolidation satisfying Doxastic Freedom does not necessarily satisfy Atom Independence.*

Proof Naive consolidation (defined later in Section 4.4.2) satisfies Atom Independence but violates Doxastic Freedom. For the other direction, consider a set of atoms $At = \{p_1, p_2, \dots\}$, and a consolidation similar to Policy V (also defined later, Section 4.5), but which instead of deciding Bp_i based on p_i , does the following: if At is infinite, decides p_i based on p_{i+1} for odd i , and based on p_{i-1} for even i ; if $At = \{p_1, \dots, p_n\}$ is finite, decides belief in p_i based on p_{i+1} , except for p_n , which is decided based on p_1 . Policy V and this modification satisfy Doxastic Freedom, but this modification does not satisfy Atom Independence (and therefore is not a \mathcal{C} -consolidation – as defined later in Definition 4.25). \square

Condition 4.12 (Consensus) *If for some agent $s \in S$ and some $\varphi \in \mathcal{L}_0$ we have that $\overline{\mathcal{V}}(\varphi, s) = \{1\}$ (or $\{0\}$), and for all $t \in S$ such that sRt : $\overline{\mathcal{V}}(\varphi, t) = \{1\}$ (or $\{0\}$), then $\text{Att}(\varphi, s) \neq -1$ (or 1).*

Consensus is derived from the SCT postulate of *unanimity*: if all voters prefer one candidate over another, then so must the outcome. It says that if an agent and all her peers have unambiguous evidence about some atom, then she should not believe contrary to that. We can define *Strong Consensus* in a similar way, but instead of demanding no contrary belief, it demands belief in case of unanimous positive evidence and disbelief in case of unanimous negative evidence.

Observation 4.13 *A consolidation satisfying Strong Consensus and Consistency also satisfies Consensus.*

Proof One just has to see that $\mathcal{M}, s \models B\varphi$ implies $\mathcal{M}, s \not\models B\sim\varphi$ for a consolidations satisfying Consistency (and similarly for the $B\sim\varphi$ case). \square

Notice that this stronger variant, in combination with Proposition 4.1, entails a form of *logical omniscience*. We could also have defined the postulate differently by considering unanimity among all agents instead of one agent and her peers, but, again, we are assuming that non-peers are inaccessible/irrelevant.

Condition 4.14 (Logical Omniscience) *For all models \mathcal{M} and $s \in S$: if $\Sigma \models \varphi$ and $\mathcal{M}, s \models B\sigma$ for all $\sigma \in \Sigma$, then $\mathcal{M}, s \models B\varphi$.*

This postulate is not derived from any postulate of SCT. It is debatable whether it should be satisfied or not, but as a normative demand on real agents we consider it too strong. Notice that it implies the knowledge of all validities, as they are consequences of the empty set, and also that the doxastic state has to be consistent or it will be trivialised.

In summary, all the postulates listed in this section are expected to be satisfied by any rational consolidation (call these *core postulates*), except for Modesty and Equal Weight, whose normative status depend on the reader's philosophical commitments with respect to the debate of Section 4.3.1, and Logical Omniscience, which is also part of another long debate (Hintikka, 1962, 1979; Rantala, 1975; Fagin and Halpern, 1987). No impossibility theorem à la Arrow (1951) ensues, and consolidations satisfying all core postulates are presented. One main difference of our approach that might explain this is that we do not have preference orders over attitudes. Note also that our connection to SCT is not fully formal: our postulates are only inspired by it.

4.4 Social Consolidations

In this section we will define *consolidation policies*, that is, methods of defining belief from evidence. We expect the most reasonable consolidations to satisfy all the core postulates, and unreasonable ones to violate at least one of them.

4.4.1 Preliminaries

Before talking about consolidations, we will formally specify what are the possible ones. Now let $\mathbf{M} = \{(\mathcal{M}, s) \mid \mathcal{M} = (S, R, \mathcal{V}) \text{ is an FVEL model and } s \in S\}$ be the class of all *pointed models*. First, we draw the following definition from the literature on n-bisimulations:

Definition 4.15 (1-Bisimulation) *Consider two FVEL models $\mathcal{M} = (S, R, \mathcal{V})$ and $\mathcal{M}' = (S', R', \mathcal{V}')$, an $s \in S$ and an $s' \in S'$. We say that $(\mathcal{M}, s) \rightleftharpoons (\mathcal{M}', s')$, read (\mathcal{M}, s) is 1-bisimilar to (\mathcal{M}', s') , iff:*

atoms *For all $p \in \text{At}$, $\mathcal{V}(p, s) = \mathcal{V}'(p, s')$;*

back *For all $t' \in S'$ s.t. $s'R't'$, there is a $t \in S$ s.t. sRt and $\mathcal{V}(p, t) = \mathcal{V}'(p, t')$ for all $p \in \text{At}$.*

forth *For all $t \in S$ s.t. sRt , there is a $t' \in S'$ s.t. $s'R't'$ and $\mathcal{V}(p, t) = \mathcal{V}'(p, t')$ for all $p \in \text{At}$.*

The purpose of Definition 4.15 is to determine whether two pointed models have equivalent evidence. Since our relation R of peerhood is not transitive, we assume that our agents only have access to their own evidence and their peers'. So formulas such as $\Box p$ are relevant for consolidation, whereas $\Box\Box p$ is not.

Proposition 4.16 *$(\mathcal{M}, s) \rightleftharpoons (\mathcal{M}', s')$ implies: $\mathcal{M}, s \models \varphi$ iff $\mathcal{M}', s' \models \varphi$ for all $\varphi \in \mathcal{L}$ containing neither B nor nested \Box . The converse also holds for image-finite models (each agent has finitely many peers).*

Proof The first direction is easy to prove by induction on the structure of φ . Base: it is immediately evident (by Definition 4.15) that if $(\mathcal{M}, s) \rightleftharpoons (\mathcal{M}', s')$ then $\mathcal{M}, s \models \varphi$ iff $\mathcal{M}', s' \models \varphi$, for all $\varphi \in \mathcal{L}_1$. I.H.: For all φ' proper subformula of φ , $(\mathcal{M}, s) \rightleftharpoons (\mathcal{M}', s')$ implies $\mathcal{M}, s \models \varphi'$ iff $\mathcal{M}', s' \models \varphi'$. Step: $\varphi = \sim\psi$. By I.H. $\mathcal{M}, s \models \psi$ iff $\mathcal{M}', s' \models \psi$, but then $\mathcal{M}, s \not\models \psi$ iff $\mathcal{M}', s' \not\models \psi$. $\varphi = \psi \wedge \chi$. By I.H., $\mathcal{M}, s \models \psi$ iff $\mathcal{M}', s' \models \psi$ and $\mathcal{M}, s \models \chi$ iff $\mathcal{M}', s' \models \chi$. Then, $\mathcal{M}, s \models \psi \wedge \chi$ iff $\mathcal{M}', s' \models \psi \wedge \chi$. $\varphi = \Box\psi$, where ψ has no \Box nor B . Since $(\mathcal{M}, s) \rightleftharpoons (\mathcal{M}', s')$, for all t

such that sRt there is a t' such that $s'R't'$ with $\mathcal{V}(p, t) = \mathcal{V}'(p, t')$ for all $p \in At$. Then, since ψ has no \Box nor B , $\psi \in \mathcal{L}_1$, and for any $\psi' \in \mathcal{L}_1$ and t such that sRt , $\mathcal{M}, t \models \psi'$ implies that there exists a t' such that $s'R't'$ and $\mathcal{M}', t' \models \psi'$. The other direction follows by **back**.

Now for the other direction (the second part of the proposition). First, from $\mathcal{M}, s \models \varphi$ iff $\mathcal{M}', s' \models \varphi$ for $\varphi \in \mathcal{L}$ not containing B nor nested \Box , we can easily see that **atoms** holds. The argument for **back** and **forth** are analogous, so we just show **forth** here. Consider a t such that sRt , and consider the set $\Sigma = \{p^x \mid p \in At \text{ and } \mathcal{M}, t \models p^x, \text{ where } x \in \mathcal{P}(\{0, 1\})\}$. We want to show that there is a t' such that $s'R't'$ and $\mathcal{V}(p, t) = \mathcal{V}'(p, t')$ for all $p \in At$. For any finite conjunction γ of elements of Σ , we have $\mathcal{M}, t \models \gamma$ and therefore $\mathcal{M}, s \models \Diamond\gamma$. But then $\mathcal{M}', s' \models \Diamond\gamma$, as $\gamma \in \mathcal{L}_1$. This implies that every finite conjunction γ of elements of Σ are satisfied in some successor of s' . Assume, then, that no successor of s' satisfies all elements of Σ . Then, for each such successor t'_i there is a $p_i^{x_i} \in \Sigma$ such that $\mathcal{M}', t'_i \not\models p_i^{x_i}$. But then, the finite conjunction $p_1^{x_1} \wedge p_2^{x_2} \wedge \dots \wedge p_n^{x_n}$, where $s'R't'_1, \dots, s'R't'_n$, is not satisfied in any successor of s' . Contradiction. So Σ is satisfied in some successor of s' and therefore **forth** holds. \square

Proposition 4.17 *The relation \equiv is an equivalence relation.*

Proof That reflexivity and symmetry are satisfied is trivial. One just have to check whether \equiv is also transitive, which can be done straightforwardly by checking Definition 4.15 in the case where $(\mathcal{M}, s) \equiv (\mathcal{M}', s')$ and $(\mathcal{M}', s') \equiv (\mathcal{M}'', s'')$, to derive $(\mathcal{M}, s) \equiv (\mathcal{M}'', s'')$. \square

Then $\equiv \subseteq \mathbb{M} \times \mathbb{M}$. Denote by $[\mathcal{M}, s]$ the equivalence class of (\mathcal{M}, s) under \equiv , that is, $[\mathcal{M}, s] = \{(\mathcal{M}', s') \in \mathbb{M} \mid (\mathcal{M}, s) \equiv (\mathcal{M}', s')\}$. Let \mathbb{M}/\equiv be the quotient class of \mathbb{M} by \equiv , that is, the class of equivalence classes of \mathbb{M} under \equiv . Then, we are interested in the following:

Definition 4.18 *A consolidation is a function $\mathbb{C} : \mathbb{M}/\equiv \times \mathcal{L}_0 \rightarrow \{0, 1\}$. For any model $\mathcal{M} = (S, R, \mathcal{V})$ with $s \in S$, we set $\mathcal{M}, s \models B\varphi$ iff $\mathbb{C}([\mathcal{M}, s], \varphi) = 1$.*

With these definitions in hand, we will introduce the following:

Definition 4.19 *We say that a condition is axiomatisable when: it holds iff all $\sigma \in \Sigma$ are valid, for some $\Sigma \subseteq \mathcal{L}$. We say that a condition is*

negatively axiomatisable *when: it holds iff all $\sigma \in \Sigma$ are invalid, for some $\Sigma \subseteq \mathcal{L}$.*⁷

Proposition 4.20 *Consistency holds iff for all finite $\Sigma = \{\sigma_1, \dots, \sigma_n\} \subseteq \mathcal{L}_0$ such that $\Sigma \models p \wedge \sim p$, $\sim(B\sigma_1 \wedge \dots \wedge B\sigma_n)$ is valid.*

Proof The logic of \mathcal{L}_0 is basically classical propositional logic (as mentioned in Chapter 2), and is, therefore, compact. So for any $\Sigma \models \varphi$ with $\varphi \in \mathcal{L}_0$, there is a finite $\Sigma' \subseteq \Sigma$ such that $\Sigma' \models \varphi$. The case where $\varphi = p \wedge \sim p$ is a particular case of this. So all inconsistent subsets of \mathcal{L}_0 have a finite inconsistent subset. \square

Proposition 4.21 *Logical Omniscience holds iff for all finite $\Sigma = \{\sigma_1, \dots, \sigma_n\} \subseteq \mathcal{L}_0$ and $\varphi \in \mathcal{L}_0$ such that $\Sigma \models \varphi$, $\sim(B\sigma_1 \wedge \dots \wedge B\sigma_n \wedge \sim B\varphi)$ is valid.*

Proof The reasoning is similar to the case for Proposition 4.20. \square

Note that Propositions 4.20 and 4.21 follow from compactness of \mathcal{L}_0 . Now consider the following axioms:

$$\mathbf{C1} \quad \sim((\varphi^t \wedge \Box \varphi^t) \wedge B\sim \varphi)$$

$$\mathbf{C2} \quad \sim((\varphi^f \wedge \Box \varphi^f) \wedge B\varphi)$$

Proposition 4.22 *A consolidation satisfying Consistency satisfies Consensus iff **C1** and **C2** are valid.*

Proof (\Rightarrow) Suppose $\sim((\varphi^t \wedge \Box \varphi^t) \wedge B\sim \varphi)$ is not valid. Then there is a model \mathcal{M} and state s such that $\mathcal{M}, s \models (\varphi^t \wedge \Box \varphi^t) \wedge B\sim \varphi$. By semantics, we find that this is the case iff $\overline{\mathcal{V}}(\varphi, s) = \{1\}$ and for all t such that sRt , $\mathcal{M}, t \models \varphi^t$ and $\text{Att}(\varphi, s) = -1$ (recall that Consistency is assumed). Therefore Consensus is violated. The case for **C2** is analogous.

(\Leftarrow) Take an arbitrary \mathcal{M} and s . Since **C1** is valid, $\mathcal{M}, s \models \sim((\varphi^t \wedge \Box \varphi^t) \wedge B\sim \varphi)$. By semantics, this corresponds to $\overline{\mathcal{V}}(\varphi, s) = \{1\}$ and for all t such that sRt , $\overline{\mathcal{V}}(\varphi, t) = \{1\}$ implies $\mathcal{M}, s \not\models B\sim \varphi$, therefore $\text{Att}(\varphi, s) \neq -1$. With similar reasoning starting from **C2**, we get the other condition for Consensus, and therefore this postulate is satisfied. \square

⁷ The word *condition* here is used to mean *proposition*, in the most general sense of the word: a statement that can be true or false. It does not have to be a proposition in the language \mathcal{L} .

4.4.2 Consolidation Policies

First, we will look at the most straightforward (and naive) possibility: $\mathcal{M}, s \models B\varphi$ iff $\mathcal{M}, s \models \Box\varphi$. This possibility is appealing because it is familiar and simple. First, let us note that, in order to include the evidence of the agent itself in the consolidation, we have to require the model to be reflexive. This raises the question: is the agent a peer of herself (see Elga (2007))? If yes, then we should only work with reflexive models, if not, then only with *anti-reflexive* models (sRs holds for no s). This is not so crucial as we can (and will) use an equivalent definition for anti-reflexive models: $\mathcal{M}, s \models B\varphi$ iff $\mathcal{M}, s \models \varphi \wedge \Box\varphi$. So we assume that agents are not peers of themselves. We call this latter definition *naive consolidation*.

Proposition 4.23 *Naive consolidation satisfies Consistency, Modesty, Equal Weight, Atom Independence, Monotonicity and Strong Consensus. It does not satisfy Doxastic Freedom and Logical Omniscience.*

Proof We just show the case for Doxastic Freedom. Consider a singleton set $S = \{s\}$ and an atom p . There is no model $\mathcal{M} = (S, R, \mathcal{V})$ with $\text{Att}(p, s) = 0$. \square

Surprisingly, naive consolidation only fails one core postulate: Doxastic Freedom. This is surprising because this consolidation actually *ignores all negative evidence*.

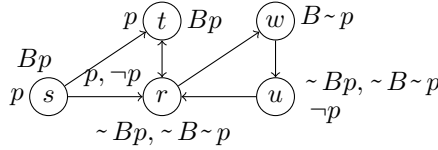


Figure 4.1: An example of naive consolidation. Agent s believes p , but not $\sim p$, since all her peers and herself satisfy p (have evidence for p), and not $\sim p$. One of the peers (r) has $\neg p$, but s ignores that. Agent w believes $\sim p$, even though she does not have evidence against p . She believes $\sim p$ only on the grounds that she and r do not have evidence for p . Agent u believes neither p nor $\sim p$, because she does not have evidence for p , but her only peer does.

Another simple consolidation we can analyse is the *sceptical consolidation*, which sets $\mathcal{M}, s \not\models B\varphi$ for all $\varphi \in \mathcal{L}_0$. Fortunately this extreme position is blocked by two of our core postulates.

Proposition 4.24 *Sceptical consolidation satisfies Consistency, Equal Weight, Atom Independence, Monotonicity and Consensus. It does not satisfy in general No Gurus (and therefore Modesty), Doxastic Freedom and Logical Omniscience.*

Now we will try a more sophisticated definition:

Definition 4.25 *Call \mathcal{C} -consolidations the policies defined by:*

$$\begin{array}{ll}
\mathcal{M}, s \models Bp & \text{iff } \mathcal{C}(V_p^s, V_{\neg p}^s, V_{\Diamond p}^s, V_{\Diamond \neg p}^s, V_{\Box p}^s, V_{\Box \neg p}^s) = 1 \\
\mathcal{M}, s \models B \sim p & \text{iff } \mathcal{C}(V_p^s, V_{\neg p}^s, V_{\Diamond p}^s, V_{\Diamond \neg p}^s, V_{\Box p}^s, V_{\Box \neg p}^s) = -1 \\
\mathcal{M}, s \models B \sim \varphi & \text{iff } \mathcal{M}, s \models B\varphi \\
\mathcal{M}, s \models B(\varphi \wedge \psi) & \text{iff } \mathcal{M}, s \models B\varphi \text{ and } \mathcal{M}, s \models B\psi \\
\mathcal{M}, s \models B \sim (\varphi \wedge \psi) & \text{iff } \mathcal{M}, s \models B \sim \varphi \text{ or } \mathcal{M}, s \models B \sim \psi
\end{array}$$

where V_χ^t is 1 if $1 \in \overline{\mathcal{V}}(\chi, t)$ and 0 otherwise; and $\mathcal{C} : \{0, 1\}^6 \rightarrow \{1, -1, 0\}$ is a function that maps evidence (in this case represented by the six binary parameters) to a belief attitude (1 for belief, -1 for disbelief and 0 for abstention).

What is a good definition for \mathcal{C} ? As we can see above, the real consolidation effort is only with respect to atomic propositions, while more complex beliefs are formed from those atomic beliefs. Some advantages of this approach are that it uses all evidence available for each atom, the agent still retains some inference power (with which it can derive other beliefs), and avoids malformed definitions, such as: $\mathcal{M}, s \models B\varphi$ iff $\mathcal{M}, s \models \varphi^t \wedge \Box \varphi^t$; $\mathcal{M}, s \models B \sim \varphi$ iff $\mathcal{M}, s \models \varphi^f \wedge \Box \varphi^f$. In words: the agent believes a formula if she and her peers have only positive evidence for it, and believes its negation if she and her peers have only negative evidence for it. This seems like a good (if too cautious) definition at first sight, but it is actually not well-formed. We can verify whether $B \sim \psi$ via the second clause, but also via the first if $\varphi = \sim \psi$. And these can sometimes give conflicting results. We avoid that by using \mathcal{C} only to decide belief for literals. Moreover:

Proposition 4.26 *All \mathcal{C} -consolidations satisfy Consistency and Atom Independence.*

Proof Consistency. By Definition 4.18, the agents can only believe a consistent set of atoms, and from that, given the “classical” nature of the rules to form beliefs in complex formulas, only classical consequences

of this consistent set of atoms can be derived, resulting in a consistent belief state. **Atom Independence.** Given that belief in an atom is only determined by \mathcal{C} , and that if \mathcal{V} does not change for an atom p , none of the parameters for \mathcal{C} will change, we conclude that $\text{Att}(p, s)$ will not change for any s . \square

Our agents under \mathcal{C} -consolidations are not necessarily omniscient, but they present some properties related to unbounded logical power:

Proposition 4.27 *Consider any \mathcal{C} -consolidation, and a maximally consistent set of literals Σ . If $\mathcal{M}, s \models B\sigma$ for all $\sigma \in \Sigma$ and $\Sigma \models \varphi$, then $\mathcal{M}, s \models B\varphi$.*

Proof This is a straightforward proof by structural induction on φ . The only thing to pay attention to here is that, in the step where $\varphi = \sim(\psi \wedge \chi)$, if we assume $\Sigma \models \sim(\psi \wedge \chi)$, we can only conclude that $\Sigma \models \sim\psi$ or $\Sigma \models \sim\chi$ (and then use the I.H.) because Σ is maximal, and therefore for any contingent formula ζ , either $\Sigma \models \zeta$ or $\Sigma \models \sim\zeta$. \square

Corollary 4.28 *Any \mathcal{C} -consolidation satisfying Doxastic Freedom also satisfies No Gurus.*

Proof First, recall that \mathcal{L}_0 is equivalent to classical logic in the sense that if $\Sigma \models \varphi$ in classical logic, then $\Sigma \models \varphi$ in \mathcal{L}_0 (see Chapter 2). Also, notice that any contingent $\varphi \in \mathcal{L}_0$ is a consequence of some consistent set of literals (of form p or $\sim p$). To see this just think about truth tables. Now with Proposition 4.27, we get that, for any \mathcal{C} -consolidation, if a maximally consistent set of literals is believed, its consequences are also believed. From this it follows that for any \mathcal{C} -consolidation satisfying Doxastic Freedom, any set of agents S with $s, t \in S$ and any contingent $\varphi \in \mathcal{L}_0$ there will be a model where $\text{Att}(\varphi, s) \neq \text{Att}(\varphi, t)$, which implies No Gurus. \square

Proposition 4.29 *Belief in \mathcal{C} -consolidations is closed under modus ponens: if $\mathcal{M}, s \models B\varphi$ and $\mathcal{M}, s \models B\sim(\varphi \wedge \sim\psi)$, then $\mathcal{M}, s \models B\psi$.*

Proof Suppose $\mathcal{M}, s \models B\varphi$ and $\mathcal{M}, s \models B\sim(\varphi \wedge \sim\psi)$. By semantics, we know that $\mathcal{M}, s \models B\sim(\varphi \wedge \sim\psi)$ iff $\mathcal{M}, s \models B\sim\varphi$ or $\mathcal{M}, s \models B\psi$. Since \mathcal{C} -consolidations satisfy Consistency, $\mathcal{M}, s \models B\varphi$ implies $\mathcal{M}, s \not\models B\sim\varphi$, therefore $\mathcal{M}, s \models \sim\sim\psi$, and by semantics $\mathcal{M}, s \models B\psi$. \square

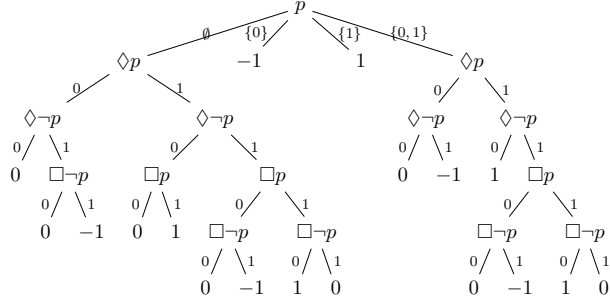


Figure 4.2: Decision trees will be used to represent \mathcal{C} -consolidations. This one represents \mathcal{C} for Policy I. Nodes are labelled by expressions that are representable with the six parameters for \mathcal{C} . The leaves are the outcomes of the consolidation: 1 for belief, -1 for disbelief and 0 for abstention of judgement.

Corollary 4.30 *Any \mathcal{C} -consolidation satisfies Logical Omniscience if we add the following clause to the semantics: if $\models \varphi$, then $\mathcal{M}, s \models B\varphi$ (where $\varphi \in \mathcal{L}_0$).*

We now return to the problem of finding a suitable \mathcal{C} function. There are $3^{(2^6)} = 3^{64} \approx 3.43 \times 10^{30}$ consolidation function candidates for \mathcal{C} . The combinations $(0, 1, 1)$ for $V_{\Diamond p}^s, V_{\Diamond \neg p}^s, V_{\Box p}^s$ and $(1, 0, 1)$ for $V_{\Diamond p}^s, V_{\Diamond \neg p}^s, V_{\Box \neg p}^s$ are impossible, though, which leaves us with “only” $3^{48} \approx 7.98 \times 10^{22}$ relevantly different candidates. In the following, we consider some promising possibilities.

Policy I. Our first social consolidation policy is in Figure 4.2. In cases of unambiguous evidence, the agent decides for belief or disbelief, accordingly. In the case of conflicting evidence, the agent already has some evidence, and since we want a consistent doxastic state, this entails that the agent will inevitably have to discard some evidence. So, in this case, the mere existence of evidence of one kind from one peer is enough to produce belief. However, when the agent has no evidence at all, even if she decides to abstain there is no waste of evidence, so she will be more demanding to change her view. In this case, unanimity of her peers is needed.

Policy II. One might consider that our previous policy still does not justify the different treatment for the problematic evidence cases, and is therefore arbitrary. Hence, we can consider a second policy where the behaviour when the evidence is *none* imitates the case for *both*: consider a decision tree identical to that of Figure 4.2 but with the subtree for none

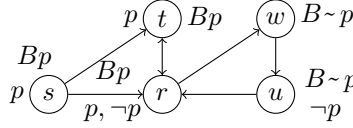


Figure 4.3: Policy I applied to the model of Figure 4.1. Here all agents except for r and w have unambiguous evidence about p , so they can easily form beliefs without looking at their peers. Agent w has no evidence whatsoever, so by the tree of Figure 4.2 she decides to believe $\sim p$ due to her only peer u satisfying $\neg p$. Agent r has evidence both for and against p . Since she has a peer with evidence for p , but no peer with evidence against p , she believes p . Note that by Figure 4.2 this decision would have been different if r had no evidence at all.

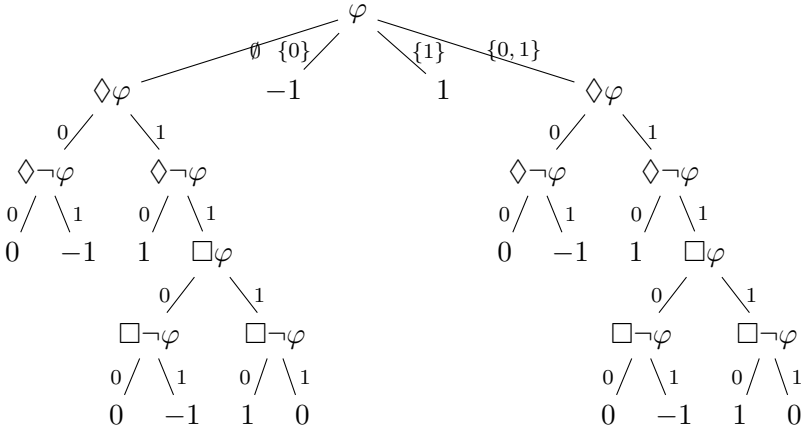


Figure 4.4: Decision tree of \mathcal{C} for Policy II.

(the leftmost subtree) just replaced by that used for both (the rightmost one), as shown in Fig. 4.4.

Proposition 4.31 *Policy I and II satisfy Monotonicity, Doxastic Freedom and Consensus. Modesty and Equal Weight are not satisfied.*

Proof Doxastic Freedom. Let $f : At \times S \rightarrow \{1, 0, -1\}$ be arbitrary. Take a model $\mathcal{M} = (S, R, \mathcal{V})$ where $R = \emptyset$ and make, for all $p \in At$ and $s \in S$, $\mathcal{V}(p, s) = \{1\}$ iff $f(p, s) = 1$, $\mathcal{V}(p, s) = \{0\}$ iff $f(p, s) = -1$ and $\mathcal{V}(p, s) = \emptyset$ otherwise.

Monotonicity. We have to check each case of variation in \mathcal{V} .

$\mathcal{V}(p, s) = \emptyset$ and $\mathcal{V}'(p, s) = \{1\}$. In this case, by the definition of \mathcal{C} , $\text{Att}'(p, s) = 1$. So s does not violate Monotonicity. Now take an arbitrary

agent $t \neq s$. $\mathcal{V}(p, t) = \mathcal{V}'(p, t)$, so V_p^t and $V_{\neg p}^t$ do not change. If not tRs , then the other values also do not change, and then Monotonicity is not violated. Even if tRs , $V_{\Diamond \neg p}^t$ and $V_{\Box \neg p}^t$ do not change. Values $V_{\Box p}^t$ and $V_{\Diamond p}^t$ may change from 0 to 1. By looking at the decision trees for Policy I and II we see that these possible changes in parameters can cause the following changes from $\text{Att}(p, t)$ to $\text{Att}'(p, t) =$: 0 to 1, -1 to 0 and -1 to 1. This last step, of determining what are the changes in the output of \mathcal{C} given the possible changes in parameters, is more reliably done computationally by a simple algorithm on the decision tree of the policy. We will not go through all the cases here, but the reasoning is similar and the last step was always checked via an algorithm.

Consensus. We will prove a stronger version of Consensus, which implies the actual postulate. **Consensus'**: If for some agent $s \in S$ and some $\varphi \in \mathcal{L}_0$ we have that $1 \in \overline{\mathcal{V}}(\varphi, s)$ (or $1 \notin \overline{\mathcal{V}}(\varphi, s)$), and for all $t \in S$ such that sRt : $1 \in \overline{\mathcal{V}}(\varphi, t)$ (or $1 \notin \overline{\mathcal{V}}(\varphi, t)$), then $\text{Att}(\varphi, s) \neq -1$ (or 1).

We prove by structural induction on φ . Base: $\varphi = p$. If $1 \in \mathcal{V}(p, s)$ and for all t with sRt also $1 \in \mathcal{V}(p, t)$, then (by looking at the decision trees of the policies) $\mathcal{M}, s \not\models B \sim p$. Similarly for negative case where $1 \notin \mathcal{V}(p, s)$ and $1 \notin \mathcal{V}(p, t)$ for all t such that sRt .

Step: $\varphi = \sim \psi$. Suppose $1 \in \overline{\mathcal{V}}(\varphi, s)$ (which in this case means just $\overline{\mathcal{V}}(\varphi, s) = \{1\}$) and $1 \in \overline{\mathcal{V}}(\varphi, t)$ for all t with sRt . But then $1 \notin \overline{\mathcal{V}}(\psi, s)$ and $1 \notin \overline{\mathcal{V}}(\psi, t)$ for all t with sRt . By I.H. $\mathcal{M}, s \not\models B \sim \psi$, but by our semantics the only way to obtain $\mathcal{M}, s \models B \sim \varphi (= \sim \sim \psi)$ is if we have $\mathcal{M}, s \models B \psi$. The negative case ($1 \notin \overline{\mathcal{V}}(\varphi, s)$...) is very similar.

$\varphi = \psi \wedge \chi$. Suppose $1 \in \overline{\mathcal{V}}(\varphi, s)$ and $1 \in \overline{\mathcal{V}}(\varphi, t)$ for all t with sRt . This implies that the valuations of ψ and χ contain 1 for s and her peers. By the I.H. $\mathcal{M}, s \not\models B \sim \psi$ and $\mathcal{M}, s \not\models B \sim \chi$. But by our semantics $\mathcal{M}, s \models \sim(\psi \wedge \chi)$ only happens if $\mathcal{M}, s \models B \sim \psi$ or $\mathcal{M}, s \models B \sim \chi$. The negative case follows similar reasoning.

Since **Consensus'** implies **Consensus**, **Consensus** is satisfied. What this proof shows is actually that: *If a \mathcal{C} -consolidation satisfies a version of Consensus' for atoms, it satisfies Consensus' (and therefore Consensus).*

Modesty. Take any atom p . If $\mathcal{V}(p, s) = \{1\}$, then $\text{Att}(p, s) = 1$ and no changes in R or \mathcal{V} can change that.

Equal Weight. Take a model $\mathcal{M} = (S, R, \mathcal{V})$ where $\mathcal{V}(p, s) = \{1\}$, $\mathcal{V}(p, t) = \{0\}$ and sRt for some $p \in At$. If we swap the values in \mathcal{V} between s and t for p , we have $1 = \text{Att}(p, s) \neq \text{Att}'(p, s) = -1$. \square

Policy III. The previous policies are in the “steadfast” category. Our agent gives more weight to her own evidence than to others’ opinions. We can

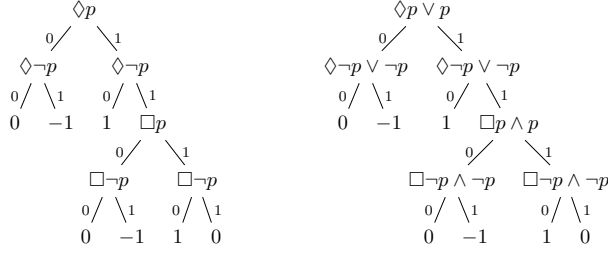


Figure 4.5: Decision trees for \mathcal{C} of Policy III for reflexive (left) and anti-reflexive (right) models. Both yield the same beliefs in their respective class of models.

devise a policy that is more in line with the “equal weight” view. In this case, we consider the relation R to be reflexive, and then “dissolve” the agents’ exceptionality in the modal expression. Starting from the consolidation of Figure 4.2, we can take its subtree for *both* as the decision tree for this policy (Figure 4.5 (left)), ignoring the inputs $V_\varphi^s, V_{\neg\varphi}^s$. This definition makes no distinction between the agent’s own evidence and her peers’. We will, however, use the definition of Figure 4.5 (right) instead, as we are working with anti-reflexive models.

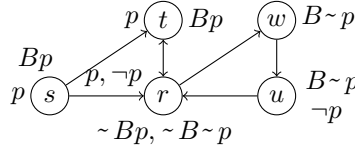


Figure 4.6: Policy III applied to the model of Figure 4.1. Agent w believes $\sim p$ because she or some peer have $\neg p$, but neither she nor her peer have p . All the other agents have evidence for and against p , either by themselves or via some peer. In this case, if the agent and all her peers have one type of evidence but not the other, a belief is formed. For example, agent s and her peers have evidence for p but not all of them have $\neg p$, so she settles with belief in p . Agent r , on the other hand, has evidence for and against p (by herself or via a peer), but they are not unanimous about neither, therefore r abstains.

Proposition 4.32 *Policy III satisfies Modesty, Equal Weight, Monotonicity, Doxastic Freedom and Consensus.*

Proof Modesty. First, we show that for any valuation of an atom, at least two distinct belief attitudes are possible for such atom. For this we need also to use the fact that this consolidation satisfies Doxastic Freedom. By cases:

$\mathcal{V}(p, s) = \emptyset$. If s has no peers, $\text{Att}(p, s) = 0$. If additionally there is an arrow to t with $\mathcal{V}(p, t) = \{0\}$, then $\text{Att}(p, s) = -1$. Or if $\mathcal{V}(p, t) = \{1\}$, then $\text{Att}(p, s) = 1$.

$\mathcal{V}(p, s) = \{0\}$. If s has no peers, $\text{Att}(p, s) = -1$. If there is an arrow to t with $\mathcal{V}(p, t) = \{1\}$, then $\text{Att}(p, s) = 0$. It is not possible to obtain $\text{Att}(p, s) = 1$.

$\mathcal{V}(p, s) = \{1\}$. If s has no peers, $\text{Att}(p, s) = 1$. If there is an arrow to t with $\mathcal{V}(p, t) = \{0\}$, then $\text{Att}(p, s) = 0$. It is not possible to obtain $\text{Att}(p, s) = -1$.

$\mathcal{V}(p, s) = \{0, 1\}$. If s has no peers, $\text{Att}(p, s) = 0$. If there is an arrow to t , then if $\mathcal{V}(p, t) = \{0\}$ we have $\text{Att}(p, s) = -1$, if $\mathcal{V}(p, t) = \{1\}$ we have $\text{Att}(p, s) = 1$.

In summary, for any atom, it is possible to abstain about it, or at least have one attitude among belief/disbelief. As is easy to see, if we make our agent abstain with respect to all atoms, $\text{Att}(\varphi, s) = 0$ for any $\varphi \in \mathcal{L}_0$. But if our agent does not abstain for any atom, she will believe a maximal set of literals, and therefore (by Proposition 4.27) she will either believe φ or $\sim\varphi$, for any $\varphi \in \mathcal{L}_0$. Since this was done with the valuation for s fixed, Modesty follows.

Equal Weight. It is easy to see by Figure 4.5, that for any atom p , if we exchange the valuation of s with that of t , all parameters will be kept the same, and therefore the attitude towards all atoms (and therefore all formulas) will be kept the same.

Monotonicity. This can be proved using the same procedure that was used in Proposition 4.31.

Doxastic Freedom. Same as Proposition 4.31.

Consensus. Same as Proposition 4.31. □

4.5 Dynamics

The dynamic operations we will study use the following models for semantics:

Definition 4.33 Consider a model $\mathcal{M} = (S, R, \mathcal{V})$. We denote by $\mathcal{M}_p^+ = (S, R, \mathcal{V}')$ any model s.t. for some $t \in S$, $\mathcal{V}'(p, t) = \mathcal{V}(p, t) \cup \{1\}$, and $\mathcal{V}'(q, r) = \mathcal{V}(q, r)$ when $q \neq p$ or $r \neq t$. We define \mathcal{M}_p^- , $\mathcal{M}_{\neg p}^+$, $\mathcal{M}_{\neg p}^-$ analogously, but with $\mathcal{V}'(p, t) = \mathcal{V}(p, t) \setminus \{1\}$, $\mathcal{V}'(p, t) = \mathcal{V}(p, t) \cup \{0\}$, $\mathcal{V}'(p, t) = \mathcal{V}(p, t) \setminus \{0\}$, respectively.

Now, with $l \in \{p, \neg p\}$ for some $p \in At$ and $\circ \in \{+, -\}$, we can define the following operator (with obvious additions to the language):

$$\mathcal{M}, s \models [\circ l]\varphi \text{ iff for every model } \mathcal{M}_l^\circ \text{ it holds that } \mathcal{M}_l^\circ, s \models \varphi$$

So, for example, $\mathcal{M}, s \models [+p]\varphi$ can be read as *if evidence for p is added for any agent, φ is the case for s* . A corresponding existential version of this operator can be defined by $\langle \circ l \rangle \varphi \stackrel{\text{def}}{=} \sim [\circ l] \sim \varphi$, with the expected semantics:

$$\mathcal{M}, s \models \langle \circ l \rangle \varphi \text{ iff for some model } \mathcal{M}_l^\circ \text{ it holds that } \mathcal{M}_l^\circ, s \models \varphi$$

We note the following interactions between modalities:

$$\mathcal{M}, s \models \Box[\circ l]\varphi \text{ iff } \mathcal{M}, s \models [\circ l]\Box\varphi \quad \mathcal{M}, s \models \Diamond\langle \circ l \rangle \varphi \text{ iff } \mathcal{M}, s \models \langle \circ l \rangle \Diamond\varphi$$

Interestingly, we can use the axioms below to define Monotonicity, revealing the hidden dynamic nature of that postulate.

M1 $\sim (Bp \wedge \langle +p \rangle \sim Bp)$ M2 $\sim (Bp \wedge \langle -\neg p \rangle \sim Bp)$ M3 $\sim (\sim B \sim p \wedge \langle +p \rangle B \sim p)$ M4 $\sim (\sim B \sim p \wedge \langle -\neg p \rangle B \sim p)$	M5 $\sim (B \sim p \wedge \langle -p \rangle \sim B \sim p)$ M6 $\sim (B \sim p \wedge \langle +\neg p \rangle \sim B \sim p)$ M7 $\sim (\sim Bp \wedge \langle -p \rangle Bp)$ M8 $\sim (\sim Bp \wedge \langle +\neg p \rangle Bp)$
--	--

Proposition 4.34 *A consolidation satisfying Consistency satisfies Monotonicity iff **M1-M8** are valid.*

Proof (\Leftarrow) If $\sim (Bp \wedge \langle +p \rangle \sim Bp)$ is valid, then for any \mathcal{M}, s , it holds that $\mathcal{M}, s \not\models Bp$ or $\mathcal{M}, s \not\models \langle +p \rangle \sim Bp$, which implies that $\mathcal{M}, s \models Bp$ implies $\mathcal{M}, s \models \langle +p \rangle \sim Bp$. This implies that if $\mathcal{M}, s \models Bp$, then there is no \mathcal{M}_p^+ such that $\mathcal{M}_p^+, s \not\models Bp$. This covers one of the cases of Monotonicity. By analogous reasoning with the other axioms, we get all the other cases.

(\Rightarrow) The axiom $\sim (Bp \wedge \langle +p \rangle \sim Bp)$ is valid if, for arbitrary \mathcal{M} and s , $\mathcal{M}, s \models Bp$ implies there is no \mathcal{M}_p^+ such that $\mathcal{M}_p^+, s \not\models Bp$. Indeed a model \mathcal{M}_p^+ satisfies the condition $\mathcal{V}(p, t) \preceq \mathcal{V}'(p, t)$ for some t (by Definition 4.33). In this case Monotonicity implies that $\text{Att}'(p, s) \geq \text{Att}(p, s)$. So indeed, if $\mathcal{M}, s \models Bp$, which by Consistency means that $\text{Att}(p, s) = 1$, we can only have $\text{Att}'(p, s) = 1$, so $\mathcal{M}_p^+, s \models Bp$. So the semantic conditions for **M1** are satisfied. Notice that the case for **M2** is similar, because a model $\mathcal{M}_{\neg p}^-$ also satisfies $\mathcal{V}(p, t) \preceq \mathcal{V}'(p, t)$ for some t . The cases for the other axioms are similar. \square

We can do something similar for Atom Independence, where $l \in \{q, \neg q\}$ and $q \neq p$:

$$\begin{array}{ll}
\mathbf{AI1} & \sim (Bp \wedge \langle ol \rangle \sim Bp) & \mathbf{AI3} & \sim (\sim Bp \wedge \langle ol \rangle Bp) \\
\mathbf{AI2} & \sim (B \sim p \wedge \langle ol \rangle \sim B \sim p) & \mathbf{AI4} & \sim (\sim B \sim p \wedge \langle ol \rangle B \sim p)
\end{array}$$

Proposition 4.35 *For image-finite models and a finite At , a consolidation satisfies Atom Independence iff **AI1-AI4** are valid. For infinite At , validity of **AI1-AI4** does not imply Atom Independence.*

Proof (\Leftarrow) Suppose **AI1-AI4** are valid. If our models are image-finite and At is finite, then for any two models \mathcal{M} and \mathcal{M}' , if there is a p such that for all $s \in S$ we have $\mathcal{V}(p, s) = \mathcal{V}'(p, s)$, then there is a finite sequence: $\mathcal{M}, \mathcal{M}_{l_1}^{\circ 1}, (\mathcal{M}_{l_1}^{\circ 1})_{l_2}^{\circ 2}, \dots, \mathcal{M}'$, where l_1, l_2, \dots do not involve p . If $\text{Att}(p, s) \neq \text{Att}'(p, s)$ (for \mathcal{M} and \mathcal{M}' , respectively), then there is one \mathcal{M}_i in this sequence such that $\text{Att}_i(p, s) \neq \text{Att}_{i+1}(p, s)$. But if **AI1-AI4** are valid, this is not possible.

(\Rightarrow) Assume that Atom Independence is satisfied, and $Bp \wedge \langle ol \rangle \sim Bp$ is satisfiable. Then there is a \mathcal{M}_l° and s such that $\mathcal{M}_l^\circ, s \not\models Bp$, while $\mathcal{M}, s \models Bp$. But then $\mathcal{V}_l^\circ(p, t) = \mathcal{V}(p, t)$ for all t , but $\text{Att}(p, s) \neq \text{Att}_l^\circ(p, s)$, and therefore Atom Independence does not hold. Contradiction. Therefore **AI1** is valid. The other cases are similar.

Now we show a consolidation which satisfies **AI1-AI4** but violates Atom Independence in a setting with infinite At . First, we will need to define some preliminary notions. Let \mathcal{M}, s have a p -canonical valuation iff $\mathcal{V}(p, s) = \{1\}$ and $\mathcal{V}(p, t) = \{1\}$ for all t with sRt , and $\mathcal{V}(q, s) = \{0\}$ and $\mathcal{V}(q, t) = \{0\}$ for all t with sRt , for all $q \neq p$. The p -canonical model of \mathcal{M}, s is a pointed model \mathcal{M}^*, s , where the valuation of \mathcal{M}^* is such that \mathcal{M}^*, s has a p -canonical valuation. For two pointed models \mathcal{M}, s and \mathcal{M}', s' , which differ only in \mathcal{V} , define the *distance* between them to be the size of the sequence (similar to the one built in the first part of this proof) needed to go from \mathcal{M} to \mathcal{M}' . If no such sequence exists, the distance is infinite. We can easily show that (*) if $\mathcal{M}, s \Leftarrow \mathcal{M}', s'$, then \mathcal{M}, s is at a finite distance from its p -canonical model iff \mathcal{M}', s' is at a finite distance from its p -canonical model. Now define a consolidation \mathbb{C} as follows: $\mathcal{M}, s \models Bp$ iff \mathcal{M}, s is at a finite distance from its p -canonical model, and $\mathcal{M}, s \not\models B\varphi$ for all non-atomic φ . This consolidation respects Def 4.18, due to (*). Moreover, this definition violates Atom Independence, for if we take a p -canonical \mathcal{M}, s (with $\text{Att}(p, s) = 1$) and change the valuation of infinitely

many atoms (without changing p) to obtain \mathcal{M}^*, s , this new pointed model is not at a finite distance from its p -canonical model \mathcal{M}, s , and therefore $\text{Att}^*(p, s) \neq 1$. This violates Atom Independence. Axioms **AI1** to **AI4**, however, are valid. Suppose $\mathcal{M}, s \models Bp$. Then \mathcal{M}, s is at a finite distance from its p -canonical model. For $\mathcal{M}, s \models \langle ol \rangle \sim Bp$ to be satisfied, there needs to be a \mathcal{M}_l°, s such that $\mathcal{M}_l^\circ, s \models \sim Bp$. But that would mean that \mathcal{M}_l° is at an infinite distance from its p -canonical model. This is impossible, for \mathcal{M}, s is p -canonical and \mathcal{M}_l° only differs from it in one atom for one agent. \square

The following formula means that *there is an agent other than myself such that if we add/remove evidence l for her, φ holds* (where $l \in \{p, \neg p\}$, for some $p \in \text{At}$):

$$\langle\langle ol \rangle\rangle \varphi \stackrel{\text{def}}{=} (p^t \wedge \langle ol \rangle (p^t \wedge \varphi)) \vee (p^f \wedge \langle ol \rangle (p^f \wedge \varphi)) \vee (p^b \wedge \langle ol \rangle (p^b \wedge \varphi)) \vee (p^n \wedge \langle ol \rangle (p^n \wedge \varphi))$$

The two following postulates could have been defined before, but now we can define them less clumsily:

ES1	$Bp \wedge \langle +\neg p \rangle \sim Bp$	ES3	$Bp \wedge \langle -p \rangle \sim Bp$
ES2	$B \sim p \wedge \langle -\neg p \rangle \sim B \sim p$	ES4	$B \sim p \wedge \langle +p \rangle \sim B \sim p$
SS1	$Bp \wedge \langle\langle +\neg p \rangle\rangle \sim Bp$	SS3	$Bp \wedge \langle\langle -p \rangle\rangle \sim Bp$
SS2	$B \sim p \wedge \langle\langle -\neg p \rangle\rangle \sim B \sim p$	SS4	$B \sim p \wedge \langle\langle +p \rangle\rangle \sim B \sim p$

Condition 4.36 (Evidence Sensitivity) *ES1-ES4 are satisfiable.*

Condition 4.37 (Social Sensitivity) *SS1-SS4 are satisfiable.*

Observation 4.38 *A consolidation satisfying Social Sensitivity also satisfies Evidence Sensitivity.*

Now from Proposition 4.20, 4.22, 4.34-4.35 and Condition 4.36-4.37, we get our main technical result:

Corollary 4.39 *A consolidation satisfies Consistency, Monotonicity, Consensus, Evidence Sensitivity and Social Sensitivity iff: $\sim (B\sigma_1 \wedge \dots \wedge B\sigma_n)$ is valid, for all finite $\Sigma = \{\sigma_1, \dots, \sigma_n\} \subseteq \mathcal{L}_0$ such that $\Sigma \models p \wedge \sim p$; **M1-M8**, **C1-C2** are valid; and **ES1-ES4**, **SS1-SS4** are satisfiable.*

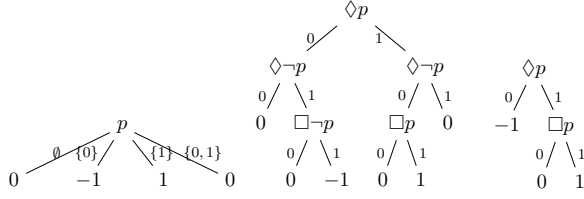


Figure 4.7: Anti-social consolidation (left), Policy IV (center), and Policy V (right).

Atom Independence can be included (with its respective axioms **AI1-AI4**) if we apply the restrictions of Proposition 4.35. The significance of Corollary 4.39 is that it characterises a class of consolidations satisfying almost all core postulates. We conjecture that Doxastic Freedom and No Gurus are not axiomatisable (nor negatively so). A hint of why that might be the case for No Gurus is that it is equivalent to saying that there is a model such that: $(\mathcal{M}, s \models B\varphi$ and $\mathcal{M}, t \not\models B\varphi)$ or $(\mathcal{M}, s \models B\sim\varphi$ and $\mathcal{M}, t \not\models B\sim\varphi)$ or $(\mathcal{M}, s \not\models B\varphi$ and $\mathcal{M}, t \models B\varphi)$ or $(\mathcal{M}, s \not\models B\sim\varphi$ and $\mathcal{M}, t \models B\sim\varphi)$. Our language, however, can only talk of belief from an agent's perspective, or modally (e.g. $\Diamond B\varphi$ – *there is a peer who believes φ*).

Figure 4.7 defines three more \mathcal{C} -consolidations which will show the importance of the new postulates. First, Social Sensitivity is the only core postulate to rule out anti-social consolidation, an unacceptable function that only takes the agent's own evidence into account.

Proposition 4.40 *Anti-social consolidation satisfies Monotonicity, Doxastic Freedom, Consensus and Evidence Sensitivity. Modesty, Equal Weight and Social Sensitivity are not satisfied.*

Now it can be speculated that Evidence Sensitivity can be forced by a combination of other postulates, such as Strong Modesty, Atom Independence and Monotonicity. Policy IV satisfies all those postulates:

Proposition 4.41 *Policy IV satisfies Strong Modesty, Monotonicity, Doxastic Freedom, Consensus and Social Sensitivity. It does not satisfy Equal Weight.*

But that logical connection between those postulates does not hold. Interestingly, Policy IV violates Equal Weight, but this time not by the agents

not giving enough importance to their peers, but by failing to appreciate their own evidence.

Policy V, which is just a modified version of naive consolidation that satisfies Doxastic Freedom, violates Evidence Sensitivity, because, as its cousin, it completely ignores negative evidence. What Evidence Sensitivity enforces is exactly this: that all evidence is taken into account at least in some occasions.

Proposition 4.42 *Policy V satisfies Strong Modesty, Monotonicity, Doxastic Freedom and Consensus. It does not satisfy Equal Weight and Evidence Sensitivity.*

Proposition 4.43 *Policies I, II and III satisfy Social Sensitivity. Naive and sceptical consolidations do not satisfy Evidence Sensitivity.*

A summary of the consolidations appears in Table 4.1. But the main conclusion is that indeed the straightforward definitions such as naive and sceptical consolidations are very unsatisfactory, and the best ones (the only ones satisfying all core postulates) are Policies I-IV, depending on whether one adheres to equal weight or steadfast views.

The $[ol]$ operators make the language more expressive, so we cannot use reduction axioms to obtain equivalent non-dynamic formulas. With these operators we gain the power to *count peers*.⁸ Let us abbreviate $\langle ol \rangle \dots \langle ol \rangle$, repeated n times, by $\langle ol \rangle^n$, with $\langle ol \rangle^0 \varphi \stackrel{\text{def}}{=} \varphi$. Then, with $l \in \{p, \neg p\}$ for some $p \in At$, we have:

$$\begin{aligned} \mathcal{M}, s &\models \sim \langle -l \rangle^n \Box \sim l && \text{iff } s \text{ has more than } n \text{ peers satisfying } l \\ \mathcal{M}, s &\models \sim \langle +l \rangle^n \Box l && \text{iff } s \text{ has more than } n \text{ peers not satisfying } l \\ \mathcal{M}, s &\models \langle -l \rangle^n \Box \sim l && \text{iff } s \text{ has at most } n \text{ peers satisfying } l \\ \mathcal{M}, s &\models \langle +l \rangle^n \Box l && \text{iff } s \text{ has at most } n \text{ peers not satisfying } l \end{aligned}$$

We can abbreviate those formulas by formulas such as $[>n]x$ and $[\leq n]x$, meaning *agent has more than n peers satisfying x* and *agent has at most n peers satisfying x* , respectively, where $x \in \{p, \neg p, \sim p, \sim \neg p\}$, for $p \in At$.

We can also define $[=n]x \stackrel{\text{def}}{=} [\leq n]x \wedge [>n-1]x$, with $n \geq 1$, meaning that *the agent has exactly n peers satisfying x* . For $n = 0$, define $[=0]x \stackrel{\text{def}}{=} \Box \sim x$.

⁸ See Areces, Hoffmann, and Denis (2010); Pacuit and Salame (2004); Baltag, Christoff, Rendsvig, and Smets (2019); Baltag, Christoff, Hansen, and Smets (2013) for modal logics with notions of counting.

	Naive	Scept.	A.-S.	Pol. I	Pol. II	Pol. III	Pol. IV	Pol. V
Atom Independence	✓	✓	✓	✓	✓	✓	✓	✓
Monotonicity	✓	✓	✓	✓	✓	✓	✓	✓
Consensus	✓	✓	✓	✓	✓	✓	✓	✓
No Gurus	✓		✓	✓	✓	✓	✓	✓
Doxastic Freedom			✓	✓	✓	✓	✓	✓
Evidence Sensitivity			✓	✓	✓	✓	✓	
Social Sensitivity				✓	✓	✓	✓	
Modesty	✓					✓	✓	✓
Equal Weight	✓	✓				✓		

Table 4.1: Postulates satisfied by consolidations. A.-S. is anti-social consolidation.

Now $[=n]l \wedge [=m]\sim l$, where $l \in \{p, \neg p\}$, indicates that the agent has exactly $n + m$ peers in total. Since for any $n \in \mathbb{N}$ there are exactly $n + 1$ binary sums that equal n , we can define $[[=n]]$, meaning that an *agent has exactly n peers in total*, via a finite disjunction $([=n]p \wedge [=0]\sim p) \vee ([=n-1]p \wedge [=1]\sim p) \vee \dots \vee ([=0]p \wedge [=n]\sim p)$.

Notice that our counting abilities are limited to \neg -literals (like p and $\neg p$) and their \sim -negations, since our base modalities $[\circ l]$ deal only with \neg -literals. This indicates that a consolidation taking amounts of evidence into account would have to work on the atomic level, just as our \mathcal{C} -consolidations, but the development of such consolidations will be left for future work.

4.6 Related Work

Now we briefly put our work in context with other belief formation/update theories. There are similar works, but in general our multi-agent perspective plus the qualitative and “modal” processing of evidence set our approach apart.

The term “consolidation” employed here is inspired by the homonymous belief revision operator (Hansson, 1991, 1997), where an inconsistent belief base is transformed into a consistent one; likewise, our consolidations must respect the Consistency postulate. One of the most obvious differences between our approach and belief revision is that we are dealing with a multi-agent setting.

As for Bayesianism, the Bayesian update rule tells us how to update our beliefs, but not how to form them – those are the priors, which are usually allowed to be arbitrary. Our models, in principle, seem to be more in line with objective Bayesianism, which is a controversial position, but more research is needed in order to make a more rigorous comparison.

Dempster-Shafer theory of evidence (Dempster, 1968; Shafer, 1976) is a generalisation of probability theory where probabilities can be assigned not only to events but also to sets of events. This theory offers rules for combinations of probability assignments, which in a way can be seen as a kind of consolidation operation.

One of the main differences between our modelling and theories as Dempster-Shafer’s and Bayesianism is that the latter have a clear quantitative take on evidence. Our framework employs a more limited modal language, where such quantitative statements are not even expressible (although we lay the groundwork for such possibility in Section 4.5). In

our models, features such as unanimity and existence of at least one peer with some evidence play important roles, whereas in the other two theories mentioned above these notions are not straightforwardly expressible. Our paper illustrates that there are *some* sensible rationality constraints for formation of evidence-based beliefs even in a limited modal setting, but on the other hand shows the limitations of such a framework and gives the next step towards a quantitative, many-valued modal logical approach to the consolidation problem.

This modal/qualitative perspective is also one of the main differences between our models and opinion diffusion models such as Baltag, Christoff, Rendsvig, and Smets (2019). Although our system is very much in the spirit of other works in opinion dynamics and aggregation and social choice theory (see e.g. Endriss and Grandi (2017)), our setup and treatment of evidence is unique. This contribution does not attempt to offer a *better* formalism for multi-agent evidence-based beliefs, but to highlight how a many-valued modal logic can be used for such a task, bringing an entirely new perspective to this field.

4.7 Conclusions and Future Work

In this chapter we showed that FVEL can be used to model networks of peers, where each one may have different evidence for each atomic proposition, including conflicting and incomplete evidence. We showed that in this setup, there is a question of *consolidation*: how to form beliefs given some evidence? We delineated formally a reasonable class of possible consolidations (Definition 4.18), using a concept similar to bisimulation. Then, we proposed postulates that have to be satisfied in order for a consolidation to be rational, and we showed that (i) they are enough to block many inadequate consolidations and (ii) they are not too strong, as they are jointly satisfiable.

Moreover, we have defined one dynamic operator with the aim of adding and removing evidence. We showed that this operator is useful to formalise some postulates inside the language, and also proposed two important new postulates formulated as axioms containing this operator, without which some unreasonable consolidations would be allowed. With these axioms, we characterised a class of consolidations satisfying most core postulates – with the exception of two which are not axiomatisable. Finally, we showed that this dynamic operator makes the language strictly more expressive, giving it the ability to “count peers”, and how this lays the groundwork

for quantitative consolidations that take amounts of evidence into account – but the development of those are left for future work.

A complete tableau system for FVEL is found in Chapter 2, and Rivieccio (2014a,b) gives an axiomatisation for a language similar to it. Since we use a different version of FVEL, a calculus for it is still missing. Given that we already presented axioms for most postulates, an axiomatic system is preferable. It remains to be seen, however, if such axiomatisation is possible, given that some postulates are not axiomatisable and others are only “negatively” so. Considering that we have not defined a unique belief operator but only constrained the possibilities for such an operator, a complete axiomatisation for our variant of FVEL will probably require one particular consolidation to be chosen. Although we have not talked about public announcements, which in this setting are operations that *remove peers not satisfying some conditions regarding evidence, higher-order evidence or even beliefs*, we know that not all of the reduction axioms of Chapter 2 apply here.

Finally, in the consolidations presented here, the agents form beliefs based on their evidence and *their peers’ evidence*. Another possibility is to make the evidence private to each agent, so that they have to resort only to their own evidence and their peers’ *opinions*. We do precisely that in the next chapter.

Chapter 5

Iterative Social Consolidations: Private Evidence

5.1 Introduction

In artificial intelligence, epistemic and doxastic logics are used as tools to model the knowledge and belief of agents (Meyer and van der Hoek, 1995; van Ditmarsch, van der Hoek, and Kooi, 2007). In practical, real-world scenarios, however, these intelligent agents often have to rely on inconsistent or incomplete data to build up their representation of the world. We can think of this data as *evidence*, a looser and more general concept than that of *justification* as featured in justification logics (Artemov, 1994, 1995, 2001; Fitting, 2005; Mkrtychev, 1997).

Recently, a series of logics have emerged with the purpose of modelling agents who possess evidence (van Benthem and Pacuit, 2011b,a; van Benthem, Fernández-Duque, Pacuit, et al., 2012; van Benthem, Fernandez-Duque, and Pacuit, 2014; Fitting, 2017; Carnielli and Rodrigues, 2019; Özgün, 2017; Santos, 2018; Shi, Smets, and Velázquez-Quesada, 2018a,b). Given this setting, then, a problem is posed: how to *consolidate* this evidence into beliefs? We highlighted the relevance of this problem in Chapter 3, where we use the term *consolidation*¹ to refer to the process of forming beliefs from evidence – formally represented by functions from evidence to doxastic models. The complexity of certain epistemic tasks

¹ Borrowed from belief revision (Hansson, 1991, 1997), where it has the meaning of transforming a potentially inconsistent belief base into a consistent one.

has been studied (e.g. in Dégremont, Kurzen, and Szymanik (2014)), and, in the same vein, looking at consolidations as processes enables us to ask questions about the complexity of such operations.

As in Chapters 3 and 4, we use four-valued epistemic logic (FVEL), defined in Chapter 2, as a base. The resulting system is reminiscent of Baltag, Christoff, Rendsvig, and Smets (2019): agents are represented as nodes, peerhood relations as edges, while belief is decided iteratively. In a first moment, B_0 is defined, based solely on the agent's own evidence. Next, B_1 is defined based on the agent's evidence again plus the B_0 beliefs of her neighbors. Then, B_2 is defined similarly but taking into account B_1 beliefs of the neighbors, and so on. The reason for this choice is to make evidence *private* to each agent. So an agent can only access its own evidence plus its neighbors' *opinions* (or beliefs), but not their evidence as in Chapter 4, which allowed for belief consolidation in a single iteration.

5.2 Logical Language

In this section we explore a variant of *four-valued epistemic logic* (FVEL) proposed in the previous chapter. The only difference here is our new definitions for belief.

5.2.1 Syntax

Let At be a countable set of atoms. Below, $p \in At$; the classical part of the language is given by \mathcal{L}_0 ; the propositional part is given by \mathcal{L}_1 ; and the complete language is given by \mathcal{L} :

$$\begin{aligned}\mathcal{L}_0 \quad \psi &::= p \mid \sim \psi \mid (\psi \wedge \psi) \\ \mathcal{L}_1 \quad \chi &::= \psi \mid \sim \chi \mid (\chi \wedge \chi) \mid \neg \chi \\ \mathcal{L} \quad \varphi &::= \chi \mid \sim \varphi \mid (\varphi \wedge \varphi) \mid \Box \varphi \mid B_i \psi\end{aligned}$$

where $i \in \mathbb{N}$. We abbreviate $\varphi \vee \psi \stackrel{\text{def}}{=} \sim(\sim\varphi \wedge \sim\psi)$ and $\Diamond\varphi \stackrel{\text{def}}{=} \sim\Box\sim\varphi$. We restrict belief to classical propositional formulas (\mathcal{L}_0) because formulas with \neg refer to evidence, and we are not interested here in expressing agents holding beliefs about evidence, only about facts. Formulas such as p are read as *the agent has evidence for p* , whereas $\neg p$ is read as *the agent has evidence against p* , and $\sim\varphi$ as *it is not the case that φ* . We read $\Box\varphi$ as *φ holds for all peers* and $B_i\varphi$ as *the agent believes φ in iteration i* .²

² Our reading of belief formulas is non-standard: $B_i p$ is not *the agent believes she has evidence for p* (at iteration i), but simply *the agent believes p* (at iteration i).

5.2.2 Semantics

Models are tuples $\mathcal{M} = (S, R, \mathcal{V})$, where S is a finite non-empty set of agents, R is a binary anti-reflexive³ relation on S representing “peerhood” and $\mathcal{V} : At \times S \rightarrow \mathcal{P}(\{0, 1\})$ is a four-valued valuation representing agents’ evidence: $\{1\}$ is *true* (t), $\{0\}$ is *false* (f), $\{0, 1\}$ is *both* (b) and \emptyset is *none* (n).⁴ A satisfaction relation is defined as follows:

$$\begin{aligned}
\mathcal{M}, s \models p &\text{ iff } 1 \in \mathcal{V}(p, s) & \mathcal{M}, s \models \neg p &\text{ iff } 0 \in \mathcal{V}(p, s) \\
\mathcal{M}, s \models \sim \varphi & & \text{ iff } \mathcal{M}, s \not\models \varphi \\
\mathcal{M}, s \models (\varphi \wedge \psi) & & \text{ iff } \mathcal{M}, s \models \varphi \text{ and } \mathcal{M}, s \models \psi \\
\mathcal{M}, s \models \neg(\varphi \wedge \psi) & & \text{ iff } \mathcal{M}, s \models \neg\varphi \text{ or } \mathcal{M}, s \models \neg\psi \\
\mathcal{M}, s \models \Box\varphi & & \text{ iff for all } t \in S \text{ s.t. } sRt, \text{ it holds that } \mathcal{M}, t \models \varphi \\
\mathcal{M}, s \models \neg\sim\varphi &\text{ iff } \mathcal{M}, s \models \varphi & \mathcal{M}, s \models \neg\neg\varphi &\text{ iff } \mathcal{M}, s \models \varphi
\end{aligned}$$

An extended valuation function can be defined differently for each type of formula. If $\varphi \in \mathcal{L}_1$, then: $1 \in \overline{\mathcal{V}}(\varphi, s)$ iff $\mathcal{M}, s \models \varphi$; $0 \in \overline{\mathcal{V}}(\varphi, s)$ iff $\mathcal{M}, s \models \neg\varphi$. Otherwise: $1 \in \overline{\mathcal{V}}(\varphi, s)$ iff $\mathcal{M}, s \models \varphi$ iff $0 \notin \overline{\mathcal{V}}(\varphi, s)$. We can also define formulas discriminating which of the four truth values formula $\varphi \in \mathcal{L}_1$ has: $\varphi^n \stackrel{\text{def}}{=} (\sim\varphi \wedge \sim\neg\varphi)$; $\varphi^f \stackrel{\text{def}}{=} \sim\sim(\sim\varphi \wedge \neg\varphi)$; $\varphi^t \stackrel{\text{def}}{=} \sim\sim(\varphi \wedge \sim\neg\varphi)$; $\varphi^b \stackrel{\text{def}}{=} \sim\sim(\varphi \wedge \neg\varphi)$. If a formula $\varphi \in \mathcal{L}_1$ has valuation t or f we say the evidence for φ is *unambiguous*, otherwise we say it is *ambiguous*. If $\mathcal{V}(p, s) = \{1\}$ we say s is a *t-agent* (w.r.t. p , but this will usually be omitted as we will mostly be thinking of a fixed atom p), if $\mathcal{V}(p, s) = \{0\}$ we say s is an *f-agent*, otherwise we say s is a *b/n-agent*.⁵

We say that $\Sigma \models \varphi$ (Σ *entails* φ) when for all models \mathcal{M} and states s , if $\mathcal{M}, s \models \sigma$ for all $\sigma \in \Sigma$, then $\mathcal{M}, s \models \varphi$. We say that $\mathcal{M} \models \varphi$ if $\mathcal{M}, s \models \varphi$ for all states s of \mathcal{M} . And $\models \varphi$ (φ is *valid*) if $\mathcal{M} \models \varphi$ for all \mathcal{M} ; otherwise φ is *invalid*. If $\models \sim\varphi$, we say φ is *contradictory*, and if φ is not contradictory nor valid, it is *contingent*. If a formula is valid or contingent, it is *satisfiable*.

³ In Baltag, Christoff, Rendsvig, and Smets (2019), the authors work with symmetric, serial and irreflexive relations. Irreflexivity here means that the agents are not peers of themselves.

⁴ We stick to the standard (Belnap, 1977) in the naming of truth values. In our context, however, n is better understood as *no evidence about* φ , t as *only evidence for* φ (or *positive evidence*), f as *only evidence against* φ (or *negative evidence*), and b as *evidence both for and against* φ .

⁵ For this chapter we could have used only three values (t , f and b/n), but since this thesis is entirely based on FVEL, we chose to keep the four values.

Notice that the semantics for belief was left open. Our goal in this chapter is to discuss a number of possible definitions for the semantics of belief, taking into account that evidence is private to each agent, therefore belief can only be defined from each agent's own evidence plus their neighbors' beliefs.

5.3 Iterative Social Consolidations: Preliminaries

The following definition will be employed throughout this chapter:

Definition 5.1 (Attitude) *Let $\text{Att}_i : \mathcal{L}_0 \times S \rightarrow \{1, 0, -1\}$ be a function such that: $\text{Att}_i(\varphi, s) = 1$ iff $\mathcal{M}, s \models B_i\varphi$; $\text{Att}_i(\varphi, s) = -1$ iff $\mathcal{M}, s \models B_i\sim\varphi$; otherwise $\text{Att}_i(\varphi, s) = 0$. (The function Att_i also depends on a model \mathcal{M} , but this will be left implicit. We will write Att'_i if we are referring to a modified model \mathcal{M}' .)*

How to define beliefs from the evidence, i.e., how to *consolidate*? Before defining any consolidations, we will present the following notion, which is similar to bisimulation:

Definition 5.2 (n -Equivalence) *We say that $(\mathcal{M}, s) \equiv_n (\mathcal{M}', s')$ iff (\mathcal{M}, s) and (\mathcal{M}', s') satisfy exactly the same \mathcal{L}_1 formulas and exactly the same formulas of the form $\Box B_i\varphi$, $\Diamond B_i\varphi$, $\Box \sim B_i\varphi$, $\Diamond \sim B_i\varphi$, where $i \leq n$.*

Now we employ this equivalence to limit the space of possibilities. All consolidations have to conform to the following condition:

Definition 5.3 (Consolidation Definability Condition (CDC)) *If $(\mathcal{M}, s) \equiv_n (\mathcal{M}', s')$, then, for all $\varphi \in \mathcal{L}_0$, $\mathcal{M}, s \models B_{n+1}\varphi$ iff $\mathcal{M}', s' \models B_{n+1}\varphi$.*

What the CDC does is to make consolidations behave as functions whose input is the initial evidence and the *belief history* of peers. It will be clear later why we want to consider the history instead of just the last iteration of peers' beliefs. The purpose of the CDC is similar to that of Definition 4.18 in the previous chapter. Even in this limited space, there are many possibilities, so in this chapter we will limit ourselves to consolidations that extend the following definition:

Definition 5.4 *Call regular consolidations the policies respecting, for all $i \in \mathbb{N}$:*

$$\begin{array}{ll}
\mathcal{M}, s \models B_0 p & \text{iff } \mathcal{M}, s \models p^t \\
\mathcal{M}, s \models B_0 \sim p & \text{iff } \mathcal{M}, s \models p^f \\
\mathcal{M}, s \models B_i \sim \sim \varphi & \text{iff } \mathcal{M}, s \models B_i \varphi \\
\mathcal{M}, s \models B_i (\varphi \wedge \psi) & \text{iff } \mathcal{M}, s \models B_i \varphi \text{ and } \mathcal{M}, s \models B_i \psi \\
\mathcal{M}, s \models B_i \sim (\varphi \wedge \psi) & \text{iff } \mathcal{M}, s \models B_i \sim \varphi \text{ or } \mathcal{M}, s \models B_i \sim \psi
\end{array}$$

Behind Definition 5.4 is the idea that only beliefs in literals have to be consolidated, and from those basic beliefs others can be built by simple propositional reasoning (very much in the spirit of Definition 4.25 of the previous chapter). Moreover, the first two clauses say that if the evidence for an atom p is only positive (there is only evidence for p but not against p) or only negative, then the agent will initially believe p or $\sim p$, respectively.

Before the first iteration, the agents have not formed any beliefs, so each agent can only use their own private evidence. In the next iterations, however, every agent may have formed beliefs, and therefore, in order to use all information they have available, the agents can now combine their own evidence with the opinion of their peers to form more robust beliefs.⁶ We remark that here the iterations are not intended to model the passage of time, but are only a necessary technical device used to circumvent the lack of peers' opinions in the beginning. If the goal were to realistically model time, it would make more sense to have asynchronous updates, where one agent updates in each iteration, but we will leave this variant for future work. The beliefs under B_0, B_1, \dots here do not really mean that the agent is convinced about such beliefs at any moment; these are just steps towards the agent's actual beliefs, which we will denote by the operator B (without index), which represents the beliefs of the agent in her *point of stabilisation*. So if the agent does not stabilise her beliefs with respect to a formula φ , we cannot say that she has actually formed any (stable) beliefs on φ .

Definition 5.5 (Stabilisation) *An agent s in a model \mathcal{M} is said to be stable at iteration $i \in \mathbb{N}$ with respect to $\varphi \in \mathcal{L}_0$ if, for all $j \geq i$: $\text{Att}_i(\varphi, s) = \text{Att}_j(\varphi, s)$. A model $\mathcal{M} = (S, R, \mathcal{V})$ is said to be stable at iteration $i \in \mathbb{N}$ w.r.t. $\varphi \in \mathcal{L}_0$ if for all $s \in S$, s is stable at iteration i*

⁶ This iterative process might remind one of Google's famous PageRank algorithm (Page, Brin, Motwani, and Winograd, 1999).

w.r.t. φ . If a model/agent is not stable (w.r.t. a formula) it is unstable. The smallest i such that agent s is stable at iteration i w.r.t. φ is called the stabilisation point of agent s w.r.t. φ . The largest stabilisation point among all agents in \mathcal{M} w.r.t. φ is called the stabilisation point of \mathcal{M} w.r.t. φ . If the stabilisation point of a model/agent (w.r.t. φ) is 0, it is called static (w.r.t. φ).

5.4 Consolidation Policies

In this section we will study three regular consolidation policies. Good policies follow some general principles such as not wasting information, not being too gullible nor too skeptical, etc. In this chapter we do not follow a postulate-based approach as in the previous one, but that could be done in future work. We will highlight the qualities and flaws of each policy as we discuss them.

5.4.1 Policy I: Monotonic Belief Diffusion

Below we define our first consolidation (a complete definition of belief):

Definition 5.6 (Policy I) *Policy I is the regular consolidation with $\mathcal{M}, s \models B_{n+1}p$ iff: $\mathcal{M}, s \models p^t$ or $(\mathcal{M}, s \models p^b \vee p^n$ and $\mathcal{M}, s \models \Diamond B_n p$ and $\mathcal{M}, s \models \Box B_n p$). And analogously for $B_{n+1} \sim p$.*

Now, similarly to Baltag, Christoff, Rendsvig, and Smets (2019), we are faced with the question of what are the conditions under which this specific policy eventually stabilises. In most cases we will only talk about stability referring to some arbitrary atom p , as the dynamics are similar for all formulas.

Lemma 5.7 *For regular consolidations, if a model/agent is stable at iteration i w.r.t. all atoms $p \in At$, then this model/agent is stable at iteration i w.r.t. all formulas $\varphi \in \mathcal{L}_0$.*

Proof Follows directly from Definition 5.4. □

Actually, Policy I is guaranteed to stabilise.

Proposition 5.8 *Under Policy I, for any model \mathcal{M} and $\varphi \in \mathcal{L}_0$, the stabilisation point of \mathcal{M} w.r.t φ is at most k , where k is the length of the longest directed path in \mathcal{M} without repeated edges.*

Proof First we prove the proposition for an arbitrary atom p , which implies the general proposition due to Lemma 5.7. Let k be the length of the longest directed path without repetition, and suppose s is an unstable agent at iteration k . If $k = 0$, it is immediately obvious that this cannot be the case, so let us assume that $k > 0$. If all peers of s were stable at iteration $k - 1$, s would be stable by iteration k (Definition 5.6). So there is an agent s_1 such that sRs_1 and s_1 is unstable at iteration $k - 1$. Similar reasoning applies to s_1 : she has a neighbor s_2 who is unstable at iteration $k - 2$, and so on, until we reach agent s_k who is unstable at iteration 0. But if agent s_k is unstable, there must be an agent s_{k+1} such that s_kRs_{k+1} (which could make the beliefs of s_k change in the next iterations). But, from s to s_{k+1} there is a path of length $k + 1$, which by our assumption (regarding k) means that there is at least one repeated edge in this path, and therefore one repeated agent. This, in turn, implies that we have a cycle with at most k agents (otherwise the length of the longest path without repetition would exceed k). If s_{k+1} is one of the repeated agents, then $s_{k+1} \in \{s, s_1, \dots, s_k\}$; otherwise, the repeated agents are all in $\{s, s_1, \dots, s_k\}$. In any case, there is a cycle whose members s_i are all in $\{s, s_1, \dots, s_k\}$. But, since all $s_i \in \{s, s_1, \dots, s_k\}$ are unstable, they are all b/n -agents. But, if that is the case, then it is not hard to see that $\text{Att}_j(p, s_i) = 0$, for all $j \in \mathbb{N}$. But this means that all $s_i \in \{s, s_1, \dots, s_k\}$ are static. Contradiction. \square

Notice that for consolidations in general, due to the CDC, we cannot talk about fixpoints in the traditional sense, i.e. an iteration i where the beliefs (the output) are the same as in iteration $i - 1$. In Policy I, though, it is the case that if all beliefs are the same in iteration i and $i + 1$, then the model is stable at i .

In this policy, if the evidence is unambiguous, the agent immediately forms belief or disbelief, and never changes. Stabilisation is explained by the following:

Proposition 5.9 *In Policy I, the spread of belief is monotonic: let $l \in \{p, \sim p\}$ for some $p \in \text{At}$; and for all $i \in \mathbb{N}$, let $A_{i,l} = \{s \in S \mid \mathcal{M}, s \models B_i l\}$; then for all $i \in \mathbb{N}$, $A_{i,l} \subseteq A_{i+1,l}$.⁷*

Proof Informally: once an agent adopts belief/disbelief, it means that all her peers have also adopted such attitude (or that she had belief/disbelief

⁷ A similar monotonicity holds for the Threshold Model Update in Baltag, Christoff, Rendsvig, and Smets (2019, Definition 2.4).

from the start, due to unambiguous evidence), which in turn implies that all *their* peers have also done so, and so on... \square

Policy I is also very restrictive: the only possible change in attitude for an agent (“in time”, or relative to the progression of iterations) is from abstention to belief/disbelief.⁸ This leads us to our next definition; a more flexible consolidation.

5.4.2 Policy II: Unstable Consolidations

Definition 5.10 (Policy II) *Policy II is the regular consolidation with $\mathcal{M}, s \models B_{n+1}p$ iff: $\mathcal{M}, s \models p^t$ or $(\mathcal{M}, s \models p^b \vee p^n$ and $\mathcal{M}, s \models \Diamond B_n p$ and $\mathcal{M}, s \models \Box \sim B_n \sim p)$. And analogously for $B_{n+1} \sim p$.*

What changes now is that peers that abstain are ignored (unless all of them abstain), i.e. the agents are less cautious about forming belief/disbelief when their evidence is ambiguous. Policy II is not guaranteed to stabilise. For example, the models of Figure 5.1 do not stabilise – in that figure, agents where p has value t are marked with a t , and similarly for f ; for the other agents, the evidence for p is ambiguous. For the first model of Figure 5.1 (top left), note that the agents with unambiguous evidence adopt belief and disbelief immediately, but a (agent on the top left of the model) and b (top right) keep changing between belief (disbelief) and abstention. First, $B_0 p$ holds for the agent marked with a t . Then, since b abstains (neither $B_0 p$ nor $B_0 \sim p$ hold), the only neighbor of a believes p , therefore we get $B_1 p$ for a , and similarly $B_1 \sim p$ for b . In the next iteration, however, a has a neighbor with $B_1 p$ (the one marked with t) and one with $B_1 \sim p$ (agent b), so she abstains – similarly for b . The cycle repeats indefinitely (thence Policy II is not monotonic as in Proposition 5.9).

Instability is undesirable for consolidations. Even though it might be rational to be always open to changing our minds, specially upon the discovery of new evidence, our models are finite and they receive no new information input during the consolidation process. Therefore, rational agents are expected to decide, in a finite amount of time (or a finite number of iterations) what are their final belief states.

⁸ Similar in spirit to our work, but different in many technical aspects, Liu, Seligman, and Girard (2014) name a change from belief to disbelief (or vice-versa) *revision* and from belief/disbelief to abstention *contraction*, adopting the classical terms from belief revision (Alchourrón, Gärdenfors, and Makinson, 1985). Likewise, the change from abstention to belief/disbelief could be named *expansion*.

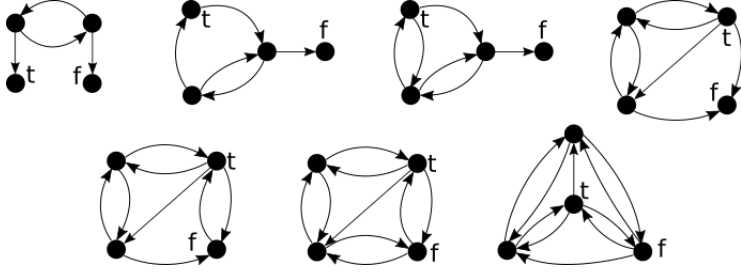


Figure 5.1: All unstable models of size 4 under Policy II have between 4 and 10 edges. This figure shows only some of them.

Now, however, all possible attitude changes between belief, disbelief and abstention are possible.

Proposition 5.11 *Stability in Policy II is decidable.*

Proof For n agents, a model has 3^n possible belief states (sets of attitudes of the agents) at each iteration. Since belief in one iteration depends only on the fixed evidence and on the belief state in the previous iteration, if the belief state in iteration 3^n differs from that of iteration $3^n - 1$, the model is unstable. \square

Proposition 5.12 ⁹ *Let R^+ be the transitive closure of R . Under Policy II, for every $s \in S$ that remains unstable there is a t such that sR^+t and t is in a cycle.*

Proof Consider an agent s such that: (*) there is no t such that sR^+t and t is in a cycle. Then all directed paths starting from this s are finite (forming a rooted directed acyclic graph). We will prove by induction on the length k of the longest path starting at s . I.H: For an agent s such that (*) holds and whose longest path starting from it has size $k \leq n - 1$, s is stable. Base: $k = 0$, then obviously s is stable. Step: $k = n$. Consider any of the longest paths from s : (s, s_1, \dots, s_{n-1}) . Then, (s_1, \dots, s_{n-1}) has length $n - 1$ and (*) holds for s_1 , which by I.H. gives us that s_1 is stable. The other peers of s that belong to smaller paths are also stable, due to the I.H. Therefore, all peers of s are stable and thus so is s . \square

⁹ Here we should be able to draw some connection to abstract argumentation frameworks (Dung, 1995), as in that theory odd cycles result in the inexistence of stable extensions.

Proposition 5.13 *A model with only one b/n -agent s is stable under Policy II.*

Proof If s is the only b/n -agent in the model, then all peers of s are static, and therefore s is stable. \square

Proposition 5.14 *In any model without a t -agent (or without an f -agent), the spread of belief is monotonic (under Policy II).*

Proof Let \mathcal{M} be a model without any f -agents. By Definition 5.10, it is impossible for any agent to have attitude -1 (disbelief). Moreover, the only way that an agent s who “adopted” (changed attitude from 0 to 1) can unadopt (revert back to 0) is when:

- i. all its peers who previously had attitude 1 also unadopted;
- ii. one of its peers changed attitude to -1 .

Since item (ii) is impossible, the only way is via item (i), but then for that to happen it is necessary that all the peers of the peers of s unadopted. This recursion cannot go on forever for our models are finite, and since we do not have an f -agent, there cannot be a first unadopter. Therefore unadoption is impossible and thus belief spread is monotonic. Similar reasoning applies for the case of no t -agent. \square

Definition 5.15 (Submodel) *We say $\mathcal{M}' = (S', R', \mathcal{V}')$ is a submodel of $\mathcal{M} = (S, R, \mathcal{V})$ if $S' \subseteq S$, $R' \subseteq R$, and for all $p \in At$ and $s \in S'$: $\mathcal{V}'(p, s) = \mathcal{V}(p, s)$.*

Definition 5.16 (Model Restriction) *Let $\mathcal{M} = (S, R, \mathcal{V})$. The restriction \mathcal{M}_Z of \mathcal{M} to $Z \subseteq S$ is the submodel $\mathcal{M}_Z = (Z, R', \mathcal{V}')$ of \mathcal{M} with $R' = R \cap (Z \times Z)$.*

Proposition 5.17 *Let $\mathcal{M} = (S, R, \mathcal{V})$, $s \in S$, R^* be the reflexive and transitive closure of R and $R^*(s) = \{t \in S \mid sR^*t\}$. Then, for all $t \in R^*(s)$, all $\varphi \in \mathcal{L}_0$ and all $i \in \mathbb{N}$: $\mathcal{M}_{R^*(s)}, t \models B_i\varphi$ iff $\mathcal{M}, t \models B_i\varphi$.*

Proof Note that $\mathcal{M}_{R^*(s)}, t \models B_i\varphi$, due to Definition 5.10, ultimately boils down to $\mathcal{M}_{R^*(s)}, t \models \psi$, for some $\psi \in \mathcal{L}$, where ψ does not have B_i operators. By our modal semantics, it is clear that $\mathcal{M}_{R^*(s)}, t \models \psi$ cannot possibly be affected by any $r \in S \setminus R^*(s)$. \square

Corollary 5.18 *Any unstable model (under Policy II) has at least one cycle (s_1, \dots, s_n) , with $n \geq 2$ and such that for all $s_i \in \{s_1, \dots, s_n\}$: s_i is a b/n -agent; there are $a, b \in S$ such that $s_i R^+ a$ and $s_i R^+ b$, a is a t -agent and b is an f -agent.*

Proof First we modify the proof of Proposition 5.12 to show that: for any unstable agent s there is a t such that $s R^+ t$ and t is in a cycle consisting only of unstable b/n -agents. We prove the contrapositive by induction as before, by assuming that no such cycle exists, and therefore any path from s to any unstable b/n -agent is finite. The rest of the induction is similar. In the base case, if the agent has no unstable b/n -peer, then all its peers are stable and therefore it is stable.

Now we just have to show that for all members s_i of this cycle, there are agents a and b as in the corollary statement. Note that for any two agents s_i, s_j in a cycle $R^+(s_i) = R^+(s_j)$. Now assume there is no t -agent a such that $s_i R^+ a$ is true. We know by Proposition 5.17 that the beliefs of s_i are the same as in $\mathcal{M}_{R^*(s_i)}$, which has no t -agent. But by Proposition 5.14 we know that in such model the spread of belief is monotonic and therefore the model stabilises. So there has to be a t -agent a with $s_i R^+ a$. Analogous reasoning applies for f -agent b . \square

Corollary 5.19 *The first model of Figure 5.1 (top left) is the smallest (in number of agents and edges) unstable model under Policy II.*

Corollary 5.18 gives necessary but not sufficient conditions for (see Figure 5.2):¹⁰

Open Problem 5.20 *What is the set of unstable models under Policy II?*

The set of such models is obviously infinite, but enumerable. For each $k \in \mathbb{N}$, we just need to generate all models of size k (which is a finite number of models), with each possible valuation (assuming here only one atom: $At = \{p\}$) and combination of edges, and compute whether the model is stable or not (decidable, by Proposition 5.11). This algorithm works as a finite description of the set of unstable models (under Policy II). It would be more interesting, however, to have a more “structural” description, such as the one of Corollary 5.18.

¹⁰ E.g. Christoff and Grossi (2017) solve this problem for a different logic. A more abstract study of oscillations (which we called instability) in logics is found in van Benthem (2015).

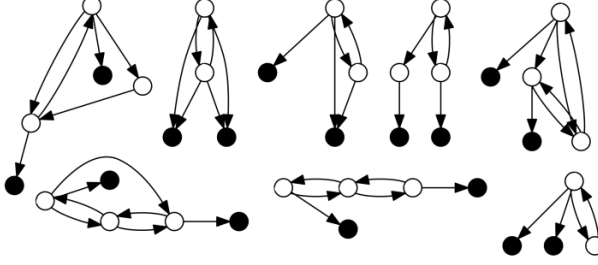


Figure 5.2: Some stable models satisfying the conditions of Corollary 5.18. In each model here, the b/n -agents are white, any one of the black agents can be taken as a t -agent and the other as an f -agent.

Unfortunately, we have not managed to find such a simple structural characterisation of unstable models (and actually we do not know if such a characterisation is even possible), but the following is our attempt at finding “simplifications” that could hopefully yield models that capture the “essence” of instability.

Definition 5.21 (Reduction) *Let \mathbb{M} be the class of all models. A relation $T \subseteq \mathbb{M} \times \mathbb{M}$ is called a semi-reduction if for all models $\mathcal{M} = (S, R, \mathcal{V})$, $\mathcal{M}' = (S', R', \mathcal{V}')$, $\mathcal{M} T \mathcal{M}'$ iff: \mathcal{M}' is stable iff \mathcal{M} is stable; and $S' \subseteq S$. Moreover, if \mathcal{M}' is a submodel of \mathcal{M} , then T is called a reduction.*

Definition 5.22 (Faithful Reduction) *A semi-reduction T is called faithful if for all models $\mathcal{M} = (S, R, \mathcal{V})$, $\mathcal{M}' = (S', R', \mathcal{V}')$, $\mathcal{M} T \mathcal{M}'$ iff: for all $i \in \mathbb{N}$, all $\varphi \in \mathcal{L}_0$ and all $s \in S'$, $\mathcal{M}', s \models B_i \varphi$ iff $\mathcal{M}, s \models B_i \varphi$.*

Note that if for all $\mathcal{M}, \mathcal{M}'$, $\mathcal{M} T \mathcal{M}'$ only if \mathcal{M}' is a restriction of \mathcal{M} , then T is a faithful reduction. Below, let arbitrary models $\mathcal{M} = (S, R, \mathcal{V})$ and $\mathcal{M}' = (S', R', \mathcal{V}')$.

Definition 5.23 *Below we define $T_1, T_2, T_3, T_4, T_5, T_6 \subseteq \mathbb{M} \times \mathbb{M}$ such that:*

- $\mathcal{M} T_1 \mathcal{M}'$ iff: \mathcal{M}' is the submodel of \mathcal{M} such that $R \setminus R' = \{(s, t)\}$, where $s, t \in S$ and s is a t -agent or an f -agent, and $S' = S$.
- $\mathcal{M} T_2 \mathcal{M}'$ iff: there is an $s \in S$ such that there is no $t \in S$ with $s R t$ or $t R s$, and \mathcal{M}' is the restriction of \mathcal{M} to $S \setminus \{s\}$.
- $\mathcal{M} T_3 \mathcal{M}'$ iff: there is a b/n -agent s , a t -agent a and an f -agent b in S such that $s R a$ and $s R b$, and \mathcal{M}' is a restriction of \mathcal{M} to $S \setminus \{s\}$.

- $\mathcal{MT}_4\mathcal{M}'$ iff: there is a b/n -agent $s \in S$ for which there is no t -agent $a \in S$ with sR^+a , and no f -agent $b \in S$ with sR^+b and \mathcal{M}' is a restriction of \mathcal{M} to $S \setminus \{s\}$.
- $\mathcal{MT}_5\mathcal{M}'$ iff: there are at least two distinct t -agents (or f -agents) $a, b \in S$; $S' = S$, $\mathcal{V}' = \mathcal{V}$ and $R' = R \cap ((S \setminus \{b\}) \times (S \setminus \{b\})) \cup Q$, with $Q = \{(a, s) \mid (b, s) \in S\} \cup \{(s, a) \mid (s, b) \in S\}$.
- $\mathcal{MT}_6\mathcal{M}'$ iff: there is a b/n -agent s for which there is no cycle (s_1, \dots, s_n) consisting only of b/n -agents in \mathcal{M} such that for an s_i in (s_1, \dots, s_n) , a t -agent $a \in S$ and an f -agent $b \in S$, it holds that s_iR^+a , s_iR^+b and s_iR^+s ; and \mathcal{M}' is a restriction of \mathcal{M} to $S \setminus \{s\}$.

Note that T_5 is the only of the above in which $\mathcal{MT}_5\mathcal{M}'$ does not require \mathcal{M}' to be a submodel of \mathcal{M} , which means that one has to apply it wisely if one wants to actually simplify a model (basically, one t -agent and one f -agent have to be chosen to concentrate all incoming arrows). It is called a semi-reduction because it does not necessarily yield simpler models.

Proposition 5.24 *The relations T_1, T_2, T_3, T_4 of Definition 5.23 are faithful reductions, T_5 is a faithful semi-reduction and T_6 is a (non-faithful) reduction.*

Proof This proof is straightforward. In some cases, one just has to use Proposition 5.17 and have in mind that peers with attitude 0 do not affect any agent's beliefs. \square

Now one can apply arbitrary sequences of the reductions above to obtain, from an arbitrary model, less cluttered counterparts which are stable if and only if the original was (see Figure 5.3). One can also use only faithful reductions to obtain a simplification where all agents have exactly the same belief history.

Reductions might be a more efficient way of checking whether a model is stable, and it certainly is an easier method for humans in many cases. But a more formal comparison of the complexity between checking stabilisation in the standard way versus using reductions is left for future work.

Just as a side note, we randomly generated (using Erdős and Rényi (1959)'s method) and tested 100,000 models of size 4 to 60 and found the percentages of unstable ones, as shown in Figure 5.4. It is clear that we can expect a 0/1 law here (e.g. as in Verbrugge (2018)), with the percentage going to zero in the limit when the size tends to infinity. This mitigates the problem

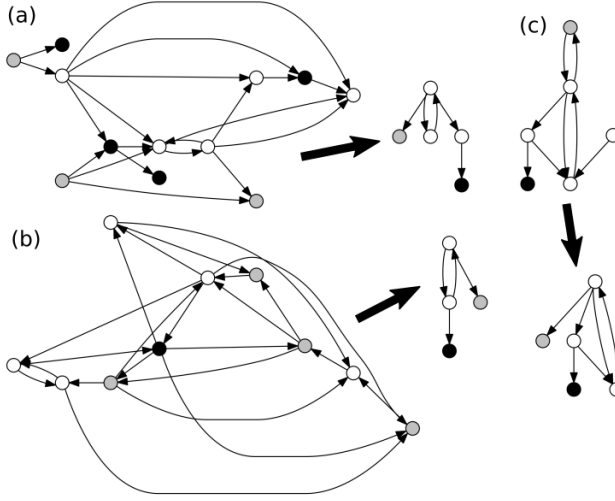


Figure 5.3: Here b/n -agents are white, t -agents are gray and f -agents are black. Many stable models reduce to a single agent model (after applying $T_1 - T_6$ as much as possible), but there are cases like (a) above where it is not reduced completely. Likewise, many unstable models reduce to the smallest unstable model, like case (b), but some do not, as in case (c). These reduced models highlight essential features behind a model’s stability or instability.

of instability for Policy II: in large enough models (such as a big network of scientists, for example), “almost never” will the agents incur an irrational consolidation infinite loop. A possible informal explanation is that, as the size increases, the number of possible structures that can be built grows much faster than the number of possibilities for “instability-inducing” structures, which are very specific: they have to respect the conditions of Corollary 5.18 plus other unknown conditions (Open Problem 5.20). Another interesting question is: why does the percentage of unstable models peak at size 11?

5.4.3 Policy III: Ignoring Unstable Peers

Our next consolidation will try to tackle the instability problem by temporarily ignoring agents who have not been stable for the last λ iterations. Formally, we define the following abbreviation:

$$\mathcal{M}, s \models \mathbf{stable}_{\lambda,p}^n \text{ (with } n \geq \lambda \geq 1 \text{ and } p \in At)$$

with the meaning: $\text{Att}_{n-1}(p, s) = \text{Att}_{n-2}(p, s) = \dots = \text{Att}_{n-\lambda}(p, s)$. (*Agent*

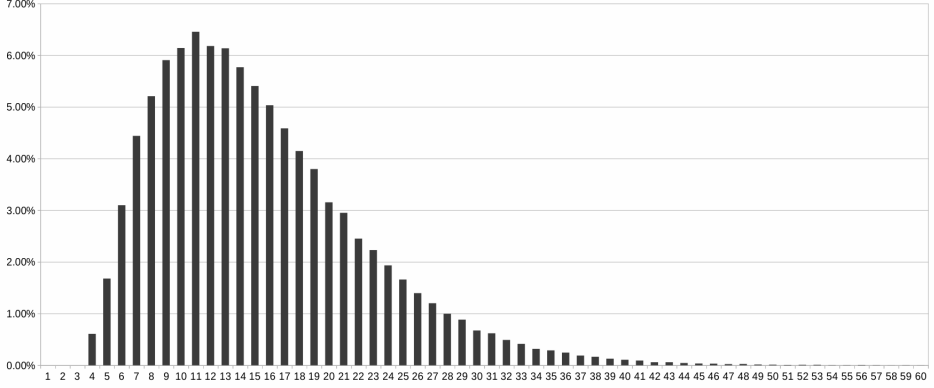


Figure 5.4: Percentage of unstable models per model size, for Policy II.

s has been stable about p in the last λ iterations preceeding iteration n), and set that if $\lambda \leq 1$, then $\mathcal{M}, s \models \mathbf{stable}_{\lambda,p}^n$; and if $\lambda > n$, then $\mathcal{M}, s \models \mathbf{stable}_{\lambda,p}^n$ is defined as $\mathcal{M}, s \models \mathbf{stable}_{n,p}^n$. This abbreviation does not increase expressivity, because it can always be defined by a finite propositional combination of conditions. For example, $\mathcal{M}, s \models \mathbf{stable}_{2,p}^{10}$ is defined as the following disjunctive condition: $(\mathcal{M}, s \models B_9p \text{ and } \mathcal{M}, s \models B_8p)$ or $(\mathcal{M}, s \models B_9\sim p \text{ and } \mathcal{M}, s \models B_8\sim p)$ or $(\mathcal{M}, s \models \sim B_9p \wedge \sim B_9\sim p \text{ and } \mathcal{M}, s \models \sim B_8p \wedge \sim B_8\sim p)$.

From the above, we conclude that $\mathcal{M}, s \models \Box \mathbf{stable}_{\lambda,p}^n$ means that for all $t \in S$ such that sRt , $\mathcal{M}, t \models \mathbf{stable}_{\lambda,p}^n$, and $\mathcal{M}, s \models \Diamond \mathbf{stable}_{\lambda,p}^n$ means that there is a $t \in S$ such that sRt and $\mathcal{M}, t \models \mathbf{stable}_{\lambda,p}^n$. To restrict the modal operators only to stable peers, we can define $\mathcal{M}, s \models \Box_{\lambda,p}^n \varphi$ as $\mathcal{M}, s \models \Box(\sim \mathbf{stable}_{\lambda,p}^n \vee \varphi)$, and $\mathcal{M}, s \models \Diamond_{\lambda,p}^n \varphi$ as $\mathcal{M}, s \models \Diamond(\mathbf{stable}_{\lambda,p}^n \wedge \varphi)$. Now we are ready for:

Definition 5.25 (Policy III- λ) Let $1 \leq \lambda \in \mathbb{N}$. Policy III- λ is the regular consolidation with $\mathcal{M}, s \models B_{n+1}p$ iff: $\mathcal{M}, s \models p^t$ or $(\mathcal{M}, s \models p^b \vee p^n \text{ and } \mathcal{M}, s \models \Diamond_{\lambda,p}^{n+1} B_n p \text{ and } \mathcal{M}, s \models \Box_{\lambda,p}^{n+1} \sim B_n \sim p)$. And analogously for $B_{n+1} \sim p$.

It is not hard to see that Definition 5.25 is compliant with the CDC (Definition 5.3), and is also the reason why we defined the CDC based on the history of peers' beliefs and not only on the last iteration. Note also that if the parameter $\lambda = 1$, Policy III- λ coincides with Policy II, so the former is a generalisation of the latter. Figure 5.5 show the evolution of belief using Policy III- λ on the model of Figure 5.1, with different values of λ .

$\lambda = 1$		$\lambda = 2$		$\lambda = 3$		$\lambda = 4$	
a	b	a	b	a	b	a	b
0	0	0	0	0	0	0	0
1	-1	1	-1	1	-1	1	-1
0	0	1	-1	0	0	0	0
1	-1	0	0	1	-1	1	-1
0	0	1	-1	1	-1	1	-1
\vdots		1	-1	1	-1	1	-1
		0	0	0	0	1	-1
		\vdots		1	-1	0	0
				1	-1	1	-1
				1	-1	1	-1
				0	0	1	-1
				\vdots		1	-1
						0	0
						\vdots	

Figure 5.5: Iterations of belief for agents a and b in the first model of Figure 5.1 (top left). Abstention is represented by 0, belief by 1 and disbelief by -1 (as in Definition 5.1).

The question that immediately surfaces is whether larger values of λ “improve” stability. Looking at Figure 5.5, we notice that larger values of λ make iterations of abstention less frequent. We have to define what “improving stability” means here, formally.

Definition 5.26 (Stability Measure) *A stability measure $<_{\mathcal{M},\varphi}$ for consolidations (w.r.t. a fixed model \mathcal{M} and an arbitrary $\varphi \in \mathcal{L}_0$) has to respect the following principles:*

- i. *If C_1 makes \mathcal{M} static but C_2 does not, then $C_2 <_{\mathcal{M},\varphi} C_1$;*
- ii. *if C_1 makes \mathcal{M} stable but C_2 does not, then $C_2 <_{\mathcal{M},\varphi} C_1$;*
- iii. *if \mathcal{M} stabilises with C_1 at iteration i and with C_2 at $j > i$, then $C_2 <_{\mathcal{M},\varphi} C_1$;*

If for all models \mathcal{M} and all $\varphi \in \mathcal{L}_0$ it is the case that $C_1 <_{\mathcal{M},\varphi} C_2$, then we say that $C_1 < C_2$, i.e. C_2 is more stable than C_1 .

For an arbitrary measure of stability, our initial hypothesis is:

Hypothesis 5.27 *Let $1 \leq \lambda, \kappa \in \mathbb{N}$. If $\lambda < \kappa$, then Policy III- $\lambda <$ Policy III- κ .*

We can start by asking whether Policy III with $\lambda = 1$ can, in some case, be more stable than with $\lambda = 2$. Definition 5.26-i cannot be used to

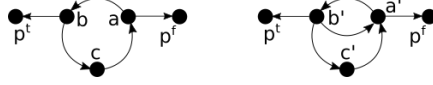


Figure 5.6: Left: Unstable with $\lambda = 1$, stable with $\lambda = 2$. Right: stable with $\lambda = 1$, unstable with $\lambda = 2$.

$\lambda = 1$			$\lambda = 2$			$\lambda = 1$			$\lambda = 2$		
a	b	c	a	b	c	a'	b'	c'	a'	b'	c'
0	0	0	0	0	0	0	0	0	0	0	0
-1	1	0	-1	1	0	-1	1	0	-1	1	0
0	1	-1	-1	1	0	0	0	-1	-1	1	0
0	0	0	0	1	-1	-1	0	0	0	0	-1
\vdots			0	1	0	-1	0	-1	-1	1	0
			<i>stabilised</i>			<i>stabilised</i>			0	0	-1
									\vdots		

Figure 5.7: Iterations of belief for the models of Figure 5.6.

violate Hypothesis 5.27, when changing λ from 1 to 2. If a model is static, changing λ will not produce any changes in belief. Perhaps surprisingly, though, our Hypothesis 5.27 can be violated by Definition 5.26-ii, that is, there is a model that stabilises when $\lambda = 1$ but does not when $\lambda = 2$, namely the model of Figure 5.6 (right). For clearness, the value of p is not shown when it is ambiguous.

The models of Figure 5.6 (left) and Figure 5.6 (right) are the smallest models (considering number of agents and arrows) that feature, respectively: (a) a change from unstable with $\lambda = 1$ to stable with $\lambda = 2$ and (b) the opposite. We tested computationally and verified that phenomena (a) and (b) do not happen in models with 4 or less agents. Another surprising result is that, among models of size 5, there are exactly the same number of models where (a) and (b) occur. Moreover, we tested with $\lambda = 1, \dots, 10$, running for at most 1000 iterations, and all models fitting (a) were stable with $\lambda = 2, \dots, 10$, and all models fitting (b) were unstable with $\lambda = 2, 3$ and stable otherwise. We suspect that phenomena (a) and (b) occur due to the qualitative difference between Policy III- λ with $\lambda = 1$, which equals Policy II, and with $\lambda \geq 2$. Moreover, we can conjecture so far that increasing λ does have a positive effect in terms of stability in general (although Hypothesis 5.27 is false), as the (b)-type models became stable with larger values of λ , despite becoming unstable with $\lambda = 2, 3$.

5.4.4 Other Policies

Other policies that have not yet been explored include the following ideas:

- i. Stopping the consolidation at a fixed iteration defined by a parameter λ . This policy would guarantee stability in a forceful manner. A drawback is that it would be *too sensitive* to the parameter λ , specially in the case of unstable models (under Policy III);
- ii. Limiting the the number of times an agent can change its attitude, e.g. after going from abstention to belief/disbelief, it cannot go back to abstention. We probably will not have problems to define this consolidation respecting the CDC, for even though the agents *do not* take into account their own belief histories directly, these are definable from their previous peers' beliefs;
- iii. Defining belief based on the number of peers holding a certain attitude. In Chapter 4, we show that by introducing a dynamic operator (which increases expressivity) we can count peers with certain attitudes. This would probably be a more realistic way of consolidating beliefs, but demands a language richer than the modal logic used here.
- iv. Allowing t and f -agents to change attitudes. For this, item (iii) above might be helpful. Or, b/n -agents could have a policy similar to Policy II, whereas t and f -agents could be more resistant to change, adopting a strategy in line with Policy I.

5.5 Conclusions and Future Work

In this chapter we used a many-valued modal logic (FVEL) to represent a multi-agent network of peers and their evidence, and defined belief based on this network and on the evidence. This chapter is an alternative view to the previous one, where the agents can access each other's evidence, therefore allowing for consolidations done in one step. By making the evidence private, we triggered an iterative process, through which the agents, always (in Policy I) or almost always (Policy II), approximate a final belief. The exception to this is in the problematic cases of unstable models, which were one of the main topics explored here.

From a practical or realistic point of view, the consolidations presented here might not look very rational. But, within the abstractions and limitations of our modelling (i.e. respecting the CDC), these approaches might be among the most rational possibilities. However, a more in-depth,

formal and argumentative defense of why they are rational in this context is needed, and left for future work (but Section 3.2.2 underlies some of our design choices here). One philosophical point that has to be better defended is the non-temporal aspect of these iterations. We are not trying to represent here agents who are updating their beliefs as the days and years go by. This time can be seen as just a “processing time”. It can also be viewed as passage of time in a very restricted situation where agents cannot do anything besides communicating with their peers – they do not have the opportunity to consult other sources of information or to make deep reflections – as if they are “deliberating” in an isolated room. This deliberation, of course, would be a very simplified one, in which all they are allowed to do is to ask the opinions of their peers, at discrete moments of time, or “turns”.

In the problem of informational cascades (Banerjee, 1992; Bikhchandani, Hirshleifer, and Welch, 1992), rational individual behaviour might lead to bad doxastic outcomes. In that case, this happens due to the evidence being private to each agent, who can only access the others’ final judgements, which is also a feature of our models. The instability, but possibly other irrational behaviours in our system, is partially explained by that feature. A comparison with the system of Chapter 4, where others’ evidence is public, would help to elucidate the impact of private evidence.

There is still much work to be done on these consolidations, especially Policy III, which has not yet been explored in great depth. Other policies like the ones in Section 5.4.4 can also be studied. Moreover, as remarked earlier, here a three-valued logic (such as the ones in Priest (2008, Ch. 7)) of evidence would suffice, but consolidations that distinguish n and b could be developed in future work.

Chapter 6

Conclusions

A summary of each chapter’s accomplishments, shortcomings and plans for future work has already been given in the last section of each chapter, so in this chapter we will look at the work developed in this thesis as a whole, discuss what we have learnt and what deserves some more attention, and try to tie up some loose ends.

6.1 The Logic FVEL

We started in Chapter 2 by presenting *four-valued epistemic logic* (FVEL). Despite being the main technical tool used in the thesis, FVEL itself is not the main contribution of this work, as it is a small deviation from other logics such as BPAL (Rivieccio, 2014a), BK (Odintsov and Wansing, 2010) and K_{FDE} (Priest, 2008). Nevertheless, we bring new technical results on FVEL, such as the tableaux, the correspondence results, the reduction of public announcements and basic propositional results such as the possible truth ranges for certain fragments of the language (Proposition 4.1). Most novel on FVEL, however, is the interpretation we have given, which makes clear what is the role of each negation, how a public announcement should be interpreted, etc.

Chapter 2 shows how the extra negation (\sim) gives FVEL more expressive power than BK and K_{FDE} ; without this, much of the work that follows could not have been done. As for BPAL, it is almost an extension of FVEL. It would be interesting to see how our interpretation extends to BPAL’s additional connectives. It is perhaps unfortunate that we have developed FVEL before becoming aware of BPAL, and later found out how they are actually very similar. Chapter 2 offers a comparison of FVEL with each

of these logics (BPAL, BK and K_{FDE}), and a new interpretation that can to some extent be used for those logics too. This work strengthens the case not only for FVEL, but also for BPAL, as a pair of useful four-valued modal logics.

The philosophical takeaway from Chapter 2 is that the epistemic character of the four-valued valuation and of the modal operator can be reconciled as meaning evidence and uncertainty about this evidence, respectively. And from Chapters 4 and 5, we learn that we can also look at the states as agents, and keep only the valuation as epistemic – but that is also seen in other works on opinion/belief diffusion. Anyhow, these are possible natural interpretations for many-valued modal logics, which can be used to model realistic situations.

Before entering on the topic of consolidations, it is important to ask how FVEL compares to other logics of evidence. In this area, the most prominent alternatives to FVEL are B&P evidence logic (van Benthem and Pacuit, 2011b), the topology-based approach of Baltag, Bezhanishvili, Özgün, and Smets (2016a), and its extensions such as Shi, Smets, and Velázquez-Quesada (2018a). These two types of models (FVEL/BPAL and B&P/topology-based ones) are very different; their main similarity is that, in both, evidence has a four-valued character. Other than that, they are very different formally. In FVEL, evidence has no structure, whereas in the other models it comes with a neighborhood/topological structure. Each of these approaches are suitable for different modelling purposes. One advantage of FVEL is being multi-agent, which may make it more adequate if one has multi-agent systems applications in mind.

Finally, on the comparison of FVEL (in the interpretation of Chapters 4 and 5) with opinion/belief diffusion models (e.g. Baltag, Christoff, Rendsvig, and Smets (2019)), a main difference is that the latter usually do not represent evidence and belief, but just belief (or opinion/behaviour/etc.). To my knowledge, FVEL is the only formalism that can model such scenarios. The uniqueness is more evident in Chapter 4, where the models look like opinion diffusion models, but there are no iterations of belief. That shows that there is just a rough similarity between our models and opinion diffusion models.

6.2 Consolidations

The main contribution of this thesis is certainly everything that has to do with consolidations. We have given a name to this concept, but the

concept itself is not new. Consolidation is the process of forming beliefs from evidence. Of course, that is not a new problem at all. However, it is a relatively modern concept in logics. Doxastic and epistemic logics lack consolidations. Beliefs are held not because agents have evidence to support them, but simply because they are either initially given, or a consequence of other beliefs, or simply tautologies. Some form of consolidation appears in justification logic, but most prominently it seems to have appeared for the first time in van Benthem and Pacuit (2011b).

Taking consolidations seriously can be useful for two main reasons. First, logicians have been fighting logical omniscience for a long time now, and perhaps one of the most natural solutions will involve considering agents who build their beliefs (step by step) from their evidence. The “step by step” part is seen here only in Chapter 5, and even then the agents are quite powerful: in each step, they can decide their beliefs for all atoms. We have not focused much on the computational complexity side of consolidations, but that is a major avenue for future work in this project. Agents can still be omniscient in logics of evidence, but this shift in approach towards “building” beliefs favours stepwise, process-oriented and resource-bounded methods, which would probably help to mitigate the problem of logical omniscience. Combining this resource-bounded perspective with a normative take might pose some challenges. Second, and more important, by explicitly focusing on consolidations, we frame the problem of defining belief in logics in a way that emphasises evidence and its connection to belief. Even if we are not thinking about resource-bounded agents, it is still a relevant question how we should form beliefs from a given body of evidence – including evidence coming from our peers. In summary: consolidations (a) naturally lend themselves to modelling resource-bounded agents, and (b) naturally put evidence and its connection to belief in the spotlight.

Besides remarking that consolidations have to be taken seriously, what have we actually learnt about consolidations?

Chapter 3 Starting in this chapter, we discuss some rationality principles for these operations. Function h_1 (which maps *true* to belief, *false* to disbelief and the other values to abstention; cf. Figure 3.1) is a quite simple consolidation function, maybe even obvious. We gave it, however, a substantial justification via a series of postulates. Postulates Respect for Evidence (RE) and Unanimity Dominance (UD), in particular, seem to be universal and general enough to transcend our specific formal setting. The

other postulates are also reasonable, but depend on the logic being used. They work for FVEL, but probably also for other logics.

Then, we offered a formal construction for cluster consolidations. A similar construction turned out to be useful also for converting FVEL into B&P models. With this consolidation, we were able to obtain a result that is in line with function h_1 , and that maps four-valued evidence models (FVEL models) into familiar bivalent doxastic models (Kripke structures). Our method is able to obtain a representation of belief that is considered standard in logic: an S5 or KD45 Kripke model. And not an arbitrary model, but one that respects h_1 – so the consolidation is somewhat reasonable. Another argument in defense of cluster consolidation is that, by analysing its construction method, we see that the worlds generated in the consolidated model are all worlds with valuations that are “accepted” by the agents, according to their evidence (see Definition 3.2), which means that it makes sense for them to consider those worlds possible. Still, more investigation should be done to find out if there are better consolidations, and if not, why not.

Due to the simplicity of our models, it might not seem that our results have a practical impact on reality, but it is a legitimate practical question whether cautious consolidation (the cluster consolidation implementing h_1) is a rational way of forming beliefs – given that the representation of evidence at hand is as simple as in FVEL models. And this thesis advances the statement that it is. In other words, if you find yourself in a multi-agent setting such as in the “coffee example” (Example 2.7), where some agents are not certain of which is the actual state of the evidence on the topic, cautious consolidation might be your best belief-forming strategy. Actually examining and weighing the evidence might be a more realistic alternative, though. But that requires additional aspects of evidence, such as internal structure, reliability, amount, etc. which are not present in FVEL models.

Chapter 4 In this chapter, once more, we have a long discussion of rationality principles and a list of postulates for consolidations, but this time for a different setting, where the opinions/evidence of others is taken into account. A first loose end that has to be tied in future work is to make a comparison of these postulates across these different settings – if they are general enough, they should not be violated. The postulates of Chapter 3 result in h_1 , which in turn is used as a motivation for consolidations in Chapters 4 and 5. Notice that, for example, Policy I and II just try to do something in the case when h_1 yields abstention, while in the other cases

it just sticks to the judgement provided by h_1 .

The postulates of this chapter are reasonable for many reasons. First, they are not that strong: most good consolidations satisfy all of them. Policies I-IV satisfy Atom Independence, Monotonicity, Consensus, No Gurus, Doxastic Freedom, Evidence Sensitivity and Social Sensitivity. Some of those policies violate Equal Weight and/or Modesty, which are not in the group of so-called *core postulates*, because they reflect the equal weight view in the peer disagreement debate. Second, they are based on already accepted principles of Social Choice Theory (SCT) for voting, now adapted to a doxastic context. For example, Atom Independence is inspired by Independence of Irrelevant Alternatives, Monotonicity by SCT's Monotonicity and Doxastic Freedom by Non-Imposition. Still, extra postulates are probably needed to really rule out any irrational consolidation. On the other hand, even sceptical consolidation, which was tolerated in the form of an H function in Chapter 3, now is rejected by a number of postulates.

So, similarly to the previous chapter, the postulates here aim to be normative, as long as the setting is as depicted by our models. Therefore, in those conditions, the consolidation policies presented are offered as rational ways of forming beliefs.

Chapter 5 This chapter has perhaps the least practical impact, compared to the previous ones, due to our synchronicity assumption – that the agents update their beliefs simultaneously in iterations. In fact, situations such as those are not completely artificial, as some algorithms such as Google's PageRank used to work in a similar way. In a multi-agent system with synchronous time, or in a web-based system where people could only update their public opinions in specific moments, a setting like the one in this chapter could arise. In addition, this chapter has some of the most interesting mathematical and computational puzzles of this thesis, some of which are left open.

Moreover, we learn that Policy I is monotonic and always stabilises, and that Policy II can lead to unstable models, but only rarely as we increase the size of the network. As in the case of informational cascades, private evidence and public access only to opinions/beliefs lead to irrational behaviour (in our case, an infinite loop of consolidation).

6.3 The Magic Word: Rationality

Many design decisions have been justified in the name of *rationality* in this thesis. The reader might think: well, that is not a real justification. What is rationality, anyway? Is it not a subjective concept?

Indeed, the concept of rationality is a complicated one, and there is a lot of discussion about it in philosophy. If we just look at the economic concept of rationality, we find that rational behaviour is one that maximises one's utility. Acting rationally is acting in an optimised (or nearly optimised, if we do not want to be too strict) way. But what are our agents' utility functions? What are they trying to optimise? Since our context is purely doxastic/epistemic, the utility of an agent is entirely determined by her doxastic state. There are two factors that increase this "epistemic utility": (a) having fewer false beliefs, and (b) having more true beliefs. The importance of not having false beliefs is already mentioned in the opening of this thesis. In general, false beliefs can lead to suboptimal actions, which will handicap the agent's utility, whatever it might be. Similarly, having true beliefs helps the agents to make better decisions, which in turn improves their utility. One could also consider other factors for epistemic utility, such as "clutter avoidance" (Harman, 1986) and other cost-related factors. After all, thinking too much without need might be detrimental to one's behaviour. We have not considered such factors here.

Then, how does this type of rationality serve as background for the postulates proposed here? If we take, for example, the Respect for Evidence (RE) postulate, the rationale is that believing in stark opposition to what evidence tells is likely going to lead to false beliefs, as long as evidence bears any connection to reality. Going further, in Chapter 4, we introduced the social dimension to consolidations, because ignoring social evidence is suboptimal behaviour. It is a well-known principle of epistemology that one should always consider the *total evidence* (again, if we disregard cost-related factors in the agent's epistemic utility). All the other postulates have an underlying motivation that ultimately resorts to this concept of rationality.

In conclusion, rationality has been the major guiding principle for all the consolidations proposed here, and that is the reason why this work is in the normative, and not descriptive, category. That is also one of the main reasons why comparisons with opinion diffusion models are superficial: there is a technical similarity, but the motivation is completely different.

6.4 The Mainstream View

As mentioned above, the problem of consolidation is as old as humanity itself. There are important mainstream theories that deal with the question of how to process evidence, for example: Bayesianism, belief revision (Alchourrón, Gärdenfors, and Makinson, 1985), and Dempster-Shafer theory (Dempster, 1968; Shafer, 1976).

The most important next step for this project is probably a thorough comparison with these other theories. As mentioned in Chapter 4, there are significant differences between these formalisms, but just as in the case of a comparison of our postulates across the different settings offered in this thesis, it should be possible to say *something* about how compatible Bayesianism and belief revision are with our work. We know that our theory is at odds with Bayesianism and AGM belief revision, for example, when it comes to tautologies. Our agents do not necessarily believe all tautologies, but in Bayesian epistemology logical truths are assigned probability 1, and in belief revision all belief sets contain all tautologies.

We mentioned earlier that while Bayesianism tries to define how beliefs are updated in the face of new evidence, our theory tries to define how beliefs are formed in the first place, given some evidence. In Bayesianism, there are priors, which represent the agent's initial beliefs. In our theory, the agents start without any beliefs. A consolidation in Bayesianism would be the result of applying a series of updates with all evidence found in some body of evidence. The result might differ depending on the priors, which does not happen in our theory. This puts FVEL consolidations more in line with objective Bayesianism, where not all initial values are acceptable for the priors. This is controversial in itself, but this view has been defended before.¹ Another obstacle for comparing our consolidations with Bayesianism is that we lack the quantitative aspects mentioned in Chapter 3: reliability, amount of evidence, etc. On the other hand, our method relies on notions such as unanimity and existence of *some* evidence, which, in principle, are not expressible in Bayesian epistemology.

Belief revision, on the other hand, is a qualitative theory. The main hurdle for comparing it with our theory is that the most relevant things that belief revision has to say regards input that contradicts a previous belief base. Again, in our case, the agent starts as a blank slate. One could consider, as in the case of Bayesianism, a series of belief revisions, starting from an empty belief set. That would roughly correspond to

¹ See Williamson (2010) for more on objective Bayesianism.

a consolidation. A starting point for comparison would be to look at iterated belief revision frameworks, such as Darwiche and Pearl (1997); Rott (2009).²

6.5 Closing Thoughts

In this thesis we offered a new multi-agent four-valued modal logic that can be used to model evidence scenarios, and a set of results related to this logic. Moreover, we discussed the concept of consolidations, and formulated formal operations that implement them in a variety of formal settings. The main question is how evidence determines belief. We argued for the operations we proposed, and offered rationality postulates that they respect.

The main directions for future work involve generalising the idea of consolidations and the postulates proposed for other settings, and making a detailed comparison with other well-established theories of belief update. The long lists of ideas in the “future work” sections of each chapter attest that there is a long road ahead when it comes to evidence logics and consolidations.

² Darwiche and Pearl (1997) show that their belief revision operation is compatible with a qualitative version of Jeffrey’s rule.

Bibliography

- T. Ågotnes and N. Alechina. The dynamics of syntactic knowledge. *Journal of Logic and Computation*, 17(1):83–116, 2007.
- 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(2):510–530, 1985.
- N. Alechina and B. Logan. Ascribing beliefs to resource bounded agents. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems: Part 2*, pages 881–888. ACM, 2002.
- N. Alechina, B. Logan, and M. Whitsey. A complete and decidable logic for resource-bounded agents. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2004)*, pages 606–613. IEEE Computer Society, 2004.
- A. R. Anderson and N. D. Belnap. *Entailment: The Logic of Relevance and Necessity, Volume 1*. Princeton University Press, Princeton, 1975.
- C. Areces, G. Hoffmann, and A. Denis. Modal logics with counting. In *International Workshop on Logic, Language, Information, and Computation (WoLLIC)*, pages 98–109. Springer, 2010.
- O. Arieli and A. Avron. Reasoning with logical bilattices. *Journal of Logic, Language and Information*, 5(1):25–63, 1996.
- D. M. Armstrong. *Belief, Truth and Knowledge*. Cambridge University Press, 1973.
- K. J. Arrow. *Social Choice and Individual Values*. John Wiley & Sons, 1951.

- S. Artemov. Logic of proofs. *Annals of Pure and Applied Logic*, 67(1-3): 29–59, 1994.
- S. Artemov. Operational modal logic. Technical Report MSI 95-29, Cornell University, December 1995.
- S. Artemov. Explicit provability and constructive semantics. *Bulletin of Symbolic Logic*, 7(1):1–36, 2001.
- F. G. Asenjo. A calculus of antinomies. *Notre Dame Journal of Formal Logic*, 7(1):103–105, 1966.
- Z. Bakhtiari, H. van Ditmarsch, and U. Rivieccio. Bilattice logic of epistemic actions and knowledge. *Journal of Annals of Pure and Applied Logic*. In press.
- P. Balbiani, H. van Ditmarsch, A. Herzig, and T. de Lima. Tableaux for public announcement logic. *Journal of Logic and Computation*, 20(1): 55, 2010.
- P. Balbiani, D. Fernández-Duque, and E. Lorini. A logical theory of belief dynamics for resource-bounded agents. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems (AAMAS 2016)*, pages 644–652, 2016.
- M. Balliu, M. Dam, and G. Le Guernic. Epistemic temporal logic for information flow security. In *Proceedings of the ACM SIGPLAN 6th Workshop on Programming Languages and Analysis for Security*. ACM, 2011.
- 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.
- A. Baltag, B. Renne, and S. Smets. The logic of justified belief change, soft evidence and defeasible knowledge. In *WoLLIC*, volume 7456 of *Lecture Notes in Computer Science*, pages 168–190. Springer, 2012.
- A. Baltag, Z. Christoff, J. U. Hansen, and S. Smets. Logical models of informational cascades. *Studies in Logic*, 47:405–432, 2013.
- A. Baltag, B. Renne, and S. Smets. The logic of justified belief, explicit knowledge, and conclusive evidence. *Annals of Pure and Applied Logic*, 165(1):49–81, 2014.

- A. Baltag, N. Bezhanishvili, A. Özgün, and S. Smets. Justified belief and the topology of evidence. In *International Workshop on Logic, Language, Information, and Computation*, pages 83–103. Springer, 2016a.
- A. Baltag, Z. Christoff, R. Rendsvig, and S. Smets. Dynamic epistemic logics of diffusion and prediction in social networks. In *Proceedings of the 12th Conference on Logic and the Foundations of Game and Decision Theory*, 2016b.
- A. Baltag, Z. Christoff, R. Rendsvig, and S. Smets. Dynamic epistemic logics of diffusion and prediction in social networks. *Studia Logica*, 107(3):489–531, 2019.
- A. V. Banerjee. A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817, 1992.
- J. Barwise and J. Etchemendy. *Language, Proof and Logic*. Seven Bridges Press, New York, 2000.
- N. D. Belnap. A useful four-valued logic. In *Modern Uses of Multiple-valued Logic*, pages 5–37. Springer, 1977.
- E. W. Beth. Semantic entailment and formal derivability. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen*, 18(13):309–342, 1955.
- P. Bieber. A logic of communication in hostile environment. In *Proceedings of The Computer Security Foundations Workshop III*, pages 14–22. IEEE, 1990.
- S. Bikhchandani, D. Hirshleifer, and I. Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5):992–1026, 1992.
- P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, 2002.
- I. Boureanu, M. Cohen, and A. Lomuscio. Automatic verification of temporal-epistemic properties of cryptographic protocols. *Journal of Applied Non-Classical Logics*, 19(4):463–487, 2009.
- G. Cantor. Ein beitrage zur mannigfaltigkeitslehre. *Journal für die Reine und Angewandte Mathematik*, 84:242–258, 1878.

- W. Carnielli and A. Rodrigues. An epistemic approach to paraconsistency: a logic of evidence and truth. *Synthese*, 196(9):3789–3813, 2019.
- R. M. Chisholm. The logic of knowing. *The Journal of Philosophy*, 60(25):773–795, 1963.
- D. Christensen. Higher-order evidence. *Philosophy and Phenomenological Research*, 81(1):185–215, 2010.
- Z. Christoff and D. Grossi. Stability in binary opinion diffusion. In A. Baltag, J. Seligman, and T. Yamada, editors, *International Workshop on Logic, Rationality and Interaction (LORI)*, pages 166–180. Springer, 2017.
- Z. Christoff and J. U. Hansen. A logic for diffusion in social networks. *Journal of Applied Logic*, 13(1):48–77, 2015.
- N. C. A. da Costa. On the theory of inconsistent formal systems. *Notre Dame Journal of Formal Logic*, 15(4):497–510, 1974.
- A. Darwiche and J. Pearl. On the logic of iterated belief revision. *Artificial Intelligence*, 89(1):1–29, 1997.
- F. Dechesne and Y. Wang. Dynamic epistemic verification of security protocols: Framework and case study. In *Proceedings of the Workshop on Logic, Rationality and Interaction (LORI)*, pages 129–144, 2007.
- C. Dégremont, L. Kurzen, and J. Szymanik. Exploring the tractability border in epistemic tasks. *Synthese*, 191(3):371–408, 2014.
- A. P. Dempster. A generalization of Bayesian inference. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(2):205–232, 1968.
- F. Dietrich and C. List. Probabilistic opinion pooling. In A. Hájek and C. Hitchcock, editors, *The Oxford Handbook of Probability and Philosophy*. Oxford University Press, 2016.
- P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–357, 1995.
- J. Dunn. Intuitive semantics for first-degree entailments and ‘coupled trees’. *Philosophical Studies*, 29(3):149–168, 1976.

- A. Elga. Reflection and disagreement. *Noûs*, 41(3):478–502, 2007.
- U. Endriss and U. Grandi. Graph aggregation. *Artificial Intelligence*, 245: 86–114, 2017.
- P. Erdős and A. Rényi. On random graphs I. *Publicationes Mathematicae*, 6:290–297, 1959.
- R. Fagin and J. Halpern. Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34(1):39–76, 1987.
- R. Fagin, J. Y. Halpern, Y. Moses, and M. Vardi. *Reasoning about Knowledge*. MIT press, 1995.
- R. Feldman. Reliability and justification. *The Monist*, 68(2):159–174, 1985.
- R. Feldman. Respecting the evidence. *Philosophical Perspectives*, 19(1): 95–119, 2005.
- R. Feldman and E. Conee. Evidentialism. *Philosophical Studies*, 48(1): 15–34, 1985.
- M. Fitting. *Proof Methods for Modal and Intuitionistic Logics*, volume 169. Springer Science & Business Media, 1983.
- M. Fitting. Many-valued modal logics. *Fundamenta Informaticae*, 15(3-4): 235–254, 1991.
- M. Fitting. The logic of proofs, semantically. *Annals of Pure and Applied Logic*, 132(1):1–25, 2005.
- M. Fitting. Reasoning with justifications. In *Towards Mathematical Philosophy*, pages 107–123. Springer, 2009.
- M. Fitting. Paraconsistent logic, evidence and justification. *Studia Logica*, 105(6):1149–1166, 2017.
- S. C. Fletcher, J. Landes, and R. Poellinger. Evidence amalgamation in the sciences: An introduction. *Synthese*, 196(8):3163–3188, 2018.
- R. Foley. Self-trust and the authority of others. In R. Foley, editor, *Intellectual Trust in Oneself and Others*, pages 83–130. Cambridge University Press, 2001.
- G. Frege. *Grundgesetze der Arithmetik*. Hermann Pohle, Jena, 1893.

- E. Fricker. Testimony and epistemic autonomy. In J. Lackey and E. Sosa, editors, *The Epistemology of Testimony*, pages 225–250. Oxford University Press, 2006.
- J. Gerbrandy. *Bisimulations on Planet Kripke*. PhD thesis, Institute for Logic, Language and Computation, University of Amsterdam, 1999.
- J. Gerbrandy and W. Groeneveld. Reasoning about information change. *Journal of Logic, Language, and Information*, 6(2):147–169, 1997.
- E. L. Gettier. Is justified true belief knowledge? *Analysis*, 23(6):121–123, 1963.
- S. Ghosh and D. de Jongh. Comparing strengths of beliefs explicitly. *Logic Journal of the IGPL*, 21(3):488–514, 2013.
- A. Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, 41(4):587–601, 1973.
- P. Girard and K. Tanaka. Paraconsistent dynamics. *Synthese*, 193(1):1–14, 2016.
- K. Gödel. Über formal unentscheidbare sätze der Principia Mathematica und verwandter systeme I. *Monatshefte für Mathematik*, 1930.
- A. I. Goldman. Innate knowledge. In S. P. Stich, editor, *Innate Ideas*, pages 111–120. University of California Press, Berkeley, 1975.
- A. I. Goldman. What is justified belief? In G. Pappas, editor, *Justification and Knowledge*, pages 1–23. Springer, 1979.
- A. I. Goldman. Internalism, externalism, and the architecture of justification. *The Journal of Philosophy*, 106(6):309–338, 2009.
- J. Y. Halpern and Y. Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54(3):319–379, 1992.
- J. Hansen. Terminating tableaux for dynamic epistemic logics. *Electronic Notes in Theoretical Computer Science*, 262:141–156, 2010.
- S. O. Hansson. *Belief base dynamics*. PhD thesis, Uppsala University, 1991.

- S. O. Hansson. Semi-revision. *Journal of Applied Non-Classical Logics*, 7 (1-2):151–175, 1997.
- J. Hardwig. Epistemic dependence. *Journal of Philosophy*, 82(7):335–349, 1985.
- G. Harman. *Change in View*. MIT Press, Cambridge, Massachusetts; London, England, 1986.
- A. Heyting. *Intuitionism: an introduction*, volume 41. Elsevier, 1966.
- D. Hilbert and P. Bernays. *Grundlagen der Mathematik*. Springer, 1934.
- J. Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*, volume 4 of *Contemporary Philosophy*. Cornell University Press, Ithaca, 1962.
- J. Hintikka. Impossible possible worlds vindicated. In *Game-Theoretical Semantics*, pages 367–379. Springer, 1979.
- S. Jaśkowski. Rachunek zdań dla systemów dedukcyjnych sprzecznych. *Studia Societatis Scientiarum (Sectio A)*, 1(5):55–77, 1948.
- J. M. Joyce. The development of subjective Bayesianism. In *Handbook of the History of Logic*, volume 10, pages 415–475. Elsevier, 2011.
- A. Jung and U. Rivieccio. Kripke semantics for modal bilattice logic. In *Proceedings of the 2013 28th Annual ACM/IEEE Symposium on Logic in Computer Science*, pages 438–447. IEEE Computer Society, 2013.
- T. Kelly. Peer disagreement and higher order evidence. In A. I. Goldman and D. Whitcomb, editors, *Social Epistemology: Essential Readings*, pages 183–217. Oxford University Press, 2010.
- S. Kripke. A completeness theorem in modal logic. *Journal of Symbolic Logic*, 24(1):1–14, 1959.
- S. Kripke. Semantical analysis of modal logic I: Normal modal propositional calculi. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 9:67–96, 1963a.
- S. Kripke. Semantical considerations on modal logic. *Acta Philosophica Fennica*, 16:83–94, 1963b.

- M. Lasonen-Aarnio. Higher-order evidence and the limits of defeat. *Philosophy and Phenomenological Research*, 88(2):314–345, 2014.
- K. Lehrer. Social information. *The Monist*, 60(4):473–487, 1977.
- H. Levesque. A logic of implicit and explicit belief. In *AAAI*, pages 198–202, 1984.
- C. I. Lewis and C. H. Langford. *Symbolic Logic*. Century Company, New York, 1932.
- F. Liu, J. Seligman, and P. Girard. Logical dynamics of belief change in the community. *Synthese*, 191(11):2403–2431, 2014.
- J. MacFarlane. In what sense (if any) is logic normative for thought. Unpublished manuscript, 2004.
- C. Martini, J. Sprenger, and M. Colyvan. Resolving disagreement through mutual respect. *Erkenntnis*, 78(4):881–898, 2013.
- B. K. Matilal. *The Character of Logic in India*. State University of New York Press, 1999.
- V. B. Matos, R. Guimarães, Y. D. Santos, and R. Wassermann. Pseudo-contractions as gentle repairs. In *Description Logic, Theory Combination, and All That. Essays Dedicated to Franz Baader on the Occasion of His 60th Birthday*, pages 385–403. Springer, 2019.
- J.-J. C. Meyer and W. van der Hoek. *Epistemic Logic for AI and Computer Science*. Cambridge University Press, 1995.
- A. Mkrtychev. Models for the logic of proofs. In *International Symposium on Logical Foundations of Computer Science*, pages 266–275. Springer, 1997.
- C. Morgan. Local and global operators and many-valued modal logics. *Notre Dame Journal of Formal Logic*, 20(2):401–411, 1979.
- O. Morikawa. Some modal logics based on a three-valued logic. *Notre Dame Journal of Formal Logic*, 30(1):130–137, 1988.
- S. P. Odintsov and H. Wansing. Modal logics with Belnapian truth values. *Journal of Applied Non-Classical Logics*, 20(3):279–301, 2010.

- P. Ostermann. Many-valued modal propositional calculi. *Mathematical Logic Quarterly*, 34(4):343–354, 1988.
- A. Özgün. *Evidence in epistemic logic: a topological perspective*. PhD thesis, University of Amsterdam, 2017.
- E. Pacuit and S. Salame. Majority logic. In D. Dubois, C. Welty, and M.-A. Williams, editors, *Principles of Knowledge Representation and Reasoning, KR 2004*, volume 4, pages 598–605. AAAI Press, 2004.
- L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November 1999. Previous number = SIDL-WP-1999-0120.
- J. 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 (ISMIS 1989): Poster Session Program*, pages 201–216, Charlotte, NC: Oak Ridge National Laboratory ORNL/DSRD-24, 1989.
- J. Plaza. Logics of public communications. *Synthese*, 158(2):165–179, 2007.
- G. Priest. *An Introduction to Non-classical Logic: From If to Is*. Cambridge University Press, 2 edition, 2008.
- G. Priest and R. Routley. Applications of paraconsistent logic. In G. Priest, R. Routley, and J. Norman, editors, *Paraconsistent Logic: Essays on the Inconsistent*, chapter 13. Philosophia Verlag, München, 1989a.
- G. Priest and R. Routley. The philosophical significance and inevitability of paraconsistency. In G. Priest, R. Routley, and J. Norman, editors, *Paraconsistent Logic: Essays on the Inconsistent*, chapter 18. Philosophia Verlag, München, 1989b.
- F. P. Ramsey. Knowledge. In R. B. Braithwaite, editor, *Foundations of Mathematics and Other Logical Essays*, pages 126–128. Kegan Paul, London, 1931.
- V. Rantala. Urn models: A new kind of non-standard model for first-order logic. *Journal of Symbolic Logic*, 4:455–474, 1975.
- B. Renne. Evidence elimination in multi-agent justification logic. In A. Heifetz, editor, *Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge*, pages 227–236. ACM, 2009.

- U. Rivieccio. Bilattice public announcement logic. In R. Goré, B. Kooi, and A. Kurucz, editors, *Proceedings of Advances in Modal Logic (AiML)*, volume 10, pages 459–477. College Publications, 2014a.
- U. Rivieccio. Algebraic semantics for bilattice public announcement logic. *Studia Logica, Proc. Trends in Logic XIII*, Springer, 2014b.
- H. Rott. Shifting priorities: Simple representations for twenty-seven iterated theory change operators. In D. Makinson, J. Malinowski, and H. Wansing, editors, *Towards Mathematical Philosophy*, pages 269–296. Springer, 2009.
- Y. D. Santos. A dynamic informational-epistemic logic. In A. Madeira and M. Benevides, editors, *Dynamic Logic. New Trends and Applications*, volume 10669 of *Lecture Notes in Computer Science*, pages 64–81. Springer, 2018.
- Y. D. Santos. Consolidation of belief in two logics of evidence. In P. Blackburn, E. Lorini, and M. Guo, editors, *International Conference on Logic, Rationality and Interaction (LORI)*, volume 11813 of *Lecture Notes in Computer Science*, pages 57–70. Springer, 2019.
- Y. D. Santos. A four-valued dynamic epistemic logic. *Journal of Logic, Language and Information*, 2020a.
- Y. D. Santos. Iterative social consolidations: Forming beliefs from many-valued evidence and peers’ opinions, 2020b.
- Y. D. Santos. Social consolidations: Rational belief in a many-valued logic of evidence and peerhood. In A. Herzig and J. Kontinen, editors, *Foundations of Information and Knowledge Systems*, pages 58–78, 2020c.
- Y. D. Santos, V. B. Matos, M. M. Ribeiro, and R. Wassermann. Partial meet pseudo-contractions. *International Journal of Approximate Reasoning*, 103:11 – 27, 2018.
- M. A. Satterthwaite. Strategy-proofness and Arrow’s conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10(2):187–217, 1975.
- F. F. Schmitt. Reliability, objectivity and the background of justification. *Australasian Journal of Philosophy*, 62(1):1–15, 1984.

- P. Schotch, J. Jensen, P. Larsen, and E. MacLellan. A note on three-valued modal logic. *Notre Dame Journal of Formal Logic*, 19(1):63–68, 1978.
- K. Segerberg. Some modal logics based on a three-valued logic. *Theoria*, 33(1):53–71, 1967.
- G. Shafer. *A mathematical theory of evidence*, volume 42. Princeton university press, 1976.
- C. Shi, S. Smets, and F. R. Velázquez-Quesada. Beliefs supported by binary arguments. *Journal of Applied Non-Classical Logics*, 28(2-3): 165–188, 2018a.
- C. Shi, S. Smets, and F. R. Velázquez-Quesada. Beliefs based on evidence and argumentation. In *International Workshop on Logic, Language, Information, and Computation (WoLLIC)*, pages 289–306. Springer, 2018b.
- T. J. Smiley. Entailment and deducibility. *Proceedings of the Aristotelian Society*, 59:233–254, 1959.
- A. Solaki. Steps out of logical omniscience. Master’s thesis, Universiteit van Amsterdam, 2017.
- R. C. Stalnaker. *Inquiry*. MIT Press, Cambridge, MA, 1984.
- K. Stenning and M. van Lambalgen. Semantics as a foundation for psychology: A case study of Wason’s selection task. *Journal of Logic, Language and Information*, 10(3):273–317, 2001.
- A. J. Stewart, M. Mosleh, M. Diakonova, A. A. Arechar, D. G. Rand, and J. B. Plotkin. Information gerrymandering and undemocratic decisions. *Nature*, 573(7772):117–121, 2019.
- P. F. Syverson and P. C. van Oorschot. On unifying some cryptographic protocol logics. In *Proceedings of 1994 IEEE Computer Society Symposium on Research in Security and Privacy*, pages 14–28. IEEE, 1994.
- A. Tamminga. *Belief Dynamics: (Epistemo)logical Investigations*. PhD thesis, ILLC, Amsterdam, 2001.
- S. Thomason. Possible worlds and many truth values. *Studia Logica*, 37(2):195–204, 1978.

- A. S. Troelstra and D. van Dalen. *Constructivism in Mathematics, Volume 1*. North Holland, Amsterdam, 1988.
- J. van Benthem. Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics*, 17(2):129–155, 2007.
- J. van Benthem. *Logical Dynamics of Information Flow*. Cambridge University Press, 2011.
- J. van Benthem. Oscillations, logic, and dynamical systems. In S. Ghosh and J. Szymanik, editors, *The Facts Matter. Essays on Logic and Cognition in Honour of Rineke Verbrugge*, pages 9–22. College Publications, 2015.
- J. van Benthem and E. Pacuit. Dynamic logics of evidence-based belief. Technical report, University of Amsterdam, ILLC, 2011a.
- J. van Benthem and E. Pacuit. Dynamic logics of evidence-based beliefs. *Studia Logica*, 99(1):61–92, 2011b.
- J. van Benthem, J. van Eijck, and B. Kooi. Logics of communication and change. *Information and Computation*, 204(11):1620–1662, 2006.
- J. van Benthem, D. Fernández-Duque, E. Pacuit, et al. Evidence logic: A new look at neighborhood structures. In T. Bolander, T. Braäuner, S. Ghilardi, and L. Moss, editors, *Advances in Modal Logic*, volume 9, pages 97–118. College Publications, 2012.
- J. van Benthem, D. Fernandez-Duque, and E. Pacuit. Evidence and plausibility in neighborhood structures. *Annals of Pure and Applied Logic*, 165(1):106–133, 2014.
- H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*, volume 337 of *Synthese Library*. Springer Science & Business Media, 2007.
- B. C. van Fraassen. *The Scientific Image*. Oxford University Press, 1980.
- F. R. Velázquez-Quesada. Inference and update. *Synthese*, 169(2):283–300, 2009.
- R. Verbrugge. Zero-one laws with respect to models of provability logic and two Grzegorczyk logics. In G. Bezhanishvili, G. D’Agostino, G. Metcalfe, and T. Studer, editors, *Advances in Modal Logic (AiML): Accepted Short Papers*, volume 12, pages 115–120, 2018.

- G. H. von Wright. *An Essay in Modal Logic*. North Holland, Amsterdam, 1951.
- P. C. Wason. Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 20(3):273–281, 1968.
- J. O. Weatherall, C. O’Connor, and J. Bruner. How to beat science and influence people: Policy makers and propaganda in epistemic networks. *arXiv preprint arXiv:1801.01239*, 2018.
- A. Whitehead and B. Russell. *Principia Mathematica*. Cambridge University Press, 1910.
- J. Williamson. *In Defence of Objective Bayesianism*. Oxford University Press, 2010.
- T. Williamson. *Knowledge and its Limits*. Oxford University Press, 2002.
- J. Zhang and F. Liu. Some thoughts on Mohist logic. In J. van Benthem, S. Ju, and F. Veltman, editors, *A Meeting of the Minds: Proceedings of the First Workshop on Logic, Rationality and Interaction (LORI)*, pages 85–102, London, 2007. College Publications.

Summary

With the advent of the internet and social media, we are exposed to more information than ever in human history. Being well-informed, however, is not an easy task, as the amount of untrustworthy information is also enormous. In our age, individuals need to be able to search, filter, combine and separate reliable from unreliable evidence, and draw sensible conclusions from it. All this has to be done with limited time and cognitive resources. The same hurdles are faced by software agents, which have an ever increasing presence in our times.

Rational agents, humans or otherwise, build their beliefs from evidence – a process which we call consolidation. But how should this process be carried out? In this thesis, we study a multi-agent logic of evidence and the question how agents should form beliefs in this logic.

We begin by formalising four-valued epistemic logic (FVEL), a multi-agent modal logic that describes public evidence and what agents know about this evidence. A key concept in this logic is that propositions can take four values: true, false, both or none. The meaning of these values is not ontic, but epistemic: *true* means that there is only positive evidence for a proposition, *false* means that there is only negative evidence (evidence against the proposition), *both* means that there are both types of evidence, and *none* means that there is no evidence at all. We offer a complete tableau proof system, prove basic properties of this logic, and also add public announcements to it.

Based on the formalism developed for FVEL, the next step is to think about how agents can use the evidence they have to form beliefs. Some principles are discussed and methods for forming beliefs, the so-called consolidations, are presented. The properties of this operation are studied, and it is then compared with an “implicit” consolidation found in another evidence logic in the literature (by van Benthem and Pacuit).

After this, another dimension of evidence is considered: social evidence, that is, evidence coming from peers. Until this point, the evidence other

agents have was not very relevant for each individual agent in its consolidation process. Now, each agent will take not only its own evidence into account, but also the evidence of its peers. Here a new interpretation of FVEL is employed, where each state represents one agent. Ideas from the social epistemology literature on peer disagreement are used to guide the principles for these social consolidations. A set of rationality postulates inspired also by Social Choice Theory is devised, and we then show that some consolidations fail these postulates, while others can satisfy all of them.

Finally, a modification of the previous social consolidations is explored. There, an agent could access the evidence of its peers, as if it were public. In this final part of the thesis, the evidence of each agent is considered to be private, and what is used instead is the beliefs or testimony of the peers, plus each agent's own evidence. This makes the consolidations iterative: in the first moment, each agent forms initial beliefs based on its own evidence only; subsequently, they change their beliefs based on their evidence plus their peers' beliefs; and so on. Eventually, the process may reach an equilibrium, or it can go on forever. This is the problem of stabilisation, and is the main topic explored in the thesis with respect to this type of consolidations.

To put it simply, the main contributions of this thesis are twofold. First, we present and study a many-valued modal logic, and show how it can be suitable for modelling multi-agent scenarios where each agent has access to some evidence, which in turn can be processed into beliefs. This is a technical and practical contribution to many-valued modal logics. Second, we open new paths for research in the field of evidence logics: we show a new approach based on many-valued logics, we highlight the concept of consolidations and the importance of looking at their dynamic nature, and build a methodology based on rationality postulates to evaluate them.

With this work, we hope to have advanced the knowledge on the field of logics of evidence and consolidations, giving new insights on the requirements of proper "epistemic machinery", which is essential for both human and artificial agents in dealing with the complex sea of information of the current era.

Samenvatting

Sinds de opkomst van het internet en sociale media worden we blootgesteld aan meer informatie dan ooit in de menselijke geschiedenis. Goed geïnformeerd blijven is echter geen gemakkelijke taak, omdat de hoeveelheid onbetrouwbare informatie enorm is. Tegenwoordig moeten individuen in staat zijn bewijsmateriaal te zoeken, te filteren en te combineren – terwijl betrouwbaar van onbetrouwbaar bewijsmateriaal wordt gescheiden – en er de juiste conclusies uit te trekken. Dit alles moet worden gedaan met beperkte tijd en cognitieve middelen. Software actoren, die een steeds grotere rol spelen in de huidige samenleving, worden met dezelfde hindernissen geconfronteerd.

Rationele actoren, menselijk of anderzins, baseren hun overtuigingen op bewijsmateriaal – een proces dat we consolidatie noemen. Maar hoe moet dit proces worden uitgevoerd? In dit proefschrift bestuderen we een multi-agent logica van bewijsmateriaal en de vraag hoe actoren overtuigingen zouden moeten vormen binnen deze logica.

We beginnen met het formaliseren van *four-valued epistemic logic* (FVEL), dat wil zeggen vierwaardige epistemische logica, een multi-agent modale logica die openbaar bewijsmateriaal, en wat actoren weten over dit bewijsmateriaal, weergeeft. Een kernidee in deze logica is dat proposities vier waarden kunnen aannemen: *waar*, *onwaar*, *beide* of *geen*. De betekenis van deze waarden is niet ontisch, maar epistemisch: *waar* betekent dat er alleen positief bewijsmateriaal is voor een propositie, *onwaar* betekent dat er alleen negatief bewijsmateriaal is (bewijsmateriaal tegen de propositie), *beide* betekent dat er beide soorten bewijsmateriaal zijn, en *geen* betekent dat er helemaal geen bewijsmateriaal is. We bieden een volledig tableau-bewijssysteem, bewijzen basiseigenschappen van deze logica en voegen er openbare aankondigingen aan toe.

Uitgaande van het formalisme ontwikkeld voor FVEL, is onze volgende stap om na te denken over hoe actoren beschikbaar bewijsmateriaal kunnen gebruiken om overtuigingen te vormen. Bepaalde principes worden be-

sproken en methoden voor het vormen van overtuigingen, de zogenaamde consolidaties, worden gepresenteerd. De eigenschappen van deze operatie worden bestudeerd en vervolgens wordt de operatie vergeleken met een “impliciete” consolidatie in een andere logica van bewijsmateriaal beschikbaar in de literatuur (ontwikkeld door Van Benthem en Pacuit).

Hierna wordt een andere dimensie van bewijs overwogen: sociaal bewijsmateriaal, dat wil zeggen, bewijsmateriaal van *peers* (d.w.z. epistemisch gelijkwaardige actoren). Tot dit punt was het bewijsmateriaal dat andere actoren hadden, niet erg relevant voor het consolidatieproces van elke afzonderlijke actor. Nu zal elke actor niet alleen rekening houden met zijn eigen bewijsmateriaal, maar ook met het bewijsmateriaal van zijn *peers*. Hier wordt een nieuwe interpretatie van FVEL gebruikt, waarbij elke toestand één actor vertegenwoordigt. Ideeën uit de sociaal-epistemologische literatuur over meningsverschillen tussen *peers* worden gebruikt als leidraad voor de principes voor deze sociale consolidaties. Een verzameling van rationaliteitspostulaten, geïnspireerd door ook de *Social Choice Theory*, wordt ontwikkeld, en we laten zien dat, terwijl sommige consolidaties niet voldoen aan deze postulaten, andere juist aan alle postulaten voldoen.

Ten slotte wordt een modificatie van de eerder genoemde sociale consolidaties onderzocht. Eerder kon een actor toegang krijgen tot het bewijs van zijn *peers* alsof het publiekelijk bekend was. In het laatste deel van het proefschrift wordt het bewijsmateriaal van elke actor als privé beschouwd, en wat in plaats daarvan wordt gebruikt, zijn de overtuigingen of getuigenissen van de *peers*, plus het eigen bewijsmateriaal van elke actor. Dit maakt de consolidaties iteratief: in eerste instantie vormt elke actor initiële overtuigingen op basis van alleen zijn eigen bewijsmateriaal; vervolgens veranderen de actoren hun overtuigingen op basis van hun bewijsmateriaal plus de overtuigingen van hun *peers*; enzovoorts. Dit proces kan uiteindelijk een evenwicht bereiken of het kan voor altijd doorgaan. Dit is het probleem van stabilisatie en dit is het belangrijkste onderwerp dat in dit proefschrift wordt onderzocht met betrekking tot dit soort consolidaties.

Kortom, de belangrijkste bijdrage van dit proefschrift is tweeledig. Eerst presenteren en bestuderen we een meerwaardige modale logica en laten we zien hoe deze geschikt kan zijn voor het modelleren van multi-agent scenario's waarbij elke actor toegang heeft tot enig bewijs, dat kan worden verwerkt tot overtuigingen. Dit is een technische en praktische bijdrage aan meerwaardige modale logica. Ten tweede openen we nieuwe wegen voor onderzoek op het gebied van de logica van bewijsmateriaal: we tonen een nieuwe aanpak op basis van meerwaardige logica, we benadrukken het

concept van consolidaties en het belang om naar hun dynamische aard te kijken, en we bouwen een methodologie op basis van rationaliteitspostulaten om ze te evalueren.

Met dit proefschrift hopen we de kennis op het gebied van de logica van bewijsmateriaal en consolidaties te hebben verruimd door nieuwe inzichten te hebben geboden in de vereisten van juiste “epistemische mechanieken”, wat essentieel is voor zowel menselijke als kunstmatige actoren om zich een weg te banen door de complexe zee van informatie van het huidige tijdperk.

Resumo

Com o advento da internet e das mídias sociais, estamos expostos a mais informações do que nunca na história da humanidade. Estar bem informado, porém, não é uma tarefa fácil, pois a quantidade de informações não-confiáveis também é enorme. Na nossa era, indivíduos precisam ser capazes de pesquisar, filtrar, combinar e separar evidências confiáveis das não-confiáveis e tirar conclusões sensatas delas. Tudo isso deve ser feito com tempo e recursos cognitivos limitados. Os mesmos obstáculos são enfrentados por agentes de software, que têm uma presença cada vez maior hoje em dia.

Os agentes racionais, humanos ou não, constroem suas crenças a partir de evidências – um processo que chamamos de consolidação. Mas como esse processo deve ser realizado? Nesta tese, estudamos uma lógica de evidência multi-agentes e a questão de como os agentes devem formar crenças nessa lógica.

Começamos formalizando a lógica epistêmica de quatro valores (FVEL), uma lógica modal multi-agentes que descreve a evidência pública e o que os agentes sabem sobre essa evidência. Um conceito chave nesta lógica é que as proposições podem assumir quatro valores: verdadeiro, falso, ambos ou nenhum. O significado desses valores não é ôntico, mas epistêmico: *verdadeiro* significa que há apenas evidências positivas para uma proposição, *falso* significa que há apenas evidências negativas (evidências contra a proposição), *ambos* significa que existem ambos os tipos de evidências, e *nenhum* significa que não há nenhuma evidência. Oferecemos um sistema completo de prova por *tableaux*, provamos as propriedades básicas desta lógica e também adicionamos anúncios públicos (*public announcements*) a ela.

Com base no formalismo desenvolvido para a FVEL, o próximo passo é pensar sobre como os agentes podem usar as evidências de que dispõem para formar crenças. Alguns princípios são discutidos e métodos para formar crenças, as chamadas consolidações, são apresentados. As propriedades

dessa operação são estudadas e depois comparadas com uma consolidação “implícita” encontrada em outra lógica de evidências na literatura (de van Benthem e Pacuit).

Depois disso, outra dimensão da evidência é considerada: evidência social, ou seja, evidência vinda de pares. Até o momento, evidência de outros agentes não era muito relevante para cada agente individualmente em seu processo de consolidação. Agora, cada agente levará em consideração não apenas sua própria evidência, mas também a de seus pares. Aqui, uma nova interpretação da FVEL é empregada, onde cada estado representa um agente. Ideias da literatura de epistemologia social sobre desacordo entre pares (*peer disagreement*) são usadas para guiar os princípios dessas consolidações sociais. Um conjunto de postulados de racionalidade inspirados também pela Teoria da Escolha Social (*Social Choice Theory*) é elaborado, e então mostramos que algumas consolidações falham nesses postulados, enquanto outras podem satisfazer todos eles.

Finalmente, uma modificação das consolidações sociais anteriores é explorada. Naquelas, um agente podia acessar as evidências de seus pares, como se fossem públicas. Nesta parte final da tese, a evidência de cada agente é considerada privada, e o que é usado em vez disso são as crenças ou o testemunho dos pares, mais a própria evidência de cada agente. Isso torna as consolidações iterativas: num primeiro momento, cada agente forma crenças iniciais com base apenas em suas próprias evidências; subsequentemente, eles mudam suas crenças com base em suas evidências mais as crenças de seus pares; e assim por diante. Eventualmente, o processo pode atingir um equilíbrio ou pode durar para sempre. Este é o problema da estabilização, e é o principal tópico explorado na tese com respeito a este tipo de consolidações.

Resumidamente, as principais contribuições desta tese são duas. Primeiro, apresentamos e estudamos uma lógica modal multi-valorada e mostramos como ela pode ser adequada para modelar cenários multi-agentes onde cada agente tem acesso a alguma evidência, que por sua vez pode ser processada e transformada em crenças. Esta é uma contribuição técnica e prática para lógicas modais multi-valoradas. Em segundo lugar, abrimos novos horizontes para a investigação no campo das lógicas de evidências: mostramos uma nova abordagem baseada em lógicas multi-valoradas, destacamos o conceito de consolidações e a importância de olhar para a sua natureza dinâmica e construímos uma metodologia baseada em postulados de racionalidade para avaliá-las.

Com este trabalho, esperamos ter avançado o conhecimento sobre o

campo de lógicas de evidências e consolidações, trazendo novos insights sobre os requisitos de um “maquinário epistêmico” adequado, que é essencial para agentes humanos e artificiais para lidar com o complexo mar de informações da era atual.