Stable Beliefs and Conditional Probability Spaces

MSc Thesis *(Afstudeerscriptie)*

written by

Konstantinos Gkikas
(born 16th of December 1988 in Athens, Greece)

under the supervision of **Dr. Alexandru Baltag**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:	**Members of the Thesis Committee:**
3 September 2015	Dr. Alexandru Baltag
	Prof. Dr. Johan van Benthem
	Prof. Dr. Jan van Eijck
	Dr. Jakub Szymanik

Institute for Logic, Language and Computation

Abstract

This thesis aims to provide a conceptual framework that unifies Leitgeb's theory of stable beliefs, Battigalli-Siniscalchi's notion of strong belief and the notions of core, a priori, abnormal and conditional belief studied by Van Fraassen and Arló-Costa.

We will first present the difficulties of modeling qualitative notions of belief and belief revision in a quantitative probabilistic setting. On one hand we have the probability 1 proposal for belief, which seems to be materially wrong and on the other hand we have the Lockean thesis (or any version of it) which deprives us of the logical closure of belief (Lottery Paradox). We will argue that Leitgeb's ([35]) theory of stability of belief is a path between this Scylla and Charybdis.

The first goal of this thesis is to provide an extension of Leitgeb's theory into non-classical probability settings, where we can condition on events with measure 0. We will define the notion of r-stable sets in Van Fraassen's setting ([25]), using two-place functions to take conditional probability as primitive. We will then use the notion of r-stability to provide a definition of conditional belief similar to Leitgeb's.

The second goal of this thesis is to develop a formal language that will express the notion of conditional beliefs. To do so, we will first define the structures called conditional probabilistic frames that will give us the semantics for the logic of conditional beliefs (**rCBL**) that we will present.

Finally, we will use the operators $\Box$ (safe belief) and C (certainty) to lay the foundations of developing a logic of stable beliefs.

Acknowledgements

I would first like to thank my supervisor Alexandru Baltag for his help, patience and guidance. I deeply admire his enthusiasm and love for what he does and I am truly grateful that he was always there, finding time to meet and work with me even in really hectic and busy periods.

I am also indebted to Prof. Dr. Johan van Benthem, Prof. Dr. Jan van Eijck and Dr. Jakub Szymanik for taking the time to read my thesis.

Moreover, thanks goes to Paolo Galeazzi for the great project that he organized, his excellent supervision and all the football matches he arranged during the past two years.

On a personal note, Marcel Proust wrote "Let us be grateful to the people who make us happy; they are the charming gardeners who make our souls blossom". In this spirit, I would like to thank Fernando, Nigel, Julian, KH, Paolo, Rick, Pablo, Dan, Pietro, Eileen, Ana and all the rest of my fellow logic students for filling the past two years with the most wonderful memories. Special thanks goes to Aldo and Moo for being the best friends one could hope for. I would also like to thank Lorena for her support throughout the past two years.

Finally, this thesis is dedicated to Christina; her smile alone was all the inspiration I needed.

Contents

CHAPTER 1

Introduction

1. Introductory remarks

The general approach to formalizing belief in a quantitative probabilistic manner is the probability 1 principle. According to this principle, an agent believes a proposition if he assigns probability 1 to it. Van Fraassen argued that this principle does not allow for distinctions to be drawn among the maximally likely propositions ([25]). Furthermore, Leitgeb claims that this principle is materially wrong ([34, p. 1344]) since it identifies "believing" in a proposition A with "being certain" of proposition A.

However, theories attempting to provide a quantitative probabilistic definition of belief come face to face with the fact that drifting away from the probability 1 principle (or in other words adapting a version of the so - called Lockean Thesis ([24])) is incompatible with maintaining the logical closure of belief. Any theory that attempts to do so is confronted by the lottery paradox.

In his papers [34], [35] Leitgeb develops a theory which avoids this problem by incorporating a notion of stability. He does not equate belief with probability one and at the same time maintains its logical closure without running into the lottery paradox, as long as rational belief is equivalent to the assignment of a stably high rational degree of belief ([35, p. 1]). We will argue that Leitgeb's theory gives us a more intuitive notion of belief, in comparison to the dominant probability 1 principle. Moreover, by adopting Leitgeb's theory we also maintain the logical closure of belief — as Leitgeb shows the lottery paradox can be avoided — and therefore it seems to be a win - win situation. Leitgeb develops his theory in a classical probability setting. We will argue that classical probabilities are unable to do belief revision. This is because conditioning on events of measure 0 can not be defined in a classical probabilistic setting.

In this thesis, we will extend Leitgeb's theory into conditional probability spaces, where conditioning on events with measure 0 is allowed, having as our final goal the use of Leitgeb's conditional and r-stable beliefs in epistemic game theory. More specifically, we will work in Van Fraassen's framework as developed in [25], [3], [21], defining the notion of r-stable beliefs in a way (almost) identical to Leitgeb. Furthermore, we will define the structures called probabilistic frames and an operator for conditional belief.

In the introduction of his paper [18], Board argues that epistemic logic is the path between the obscurity of Aumann structures and the complexity of belief hierarchies. He argues that a logical approach offers simplicity and transparency. Simplicity, due to the semantic structures used to provide truth conditions for formulas of the formal language; structures which can be easily adapted to provide epistemic models for games. And transparency because of the straightforward interpretation of the language and axiom system that provide the syntax of the logic ([18, p. 53]).

Following Board, we will present our logic of conditional belief. The language of this logic will be the language of epistemic logic, augmented by adding the modal operator $B_i^{\phi}\psi$ that will stand for "agent i's belief in ψ conditioning on ϕ". Our axiom system will be similar to Board's ([18]). Finally, our semantics will be given by the structures called probabilistic models. We will prove soundness and completeness of our logic with an interesting completeness proof in which we will represent r-stable beliefs as Grove spheres, obtain a probabilistic model from one of Board's structures and prove a truth-preserving lemma between Board's semantics and ours. After this lemma, completeness of our axioms w.r.t. Board's semantics will give us completeness w.r.t our probabilistic semantics.

However, the language of this logic will be unable to express statements such as: "agent i has an r-stable belief that agent j ...". This in turn prevents us from being able to define notions such as Common or Mutual r-stable belief in rationality. A straightforward introduction of an operator in order to express r-stable beliefs syntactically would not be a good idea. This is because such an operator would not satisfy the **K**-axiom. Therefore, we will introduce two other modalities that satisfy the **K**-axiom and express our operator for r-stable beliefs in terms of those.

We will turn to Baltag and Smets' *safe belief* modality: $\Box$ ([4]) and we will also define the operator C to express *certainty.* However, in contrast to Baltag and Smets, our safe belief and certainty modalities are not going to be truthful, as explained in chapter 8. With these operators at hand, we will be able to define a notion of stability of belief, minimally different from Leitgeb's: this notion will be called *quasi*-stability. Finally, in the conclusion of this thesis, we will propose an axiomatization of these modalities hoping to develop a logic of certainty and safe belief in a future paper.

2. Organization of the Thesis

The thesis is organized as follows.

In chapter 2 we will give a brief presentation of all the background work. We will first discuss Leitgeb's stability theory of belief in a classical probability setting. Furthermore, we will also present Battigalli-Siniscalchi's notions of conditional and strong belief.

In chapter 3 we will develop our framework. We will work in Van Fraassen's setting, taking conditional probability functions as primitive. Moreover, we will use Van Fraassen's notions of a priori, normal and abnormal sets and prove useful properties about them.

In chapter 4 we will define our notion of r-stable beliefs. Our definition will be similar to Leitgeb's, but we will allow conditioning on sets with measure 0 as well. We will also prove some important properties of r-stable sets. We will show that our r-stable sets are well-founded w.r.t. inclusion, as long as they are not a priori and we will also show that they are well-ordered w.r.t. a new relation we will define, the *quasi*-subset relation.

In chapter 5 we will define our notion of conditional belief, based on the notion of r-stability. We will also prove that our conditional belief operator is closed under conjunction and is consistent w.r.t. normal sets.

In chapter 6 we will define the structures called probabilistic frames. We will also make a brief comparison of our stable beliefs with Battigalli-Siniscalchi's strong belief.

In chapter 7 we will present the logic of conditional belief. Along with soundness, we will also provide a completeness proof based on Board's *belief revision structures* ([18]).

In chapter 8, we will define two new operators: $\Box$ (safe belief) coming from Baltag and Smets ([4]) and C (certainty). Moreover, we will define a new notion of r-stability, that of *quasi*-stability. Finally, we will show that with certain restrictions it is possible to express both *quasi*-stability and conditional belief in terms of $\Box$ and C.

Finally, in chapter 9 we will present a possible axiomatization of $\Box$ and C and discuss about future work and possible applications of r-stable beliefs.

CHAPTER 2

Background

1. Probability 1 vs Lockean Thesis

It is well known that there are severe strains between probabilities and belief ([25]). As Van Fraassen writes ([25]) "they (probability and belief) seem too intimately related to exist as separate but equal; yet if either is taken as the more basic, the other may suffer." In this section of the thesis, we will explain why modeling qualitative notions of belief and belief revision in a quantitative probabilistic setting is far from trivial.

By *qualitative belief*, we usually mean belief in the sense that either it is believed that A is the case, or that $\neg A$ is the case, or that neither of these is the case, i.e. that our agent is agnostic w.r.t. A.

By *quantitative belief*, we refer to the assignment of a degree (a numerical degree, for example a probability number) of belief to propositions, degrees which serve as measures of the "strength" of an agent's belief in a proposition. Typically, believing a proposition A with a degree of 1 means that the agent is certain that A is true, while a degree of 0 means that the agent is certain that $\neg A$ is true, i.e. that A is false [34, p. 1339].

On the qualitative side, doxastic logic usually assumes the **KD45** axioms for belief:

K $B(\phi \to \psi) \Rightarrow (B\phi \to B\psi)$
D $\neg B(\phi \wedge \neg\phi)$
4 $B\phi \Rightarrow BB\phi$
5 $\neg B\phi \Rightarrow B\neg B\phi$

with B being the modal operator for belief.

The **K** axiom implies that the agent will believe all the logical consequences of his beliefs. Axiom **D** is also called "Consistency of Belief" and it tells us that the agent can not believe the contradictory proposition. Axiom **4** is also called "Positive Introspection" and it tells us that an agent believes his own beliefs. Finally, axiom **5** is also known as "Negative Introspection" and it tells us that if an agent does not believe ϕ then he believes that he does not believe ϕ.

On the quantitative side now, we have probability functions. According to the Bayesian view degrees of belief obey axioms of the probability calculus. Therefore, we ascribe probability numbers (as mentioned above) to sentences or formulas ([34, p. 1343]). For that, we need a probability function P:

$$P : \mathcal{L} \to [0, 1]$$

accompanied by certain axioms and rules.
E.g. if $A \in \mathcal{L}$ is a proposition which is logically true, then $P(A) = 1$.

Moreover, most writers also assume finite additivity, i.e.:
If A, B are inconsistent with each other, then:

$$P(A \vee B) = P(A) + P(B)$$

The two most popular proposals for bridging principles relating qualitative and quantitative belief are the following:

1. Probability 1 proposal:

$$B(A) \text{ iff } P(A) = 1$$

2. Lockean Thesis:

$$B(A) \text{ iff } P(A) \geq r, \text{ for some } r \in (\tfrac{1}{2}, 1).$$

The first principle is one of the standard ways of defining belief in epistemic game theory ([12], [13]). According to principle **1.**, an agent believes the proposition A if and only if he assigns probability 1 to it. Now using this interpretation of belief, one is able to derive all the classical doxastic axioms presented above. However, this definition of belief seems — as Leitgeb puts it ([34, p. 1344]) — *materially wrong*. Van Fraassen argues ([25, p. 2]) that this principle "treats belief as on a par with tautologies" and there is no distinction among the maximally likely propositions. Imagine for example that Aldo rationally believes that his friend Moritz is going to call him. In that case, Aldo would (probably) refrain from accepting a bet in which he would win one drachma if he was right and lose 100.000 euros if he was wrong. However, given that he believes this proposition to the degree of 1, he should be eager to accept such a bet. Hence, it is possible to rationally believe a proposition with a probability less than 1 and such a case is out of the reach of this principle.

The second principle, was called "Lockean Thesis" by Richard Foley in [24] and identifies belief with a high degree of probability for some threshold $r \in (\frac{1}{2}, 1)$. The big advantage of this principle is that it presents a much more intuitive definition of belief, allowing us to draw a distinction between being *certain* of A ($P(A) = 1$) and *believing* A ($P(A) > \frac{1}{2}$), in the sense of considering A to be more probable than its negation. However, this approach has its own problems as well. Adopting the Lockean Thesis, one has to give up the logical closure of belief unless $r = 1$. The argument is the infamous Lottery paradox ([25], [8]). In a lottery of 1.000 tickets believed to be fair, one agent believes with degree 0,999 that *each* single ticket is not the winning one. But then the agent should not believe with degree 0,999 the conjunction of all these, i.e.: that no ticket is the winning one.

Therefore, as Leitgeb puts it ([34, p. 1345]) the probability 1 proposal is logically fine but materially wrong, while the Lockean Thesis seems to be materially fine but logically wrong.

2. Leitgeb's Stability Theory

In his Stability Theory of Belief ([35]), Leitgeb argues that: "it might be possible to have one's cake and eat it too" ([35, p. 5]). Leitgeb's formalization of belief is based on what he calls the "Humean conception of belief" and on a modified Lockean Thesis. He argues that according to Loeb ([36]), Hume held the following view in his Treatise of Human Nature: "it is rational to believe a proposition in case it is rational to have a stably high degree of belief in it" ([35, p. 33]). The key-word here (that is absent in the Lockean thesis) is *stable*. What does it mean to have a *stably* high degree of belief in a proposition in a quantitative probabilistic setting? Leitgeb provides the following formalization of this view.

His stability theory of belief consists of the following three principles ([35, pp. 7,13]). Take W a finite set of possible worlds and:

P1 There is a (uniquely determined) consistent proposition $B_W \subseteq W$, such that for all propositions A we have: $B(A)$ iff $B_W \subseteq A$.

P2 Function P is a classical probability measure over $\mathcal{P}(W)$, the powerset of W, i.e.: for all $A, B \subseteq W$:
- $P(W) = 1$,
- if A is inconsistent with B, then $P(A \vee B) = P(A) + P(B)$,
- $P(B|A) = \frac{P(B \cap A)}{P(A)}$

P3 For all $A \subseteq W$ such that B_W is consistent with A and $P(A) > 0$ we have $P(B_W|A) > \frac{1}{2}$ and if $P(B_W) = 1$, then B_W is the least proposition $A \subseteq W$ with $P(A) = 1$.

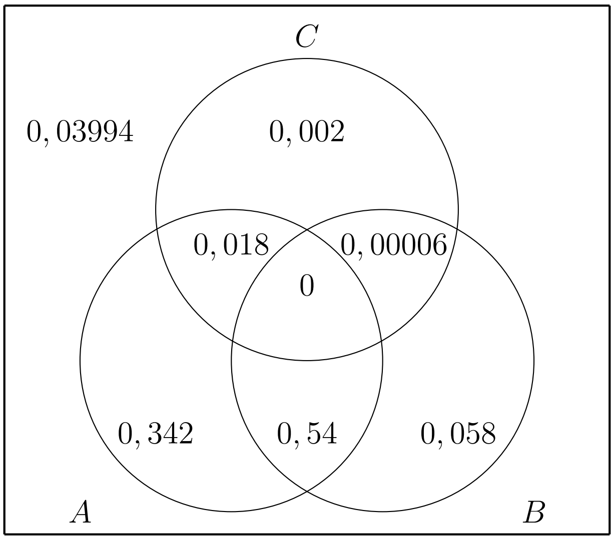
Principle **P1** tells us that there is a proposition B_W that is consistent, which has the property that all its supersets (including B_W as well) are believed. The obvious question here is: what determines B_W and how does it have this special property? The next principle **P2**, essentially defines a classical probability measure P. Finally in the last principle **P3**, this measure P is used to specify the proposition B_W. It tells us that B_W has the property that it maintains a high degree of probability, while conditioning on any proposition A that is first of all consistent itself (meaning has a probability higher than 0) and second it is consistent with B_W. Moreover, if B_W has a probability equal to 1, then it is the smallest such proposition.
What is important to note here is that principle **P3** includes Leitgeb's definition of a P^r-stable set, which is the core of his theory of belief.

Definition 2.1. *Let P be a classical probability measure on $\mathcal{P}(W)$ and let $r \in [\frac{1}{2}, 1)$ a number. A set $A \in \mathcal{P}(W)$ is P^r-stable if $P(A|B) > r$, for all $B \in \mathcal{P}(W)$ such that $A \cap B \neq \emptyset$ and $P(B) > 0$.*

If we now think of $P(A|B)$ as the degree of A under the supposition of B, then a P^r-stable proposition has the property that whatever proposition B that is consistent with proposition A and has a probability bigger than 0 (i.e. is considered probable) is supposed, the probability of proposition A will be bigger than $\frac{1}{2}$, i.e. bigger than its negation. Therefore the idea here is that a proposition is *stable* if it is considered more probable than its negation, whatever consistent comes its way. Moreover, it is also important to notice that a P^r-stable set also has a probability higher than $\frac{1}{2}$ itself, since we can always conditionalize on W, the set of all possible worlds ([34, p. 1359]).

The following example ([34, p. 1348],[35, p. 4]) will give a better idea of P^r-stability.

Example 1. Take W a set of possible worlds: $W = \{w_1, w_2, w_3, ..., w_8\}$. Take P a classical probability measure on $\mathcal{P}(W)$, such that:
$P(\{w_1\}) = 0,54$, $P(\{w_2\}) = 0,342$, $P(\{w_3\}) = 0,058$, $P(\{w_4\}) = 0,03994$, $P(\{w_5\}) = 0,018$, $P(\{w_6\}) = 0,002$, $P(\{w_7\}) = 0,00006$, $P(\{w_8\}) = 0$.
Now take $r = \frac{1}{2}$.
We can think of these eight possibilities as descriptions built from 3 propositions: A, B, C.
Take w_1 to correspond to $A \wedge B \wedge \neg C$, w_2 to $A \wedge \neg B \wedge \neg C$, w_3 to $\neg A \wedge B \wedge \neg C$, w_4 to $\neg A \wedge \neg B \wedge \neg C$, w_5 to $A \wedge \neg B \wedge C$, w_6 to $\neg A \wedge \neg B \wedge C$, w_7 to $\neg A \wedge B \wedge C$ and w_8 to $A \wedge B \wedge C$.
This probability space looks like this:



Now take $r = 3/4$ to be our threshold.
Then we get the following $P^{3/4} - stable$ sets:
$\{w_1, ..., w_5\}$, $\{w_1, ..., w_6\}$, $\{w_1, ..., w_7\}$, $\{w_1, ..., w_8\}$.
Consider for example the set $\{w_1, ..., w_5\}$.
This set is P^r-stable because for all Y such that $Y \cap \{w_1, ..., w_5\}$ and $P(Y) > 0$ we have that $P(\{w_1, ..., w_5\}|Y) > \frac{3}{4}$.

And now we turn back to principle **P1** to obtain Leitgeb's definition of belief:
For some $H \subseteq W$ and a classical probability measure $P : \mathcal{P}(W) \rightarrow [0, 1]$, we have that:

$$B(H) \text{ holds if and only if } \exists S : P^r\text{-stable set such that } S \subseteq H.$$

Therefore, the agent believes a proposition, if it is entailed by one of his stable beliefs.

Now Leitgeb shows that his theory solves the Lottery Paradox ([35, p. 25,26]) and he essentially proves the soundness of the **KD45** axiom system. Therefore he full-fills his promise that there might be a way to have the cake and eat it too.
He shows that his stable beliefs have some interesting properties, e.g. that they are nested

and well-founded w.r.t. inclusion. This in turn entails that we can take stable beliefs as our Grove spheres for a sphere model for belief revision, a fact that we will use to provide one completeness proof later on in this thesis.

Leitgeb develops his theory and provides all these results in a classical probability setting (Kolmogorov's axioms). However, this setting is too restrictive. This is because in a classical probability setting conditioning on events with measure 0 is not defined. Suppose for example that we have a space W and $A, B \subseteq W$ such that $P(A) = 0$. Then $P(B|A) = \frac{P(A \cap B)}{P(A)}$, a fraction that is not defined since the denominator is 0. Now this restriction poses an issue in belief revision, since as Halpern writes "That makes it unclear how to proceed if an agent learns something to which she initially assigned probability 0" ([29]). Hence when an agent is confronted with the occurrence of an event she considered impossible, instead of revising her beliefs she would raise her hands in despair. As Baltag and Smets write "it is well known that simple (classical) probability measures yield problems in the context of describing an agent's beliefs and how they can be revised"([8, p. 3]). Although consideration of events with measure 0 might seem to be of little interest, Halpern argues ([29, pp.1,2]) that it plays an essential role in game theory, "particularly in the analysis of strategic reasoning in extensive form games and in the analysis of weak dominance in normal form games". See for example (among others) [12], [13], [20], [16], [17], [31], [30]. Moreover, conditioning on events with measure 0 is also crucial in the analysis of conditional statements in philosophy ([1], [37]) and in dealing with monotonicity in AI ([33]). One way of preempting this problem without giving up classical probability is to demand that only impossible events can have probability 0. However, this entails that agents never have any wrong beliefs about anything, which seems to be a rather severe constraint. Hence Baltag and Smets' conclusion that classical probabilities can not deal with any non-trivial belief revision ([8, p. 3]) seems indeed correct. And this is the main idea that motivates this thesis: to provide an extension of Leitgeb's theory into non-classical probability spaces.

3. Conditioning on sets of measure 0

In [29] J. Halpern provides an excellent overview of the three most popular approaches of dealing with conditioning on sets of measure 0: conditional probability spaces (CPS's), nonstandard probability spaces (NPS's) and lexicographic probability systems (LPS's).

The idea behind CPS's goes back to Popper ([38]) and de Finetti ([22]) and essentially is to take conditional probability as primitive. This allows for the measure $\mu(V|U)$ to be defined even if $\mu(U|W) = 0$, for W a space and $U, V \subseteq W$ ([29, pp. 2,3]).

In nonstandard probability spaces, the idea is that there are infinitesimals that can be used to model events that might have infinitesimally small probability but are still observable ([29, pp. 2,3]). This idea goes back to [40] and has been used in economics ([31], [30]), in AI ([33]) and in philosophy ([1], [37]).

Finally, lexicographic probability systems were recently introduced by Blume, Brandenburger and Dekel ([16, 17]).

A lexicographic probability system is a sequence $(\mu_0, \mu_1, \mu_2, ...)$ of probability measures. The idea is that the first measure of this sequence μ_0 is the most important one, then μ_1 is the next and so on. The probability assigned to some event E by (μ_0, μ_1) is $\mu_0(E) + \epsilon\mu_1(E)$ for some infinitesimal ϵ. Then even if $\mu_0(E) = 0$, E still has a positive probability given of course that $\mu_1(E) > 0$ ([29, pp. 2,3]).

Halpern also provides some very interesting equivalence results between these three approaches. In particular Hammond ([31]) shows that CPS's are equivalent to a subclass of LPS's, called: lexicographic conditional probability spaces (LCPS's) as long as the state space is finite and conditioning on any nonempty set is possible ([29, pp. 2,3]).

On the other hand, Halpern shows that if the state space is finite, NPS's are equivalent to LPS's ([29]). For more on these results, we refer the reader to [29], [31].

In this thesis, we adopt the Popper-Renyi theory of conditional probability systems (CPS's) ([25], [8], [39],[26], [38]), taking conditional probability as primitive.

More specifically, we will adopt Van Fraassen's setting as developed in [25], [3], [21].

4. Battigalli-Siniscalchi

In their papers [12], [13] Battigalli and Siniscalchi develop their notions of conditional and strong belief using conditional probability systems. Their theory follows the probability 1 principle of quantitative representation of belief, equating "believing H given E" with "assigning probability 1 to H when conditioning on E". Throughout the thesis, we will be comparing (whenever possible) our setting and theories of conditional and r-stable beliefs with B-S's work. In particular, we will see that there is a direct analogy between B-S's work and ours, since we will show that our notion of r-stability can express B-S's notion of strong belief. Furthermore, this comparison shows that there is a tight connection between the theory presented in this thesis and B-S's work, a connection that opens directions for future applications of stable beliefs in epistemic game theory, as we will discuss in chapter 9.

In this section, we will present B-S's notions of conditional and strong belief as developed in [12], [13].

To do so, we will need to define a lot of notions from the field of epistemic game theory (strategies, payoff types, type spaces, rationality and so on). The reader that is already familiar with these notions is asked to jump directly to Definition 2.7. On the other hand, the reader that is not acquainted with these concepts at all should be aware that we will only give a very brief presentation and is referred to [12], [13] among others.

Consider a set $I = \{1, ...|I|\}$ of players, a finite collection $\mathcal{H}$ of non-terminal histories, including the empty history $\emptyset$ and a finite collection of terminal histories $\mathcal{Z}$. Moreover, Θ_i is a finite collection of payoff types for each player $i \in I$ and $u_i : \mathcal{Z} \times \Theta \to \mathbb{R}$, where $\Theta = \Theta_1 \times \Theta_2 \times ... \times \Theta_I$.

Now each element $\theta_i \in \Theta_i$ represents player i's information about the payoff-aspects of the game. If the set Θ contains only one element, then the game has complete information.

At each stage of the game, all players are informed about the history that just occurred. However, no one is informed about each others payoff types. Therefore, each player's actions depend on previous histories that have occurred, but not on his information θ_i. We will denote the set of actions of player i depending on previous history $h \in \mathcal{H}$ as: $A_i(h)$. If there is only one active player at each $h \in \mathcal{H}$, we say that the game has perfect information.

For every $i \in I$, we will use S_i to denote the set of strategies available to him. A strategy is defined as a function $s_i : \mathcal{H} \to \bigcup\limits_{h \in \mathcal{H}} A_i(h)$, with $s_i(h) \in A_i(h)$ for all h. Now $S = \Pi_{i \in I} S_i$ and $S_{-i} = \Pi_{j \neq i} S_j$.

For any $h \in \mathcal{H} \cup \mathcal{Z}$, $S(h)$ denotes the set of strategy profiles which induce history h. Its projections on S_i and S_{-i} are denoted by $S_i(h)$ and $S_{-i}(h)$ respectively.
Now we also use $\Sigma_i = S_i \times \Theta_i$ to denote the set of strategy-payoff type pairs for player i and we let $\Sigma = \Pi_{i \in I} \Sigma_i$ and $\Sigma_{-i} = \Pi_{j \neq i} \Sigma_j$.
Finally, we have the notation required to define a payoff function $U_i : \Sigma_i \times \Sigma_{-i} \to \mathbb{R}$ as usually: for all $z \in \mathcal{Z}$, $(s_i, \theta_i) \in \Sigma_i$ and $(s_{-i}, \theta_{-i}) \in \Sigma_{-i}$, if $(s_i, s_{-i}) \in S(z)$, then $U(s_i, \theta_i, s_{-i}, \theta_{-i}) = u_i(z, (\theta_j)_{j \in I})$.
Finally, let $\mathcal{H}(s_i) = \{h \in \mathcal{H} : s_i \in S_i(h)\}$ denote the collection of histories consistent with s_i.

For a given measure space $(X_i, \mathcal{X}_i)$, consider a non-empty collection $\mathcal{B}_i \subseteq \mathcal{X}_i$ of events such that $\emptyset \notin \mathcal{B}_i$.
The collection $\mathcal{B}_i$ is a collection of observable events concerning x, which $x \in X_i$ is the "true" element (real world).

Definition 2.2. *A* conditional probability system (or CPS) *on $(X_i, \mathcal{X}_i, \mathcal{B}_i)$, is a mapping $\mu(\cdot|\cdot) : \mathcal{X}_i \times \mathcal{B}_i \to [0, 1]$ such that, for all $B, C \in \mathcal{B}_i$ and $A \in \mathcal{X}_i$, we have that:*

- *$\mu(B|B) = 1$,*
- *$\mu(\cdot|B)$ is a probability measure on $(X_i, \mathcal{X}_i)$,*
- *$A \subseteq B \subseteq C$ implies $\mu(A|B)\mu(B|C) = \mu(A|C)$.*

Denote the set of conditional probability systems on $(X_i, \mathcal{B}_i)$ by $\Delta^{\mathcal{B}_i}(X_i)$.
Now B-S obtain what they call player i's *first order (conditional) beliefs* about her opponent's behavior and payoff types by taking $X_i = \Sigma_{-i}$ and $\mathcal{B}_i = \{B \subseteq \Sigma_{-i} : B = S_{-i}(h) \times \Theta_{-i} \text{ for } h \in \mathcal{H}\}$. We denote the collection of CPSs defined on $(\Sigma_{-i}, \mathcal{B}_i)$ by $\Delta^{\mathcal{H}}(\Sigma_{-i})$. Now to obtain player i's *higher-order beliefs*, B-S introduce the notion of an extensive-form type space ([13, p.: 361]). Each agent's epistemic type $t_j \in T_j$ is used to parametrize her conditional beliefs. Therefore, a state of the world is an array $\omega = (\omega_j)_{j \in I} = (s_j, \theta_j, t_j)_{j \in I}$ of srategies, payoff types and epistemic types. We now consider a set of possible worlds $\Omega = \Pi_{j \in I} \Omega_j \subseteq \Pi_{j \in I}(\Sigma_j \times T_j)$. Player i has conditional

beliefs about the strategies, payoff types and epistemic types of her opponents.
Hence, we specify the structure $(X_i, \mathcal{B}_i)$ as: $X_i = \Pi_{j\neq i}\Omega_j = \Omega_{-i}$ and $\mathcal{B}_i = \{B \in \mathcal{X}_i : B = \{(s_{-i}, \theta_{-i}, t_{-i}) \in \Omega_{-i} : s_{-i} \in S_{-i}(h)\}$ for $h \in \mathcal{H}\}$. We use $\Delta^{\mathcal{H}}(\Omega_{-i})$ to denote the set of CPSs on $(\Omega_{-i}, \mathcal{B}_i)$.

And now we define the structure called type space:

DEFINITION 2.3. ([13],[14]) *A* type space *on* $(\mathcal{H}, S(\cdot), \Theta, I)$ *is a tuple*

$$\mathcal{T} = (\mathcal{H}, S(\cdot), \Theta, I, (\Omega_i, T_i, g_i)_{i\in I})$$

such that for every $i \in I$, T_i *is a compact topological space and:*

- Ω_i *is a closed subset of* $\Sigma_i \times T_i$ *such that* $proj_{\Sigma_i}\Omega_i = \Sigma_i$,
- $g_i = (g_{i,h})_{h\in\mathcal{H}} : T_i \to \Delta^{\mathcal{H}}(\Omega_{-i})$ *is a continuous mapping.*

Notice that $g_i(t_i) = (g_{i,h}(t_i)_{h\in\mathcal{H}})$ is agent i's conditional probability system. This according to Battigalli and Siniscalchi denotes the beliefs of the epistemic type t_i. Therefore, at any possible world $\omega = (s_i, \theta_i, t_i)_{i\in I} \in \Omega$ we specify player i's strategy (s_i), her disposition to believe $(g_i(t_i))$ and her payoff type θ_i ([13, p.: 362]).

The natural question is whether there exists a type space that encodes all "conceivable" hierarchical beliefs. This question has been answered in the affirmative ([19]) and B-S give the following definition of the complete-type space:

DEFINITION 2.4. ([13]) *A* belief-complete type space *on* $(\mathcal{H}, S(\cdot), \Theta, I)$ *is a type space* $\mathcal{T} = (\mathcal{H}, S(\cdot), \Theta, I, (\Omega_i, T_i, g_i)_{i\in I})$ *such that for every* $i \in I$, $\Omega_i = \Sigma_i \times T_i$ *and the function* g_i *maps* T_i *onto* $\Delta^{\mathcal{H}}(\Pi_{j\neq i}\Sigma_j \times T_j)$.

B-S show in [12] that it is always possible to construct a belief-complete type space.

Now the basic behavioral assumptions in game theory is that each player i chooses and carries out a strategy $s_i \in S_i$ that is optimal given her payoff type θ_i and her beliefs, conditional upon any history consistent with s_i ([13, p. 363]). Here we define the notion of *best reply*:

DEFINITION 2.5. ([13, p. 363]) *Fix a CPS* $\mu_i \in \Delta^{\mathcal{H}}(\Sigma_{-i})$. *A strategy* $s_i \in S_i$ *is a* sequential best reply *to* μ_i *for payoff type* $\theta_i \in \Theta_i$ *if and only if for every* $h \in \mathcal{H}(s_i)$ *and every* $s'_i \in S_i(h)$:

$$\sum_{(s_{-i},\theta_{-i})\in\Sigma_{-i}} (U_i(s_i, \theta_i, s_{-i}, \theta_{-i}) - U_i(s'_i, \theta_i, s_{-i}, \theta_{-i})) \times \mu(\{(s_{-i}, \theta_{-i})\}|S_{-i}(h) \times \Theta_{-i}) \geq 0$$

For any CPS $\mu \in \Delta^{\mathcal{H}}(\Sigma_{-i})$ let $r_i(\mu_i)$ denote the set of pairs $(s_i, \theta_i) \in \Sigma_i$ such that s_i is a sequential best reply to μ_i for θ_i.
We will also include the following notation to define rationality. Fix $\mathcal{T}$ a type space and for every player $i \in I$, define:

$$f_i = (f_{i,h})_{h\in\mathcal{H}} : T_i \to [\Delta(\Sigma_{-i})]^{\mathcal{H}}$$

as her first-order belief mapping, i.e. for all $t_i \in T_i$ and $h \in \mathcal{H}$:

$$f_{i,h}(t_i) = marg_{\Sigma_{-i}} g_{i,h}(t_i).$$

Finally, we define rationality of a player:

Definition 2.6. *Player i is* rational *at a state ω in $\mathcal{T}$ if and only if*

$$\omega \in R_i = \{(s, \theta, t) \in \Omega : (s_i, \theta_i) \in r_i(f_i(t_i))\}.$$

After all these concepts in place B-S move on to define their notion of (conditional) *probability-one belief*, or (conditional)*certainty* as they call it ([13, p.: 364]).

So let $\mathcal{A}_i$ denote the $\sigma - algebra$ of events $E \subseteq \Omega$ such that $E = \Omega_{-i} \times proj_{\Omega_i} E$. We can see $\mathcal{A}_i$ as the collection of events concerning Player i. The collection of events concerning Player i's opponents: $\mathcal{A}_{-i}$ is similarly defined.

Definition 2.7. *The* conditional (probability-one) belief operator *for player $i \in I$ given some history $h \in \mathcal{H}$ is a mapping $B_{i,h} : \mathcal{A}_{-i} \to \mathcal{A}_i$, defined by:*

$$\forall E \in \mathcal{A}_{-i} : B_{i,h}(E) = \{(s, \theta, t) \in \Omega : g_{i,h}(t_i)(proj_{\Omega_{-i}} E) = 1\}.$$

For any $E \in \mathcal{A}_{-i}$, $B_{i,h}(E)$ is read as: "Player i would be certain that her opponent's strategies, payoff and epistemic types are consistent with E were she to observe history h". Now B-S argue that their conditional belief operator satisfies the standard properties: *Conjunction, Monotonicity* and that for any $E \in \mathcal{A}_{-i}$, $B_{i,h}(E)$ is measurable ([13, p.: 362]).

And now for their notion of Strong Belief:

Definition 2.8. ([13, p.: 365]) *For any type space $\mathcal{T}$, define the operator $SB_i : \mathcal{A}_{-i} \to \mathcal{A}_i$ by $SB_i(\emptyset) = \emptyset$ and:*

$$SB_i(E) = \bigcap_{h \in \mathcal{H} : E \cap [h] \neq \emptyset} B_{i,h}(E)$$

for all events $E \in \mathcal{A}_{-i} - \{\emptyset\}$ and $[h] := \Pi_{j \in I} S_j(h) \times \Theta_j \times T_j$ is the event "history h occurs".

We say that player i *strongly believes* that an event $E \neq \emptyset$ is true if and only if he is certain of E at all histories consistent with E.
These are B-S's notions of conditional and strong belief. We will be referring to them at certain points throughout this thesis, in order to compare them with our work.

CHAPTER 3

Conditional Probability Spaces

In this chapter, we will define the structures called conditional probability spaces. These structures will be the core of this thesis and will be based on two-place probability functions. We will also show that classical probability spaces can be recovered from our conditional probability spaces.

Moreover, we will present Van Fraassen's notions of normal, abnormal and a priori sets ([25], [3], [21]) and prove that they have certain interesting properties.

We begin with the definition of a $\sigma - algebra$.

DEFINITION 3.1. *Let X be some set and 2^X its powerset.*
Then a subset $\Sigma \subseteq 2^X$ will be called a $\sigma - algebra$ on X if it satisfies the following:

- $X \in \Sigma$,
- $A \in \Sigma \Rightarrow X - A \in \Sigma$ *(Σ is closed under complements),*
- $A_1, A_2, A_3, ... \in \Sigma \Rightarrow A_1 \cup A_2 \cup A_3 \cup ... \in \Sigma$ *(Σ is closed under countable unions).*

It follows that the empty set: $\emptyset$ is in Σ and also that Σ is closed under countable intersections as well.

We proceed with the definition of a two-place probability function. As mentioned above the idea behind CPS's is to take conditional probability as primitive.
Let W be the set of all possible worlds, $\mathcal{F}$ a $\sigma - algebra$ on W.

DEFINITION 3.2. *The two-place function $P(\cdot|\cdot) : \mathcal{F} \times \mathcal{F} \to [0, 1]$ defined on $\mathcal{F} \times \mathcal{F}$ with the following requirements:*

- *for any fixed set $A \in F$, the map $P^A : \mathcal{F} \to [0, 1]$ such that*
$$P^A(B) = P(B|A), \forall B \in \mathcal{F},$$
is either a countably additive probability measure on $\mathcal{F}$, or $P(B|A)$ has constant value 1,
- *(Multiplication Axiom) for all $A, B, C \in \mathcal{F}$:*
$$P(B \cap C|A) = P(B|A)P(C|B \cap A),$$

will be called a two-place probability function *over $\mathcal{F} \times \mathcal{F}$.*

For some $A \in \mathcal{F}$, define: $P(A) = P(A|W)$, with W being the whole space of possible worlds.

And now we proceed with the definition of a conditional probability space:

DEFINITION 3.3. *We will call the structure $(W, \mathcal{F}, P)$ with*

- *W a set of possible worlds,*
- *$\mathcal{F}$ a $\sigma - algebra$ on W,*
- *P a two-place probability function over $\mathcal{F} \times \mathcal{F}$ as defined above,*

a conditional probability space.

Moreover, for some $A \in \mathcal{F}$, define: $A^c = W - A$, i.e. A^c is the complement of A.

OBSERVATION 3.4. *At this point, observe that classical probability spaces $(W, \mathcal{F}, P)$ with P a unary countably additive probability measure: $P : \mathcal{F} \to [0, 1]$, can be recovered as a special case of conditional probability spaces.*

PROOF. Consider $(W, \mathcal{F}, P)$ a space such that:

- W a set of possible worlds,
- $\mathcal{F}$ a $\sigma - algebra$ on W,
- $P(\cdot) : \mathcal{F} \to [0, 1]$ such that for $A \in \mathcal{F}$, $P(A)$ is a countably additive probability measure and
  - $P(B|A) = \frac{P(B \cap A)}{P(A)}$, if $P(A) > 0$,
  - $P(B|A) = 1$, if $P(A) = 0$.

Now we need to establish that the *Multiplication Axiom* $P(B \cap C|A) = P(B|A)P(C|B \cap A)$ holds for the probability function P.

So take $A, B, C \in \mathcal{F}$. We have the following cases:

- Case 1. $P(A) = 0$. Then since $B \cap A \subseteq A$ and $P(\cdot)$ is a probability measure, we get that $P(B \cap A) = 0$ as well. Therefore:
$$1 = P(B \cap C|A) = P(B|A) = P(C|B \cap A) = 1.$$
- Case 2. $P(A) > 0$. Then we have the following subcases:
  - Subcase 1. $P(B \cap A) = 0$. Then since $B \cap A \cap C \subseteq B \cap A$ and $P(\cdot)$ is a probability measure, we get that $P(B \cap A \cap C) = 0$ as well. Thus:
$$0 = \frac{P(B \cap C \cap A)}{P(A)} = P(B \cap C|A) = P(B|A) = \frac{P(B \cap A)}{P(A)} = 0.$$
  - Subcase 2. $P(A \cap B) > 0$. Then:
$$P(B \cap C|A) = \frac{P(B \cap C \cap A)}{P(A)} = \frac{P(B \cap C \cap A)}{P(A)} \frac{P(B \cap A)}{P(B \cap A)} = P(B|A)P(C|B \cap A).$$

□

We have now reached an essential part of Van Fraassen's theory, the normal and abnormal sets ([25]). Abnormal sets play the role of "absurdities". They are the sets (or events) that not only have measure 0 but are also considered so absurd by the agent that if they occur, our agent is so confused that he is willing to believe anything.

DEFINITION 3.5. *Normal and abnormal sets.*

- *Let $A \in \mathcal{F}$. If $P^A(B)$ is a probability measure, for $B \in \mathcal{F}$, we will call A* normal.
- *Let $A \in \mathcal{F}$. If $P^A(B)$ has constant value 1 for $B \in \mathcal{F}$, we will call A* abnormal.

It is important to notice that a normal set may have measure 0. This means that we do not expect the agent to believe it, but he still might. For example, the rationals as a subset of the reals have measure 0. However, it might happen that we randomly pick a real and it turns out to be a rational. Therefore, the rationals $\mathbb{Q}$ are a normal set of measure 0 ([21, p. 6]).

DEFINITION 3.6. *A conditional probability space* $(W, \mathcal{F}, P)$ *in which* W *is abnormal, will be called* trivial.

Notice that if W is abnormal, then all $P(A|B) = 1$ for all $A, B \in \mathcal{F}$.

OBSERVATION 3.7. *For* $A, B \in \mathcal{F}$*:* $P(A \cap B|A) = P(B|A)$.

PROOF. Pick $A, B \in \mathcal{F}$. We have two cases:

- Case 1. A is normal. Then the *Multiplication Axiom* tells us that: $P(A \cap B|A) = P(A|A)P(B|A \cap A)$.
  Now $P(A|A)$ is a probability measure and therefore: $P(A|A) = 1$.
  Hence $P(A \cap B|A) = P(B|A)$.
- Case 2. A is abnormal. Then $P(A \cap B|A) = 1 = P(B|A)$.

□

We now proceed with the definition of a priori sets. This notion is the exact opposite of abnormality. As Van Fraassen puts it, an a priori proposition A is not epistemically distinguishable from the tautology W (the whole space). Intuitively, this means that the negation of A is considered absurd, or even better *abnormal*.

DEFINITION 3.8. *A priori and contingent sets.*

- *We will call a set* $K \in \mathcal{F}$ a priori *if and only if* K^c *is abnormal.*
- *We will call a set* $K \in \mathcal{F}$ contingent *if and only if* K *is not a priori.*

Note that Van Fraassen has not specified non-a priori (our contingent) sets. Though it may be a bit too much to have normal, abnormal, a priori and contingent sets all together, we decided that such a distinction will be useful — especially in the next chapters. Now our normal and abnormal sets have some interesting properties that will be of use later on.

OBSERVATION 3.9. *In a classical probability space* $(W, \mathcal{F}, P)$ (as in 3.4)*, the sets with probability* 0 *are abnormal and the sets with probability* 1 *are a priori.*

PROOF. Take $(W, \mathcal{F}, P)$ a classical conditional probability space as in 3.4.

Consider $A \in \mathcal{F}$ such that $P(A) = 0$. Then $P(\cdot|A)$ has constant value 1. This entails by Definition 3.5 that A is abnormal.

Now consider $A \in \mathcal{F}$ such that $P(A) = 1$. Then since $P(A)$ is a probability measure, we have that $P(A^c) = 0$. Hence A^c is abnormal. This entails that A is a priori (Definition 3.8). □

And now we proceed with certain properties of normal and abnormal sets.

PROPERTY 3.10. *If* A *is abnormal and* B *is normal, then:* $P(A|B) = 0$

PROOF. Take $A, B \in \mathcal{F}$ such that A is abnormal and B is normal.
Then $P(\emptyset|B) = P(\emptyset|A)P(A|B)$, but B is normal thus $P(\emptyset|B) = 0$ and A is abnormal, thus $P(\emptyset|A) = 1$, hence $P(A|B) = 0$. □

PROPERTY 3.11. *Supersets of normal sets are normal.*

PROOF. Pick $A \in \mathcal{F}$ a normal set. Assume towards a contradiction that $\exists B \in \mathcal{F}$ such that $A \subseteq B$ and B is abnormal.
Then we have that: $0 = P(\emptyset|A) = P(B|A) = P(B \cap A|A) = P(A|A) = 1$.
Contradiction.

□

PROPERTY 3.12. *Subsets of abnormal sets are abnormal.*

PROOF. Pick $A \in \mathcal{F}$ an abnormal set. Consider $B \in \mathcal{F}$ such that $B \subseteq A$. Then if B is normal, the previous property entails that A is normal as well. Hence, B can only be abnormal.

□

This property is also proved in [21, p.: 7].

PROPERTY 3.13. *A countable union of abnormal sets is abnormal.*

PROOF. Consider $X, Y \in \mathcal{F}$ two abnormal sets. Assume towards a contradiction that $X \cup Y$ is normal.
Then the multiplication axiom entails that: $P(\emptyset|X \cup Y) = P(\emptyset|X)P(X|X \cup Y)$.
We know that $P(\emptyset|X) = 1$, since X is abnormal.
Moreover, we have assumed that $X \cup Y$ is normal and hence $P^{X \cup Y}(\cdot)$ is a probability measure.
Hence, $P(\emptyset|X \cup Y) = 0$. This entails that $P(X|X \cup Y) = 0$.
With an analogous argument we also get that $P(Y|X \cup Y) = 0$.
However $1 = P(X \cup Y|X \cup Y) = P(X|X \cup Y) + P(Y|X \cup Y) - P(X \cap Y|X \cup Y)$ and hence combining these we get a contradiction.

Assume now that $X_i : i \in \mathbb{N}$ are countably many abnormal sets in $\mathcal{F}$.
Consider $Z = \bigcup\limits_i X_i$.
We have that $Z \in \mathcal{F}$ by the definition of a $\sigma - algebra$.
Now assume towards a contradiction that Z is normal.
We get that $P(X_i|Z) = P(\emptyset|Z) = 0$.
But $P^Z(\cdot)$ is a probability measure, thus $P(Z|Z) = 0$.
Contradiction. □

OBSERVATION 3.14. *An intersection of normal sets (either finite intersection or not) is not necessarily normal as well.*

PROOF. Counterexample:
Consider $W = \{w_1, w_2, w_3, w_4\}$ and consider a two-place function P as defined above such that: $P(w_i|W) = \frac{1}{4}$, for all $i \in \{1, 2, 3, 4\}$.
Now take $X = \{w_1, w_2\}$ and $Y = \{w_3, w_4\}$.
Then X and Y are normal sets but $X \cap Y = \emptyset$ and hence $X \cap Y$ is abnormal. $\square$

Observe that Battigalli and Siniscalchi's Conditional Probability Systems as defined in 2.2[1] are similar to our two-place probability functions as defined in 3.2 along with our normal and abnormal sets in Definition 3.5.

First of all observe that B-S's collection $\mathcal{B}_i$ is essentially our collection of normal sets. For them, $\mathcal{B}_i$ is a collection of observable events — or "relevant hypotheses" — concerning the real world x. This implies that the agent considers the elements of $\mathcal{X}_i - \mathcal{B}_i$ as non-relevant hypotheses, which in our setting amounts to them being abnormal.

Moreover, B-S define their CPS μ on $\mathcal{X}_i \times \mathcal{B}_i$, demanding that $\mu(\cdot|B)$ is a probability measure on $(X_i, \mathcal{X}_i)$. Therefore, if $B \in \mathcal{B}_i$ then $\mu(\cdot|B)$ is a probability measure. Otherwise (i.e. if $B \notin \mathcal{B}_i$), $\mu(\cdot|B)$ is not defined.

Now in their setting, our two-place probability function P is defined on $\mathcal{X}_i \times \mathcal{X}_i$, demanding that $P(\cdot|B)$ is a probability measure if B is normal and otherwise $P(\cdot|B) = 1$. Hence, if $B \in \mathcal{B}_i$, we have that $P(\cdot|B)$ is defined but has constant value 1.

Finally, B-S replace our *Multiplication Axiom* with

$$A \subseteq B \subseteq C \text{ implies } \mu(A|B)\mu(B|C) = \mu(A|C)$$

for $A \in \mathcal{X}_i$ (i.e. A is possibly abnormal — or $\mu(\cdot|A)$ is not defined) and $B, C \in \mathcal{B}_i$.

Now if we in turn assume our *Multiplication Axiom* for a two-place probability function $P : \mathcal{X}_i \times \mathcal{X}_i \to [0, 1]$:

$$\forall A, B, C \in \mathcal{X}_i : P(B \cap A|C) = P(B|C)P(A|B \cap C),$$

and take $A, B, C \in \mathcal{X}_i$ such that B, C are normal and $A \subseteq B \subseteq C$, we derive that

$$P(A|B)P(B|C) = P(A|C).$$

[1]For Battigalli and Siniscalchi, a conditional probability system (or CPS) on $(X_i, \mathcal{X}_i, \mathcal{B}_i)$, is a mapping $\mu(\cdot|\cdot) : \mathcal{X}_i \times \mathcal{B}_i \to [0, 1]$ such that, for all $B, C \in \mathcal{B}_i$ and $A \in \mathcal{X}_i$, we have that:

- $\mu(B|B) = 1$,
- $\mu(\cdot|B)$ is a probability measure on $(X_i, \mathcal{X}_i)$,
- $A \subseteq B \subseteq C$ implies $\mu(A|B)\mu(B|C) = \mu(A|C)$.

CHAPTER 4

r-stability

In this chapter we will define stable sets. We will also prove that they have some handy properties that will be essential in the following chapters.

Our definition is similar to Leitgeb's ([34], [35]). The idea behind the notion of an r-stable proposition, is that it maintains a probability high enough (bigger than $\frac{1}{2}$) whatever *consistent with it* happens. Similarly to Leitgeb, r represents our agent's threshold.

Throughout the whole chapter we will be talking about r-stability with the number $r \in (\frac{1}{2}, 1]$ specified in the beginning. This is because we assume that an agent appears in a game with his own threshold predetermined. This means that we assume that each agent has already specified a number bigger than one half and smaller or equal to 1 that serves as his own degree of gullibility/mistrust. Our assumption is that before the game begins each agent has already picked his threshold in $(\frac{1}{2}, 1]$.

As written in the introduction Leitgeb proves that r-stable sets are nested and well-founded w.r.t. the $\subseteq$ relation. Sadly, we can not do the same unless we change the definition of r-stable sets, because in our case things are more complicated than in a classical probabilistic setting. Allowing conditioning on events with measure 0 makes it impossible to define stable beliefs similarly to Leitgeb *and* at the same time maintain nestedness w.r.t. the $\subseteq$ relation. Assume for example that X, Y are two a priori r-stable sets. Then, their relative complement $((X - Y) \cup (Y - X))$ might be abnormal, but nevertheless *non-empty*. Then, it is not the case that $X \subseteq Y$ or $Y \subseteq X$.

There are two ways out of this situation: either change the definition of r-stable sets, demanding that all r-stable sets are contingent, or prove *some version* of nestedness for r-stable sets. We decided in favor of the second way out and defined a *quasi*-subset relation: $\subseteq_q$ such that $X \subseteq_q Y$ iff X is included in Y, apart from an abnormal part of it. Finally we established nestedness of r-stable sets w.r.t. $\subseteq_q$, i.e. for X, Y r-stable sets: $X \subseteq_q Y$ or $Y \subseteq_q X$.

Another issue was proving well-foundedness of r-stable beliefs w.r.t. the $\subseteq$ relation. The difference in our framework is that it might be the case that an r-stable set has an abnormal subset. This would imply (as we will prove later on) that this r-stable set is a priori. But in this case, it is not necessary that well-foundedness holds, since it might as well be the case that we have an infinitely descending sequence of a priori sets. Once again we could either change the definition of r-stable sets or prove a *version* of well-foundedness. We could define r-stable sets by demanding that they are contingent (but we rejected that already), or by demanding that all non-empty subsets of an r-stable set are normal. [2] Aiming to have a definition of r-stable sets as general as possible, in order

[2]This property is called Finesse by Arló-Costa ([3]) and it holds for Van Fraassen's belief cores. The proof can be found in [21, p.: 7].

for them to be straight-forwardly comparable with Leitgeb's stable sets and B-S's strong beliefs, we decided to prove a version of well-foundedness, i.e. only for *contingent* r-stable sets[3].

Finally, we proved two important propositions that will be used extensively throughout the rest of the thesis. These propositions tell us that an r-stable set with a non-empty abnormal subset is a priori and that if S is an a priori set and $S \cap E$ is non-empty and abnormal then E is abnormal.

So we begin by considering W a set of possible worlds and $(W, \mathcal{F}, P)$ a conditional probability space as defined above.

Pick now $r \in (\frac{1}{2}, 1]$ our agent's threshold.
We begin with the definition of r-stable sets.

Definition 4.1. *Define the following:*

- *We will call a set $S \in \mathcal{F}$* r-stable set*, if:*
  for all $B \in \mathcal{F}$ such that $B \cap \mathcal{F} \neq \emptyset : P(S|B) \geq r$.[4]
- *We will call a set $S \in \mathcal{F}$* r-superstable set*, if:*
  for all $B \in \mathcal{F} : P(S|B) \geq r$.
- *We will call a set $S \in \mathcal{F}$* certain*, if:*
  $P(S|W) = 1$.
- *We will call a set $S \in \mathcal{F}$* absolutely certain *if:*
  for all $B \in \mathcal{F} : P(S|B) = 1$.

Observation 4.2. *Consider $S \in \mathcal{F}$:*
The following are equivalent:

1. *S is* r-superstable
2. *S is absolutely certain*
3. *S is* 1-superstable
4. *S is a priori*

Proof. We will prove the following equivalences:

- 1. $\Leftrightarrow$ 4.
  ($1 \Rightarrow 4$) Assume that $X \in \mathcal{F}$ is r-superstable and that X^c is normal.
  Then $P(X|X^c) \geq r \Leftrightarrow P(W|X^c) - P(X^c|X^c) \geq r \Leftrightarrow 0 \geq r$.
  Contradiction and thus X^c is abnormal and X is a priori.
  ($4 \Rightarrow 1$) A direct consequence of the definition of a priori sets.
- 2. $\Leftrightarrow$ 4.
  ($2 \Rightarrow 4$) Suppose $X \in \mathcal{F}$ is absolutely certain and assume that X^c is normal.
  Then $P(X|X^c) = 1 \Leftrightarrow P(W|X^c) - P(X^c|X^c) = 1 \Leftrightarrow 0 = 1$.
  Contradiction, thus X^c is abnormal and therefore X is a priori.
  ($4 \Rightarrow 2$) A direct consequence of the definition of a priori sets.

[3]Notice that Leitgeb does something similar as well. In [34, p.: 1364] he proves that there is no infinitely descending sequence of P^r-stable sets with probability less than 1.

[4]Notice that for Leitgeb $r \in [\frac{1}{2}, 1)$ and $P(S|B) > r$, while for us $r \in (\frac{1}{2}, 1]$ and $P(S|B) \geq r$.

- 3. $\Leftrightarrow$ 2.
  Obvious.

$\square$

From now on, we fix a number $r \in (\frac{1}{2}, 1]$. Unless otherwise specified, stable means r-stable.

Now we will define the *quasi*-subset relation, in order to establish a version of nestedness for our stable sets.

DEFINITION 4.3. *For some $X, Y \in \mathcal{F}$, define: $X \subseteq_q Y$ iff $\exists$ an abnormal set Z such that: $X \subseteq Y \cup Z$.*

PROPOSITION 4.4. *If for $X, Y \in \mathcal{F}$ we have that $X \subseteq Y$ then $X \subseteq_q Y$.*

PROOF. Take $X, Y \in \mathcal{F}$ such that: $X \subseteq Y$.
We have that $X \subseteq Y \cup (X - Y)$.
However, $X - Y = \emptyset$ and therefore $X - Y$ is abnormal.
Thus $X \subseteq_q Y$. $\square$

And now we prove that our stable sets are nested w.r.t. $\subseteq_q$.

PROPERTY 4.5. *If X, Y are stable sets then either $X \subseteq_q Y$ or $Y \subseteq_q X$.*

PROOF. Assume that $X, Y \in \mathcal{F}$ are stable sets.

- Case 1. $X \subseteq Y$ or $Y \subseteq X$. Then we get by Proposition 4.4 above that $X \subseteq_q Y$ or $Y \subseteq_q X$ and our property holds.
- Case 2. Neither $X \subseteq Y$ nor $Y \subseteq X$.
  Consider the set: $Z = (X - Y) \cup (Y - X)$. $\mathcal{F}$ is a $\sigma - algebra$ hence: $Z \in \mathcal{F}$.

  Claim: Z is abnormal.

  PROOF. Assume towards a contradiction that Z is normal.
  We have that $Z \cap X = X - Y \neq \emptyset$ and $Z \cap Y = Y - X \neq \emptyset$ since neither $X \subseteq Y$ nor $Y \subseteq X$.
  The definition of r-stability entails that $P(X|Z) \geq r$ and $P(Y|Z) \geq r$.
  Hence we get: $P(X|Z) + P(Y|Z) \geq 2r > 1$.
  But this is equivalent to (by Observation 3.7):

$$P(X \cap Z|Z) + P(Y \cap Z|Z) \geq 2r.$$

  Hence: $P(X - Y|Z) + P(Y - X|Z) \geq 2r > 1$.

But Z is normal and therefore $P^Z(\cdot) = P(\cdot|Z)$ is a probability measure.
Therefore we have that:

$$P(((X - Y) \cup (Y - X))|Z) - P(((X - Y) \cap (Y - X))|Z) \geq 2r.$$

Hence: $P(Z|Z) \geq 2r > 1$.
Contradiction. $\square$

Hence Z is abnormal. Now since we have that $X \subseteq Y \cup Z$, we get that $X \subseteq_q Y$.

$\square$

Corollary 4.6. *If X and Y are stable and at least one of X, Y is contingent, then either $X \subseteq Y$ or $Y \subseteq X$.*

Now the previous proposition entails the following interesting observation:

Observation 4.7. *If X, Y are stable sets, so is $X \cap Y$.*

Proof. Consider $X, Y \in \mathcal{F}$ stable sets.
Assume without loss of generality that $Y \subseteq_q X$.
Then there exists an abnormal set Z such that: $Y \subseteq X \cup Z$.
However, we also have that $Y = X \cup (Y - X)$ and therefore $Y - X$ is abnormal.
Now pick set $E \in \mathcal{F}$ such that $X \cap Y \cap E \neq \emptyset$.
Then:

$$P(X \cap Y|E) = P(X|E) + P(Y|E) - P(X \cup Y|E).$$

Now we have that: $X \cup Y = X \cup (Y - X)$.
Therefore:

$$P(X \cap Y|E) = P(X|E) + P(Y|E) - P(X|E) - P(Y - X|E) + P(X \cap (Y - X)|E).$$

But $Y - X$ is abnormal and therefore:

$$P(X \cap Y|E) = P(Y|E) \geq r.$$

Hence $X \cap Y$ is stable. $\square$

We also have the following property for unions of stable sets.

Property 4.8. *A countable union of stable sets is stable.*

Proof. Let $\{K_i\}$ a countable family of stable sets $K_i \in \mathcal{F}$.
Then $B = \bigcup\limits_i K_i$ is in $\mathcal{F}$ as well.
Now let $E \in \mathcal{F}$ a normal set such that $B \cap E \neq \emptyset$.
We want to show that: $P(B|E) \geq r$.
First notice that if E was abnormal, then $P(B|E) \geq r$ would hold trivially.

Since now $B \cap E = (\bigcup_i K_i) \cap E = \bigcup_i (K_i \cap E)$ is non-empty, there must be some i such that $K_i \cap E \neq \emptyset$.
Since K_i is stable, we get that: $P(K_i|E) \geq r$.
But then, we also have that $P(B|E) \geq r$ since $K_i \subseteq B$ and E is a normal set.
Hence B is indeed a stable set. □

As mentioned above, we will show that contingent stable sets are well-founded. This is the strongest result we can get without imposing any constraints on the definition of stable sets.

PROPOSITION 4.9. *There is no infinitely descending chain of sets $K_1 \supset K_2 \supset K_3 \supset ...$ such that all K_i are contingent stable. (Note that the relation $\supset$ here is proper.)*

PROOF. Assume towards a contradiction that there exists an infinitely descending chain of sets $K_1 \supset K_2 \supset K_3 \supset ...$ such that all K_i are contingent stable sets.
Now consider the set $X = (K_1 - \bigcap_i K_i)$.

Claim: X is normal.

PROOF. Assume otherwise towards a contradiction. Then consider the set:

$$Y = (W - K_1) \cup X = K_1^c \cup X.$$

Now $Y \in \mathcal{F}$ and $Y \cap K_1 \neq \emptyset$. Now since K_1 is stable, we get that:

$$P(K_1|Y) \geq r > \frac{1}{2}.$$

Now by Observation 3.7: $P(K_1|Y) = P(K_1 \cap Y|Y) = P(X|Y) \geq r$.
Now K_1^c is normal (since K_1 is contingent) and therefore by our properties above Y is normal as well, as its superset.
Now we have assumed that X is abnormal and hence: $P(X|Y) = 0$. But this is a contradiction. □

Now that we have established that X is normal, we go on with our proof:
We have that:

$$\begin{aligned}
1 &= P(X|X) \\
&= \sum_{i=1}^{\infty} P(K_i - K_{i+1}|X) \\
&= \lim_{N\to\infty} \sum_{i=1}^{N-1} P(K_i - K_{i+1}|X) \\
&= \lim_{N\to\infty} P(K_1 - K_N|X) \\
&= P(K_1|X) - \lim_{N\to\infty} P(K_N|X) \\
&= 1 - \lim_{N\to\infty} P(K_N|X)
\end{aligned}$$

Now every K_N is stable and $K_N \cap X \neq \emptyset$.
This is because $K_N - K_{N+1} \subseteq K_N \cap X$, which is non-empty because K_{N+1} is strictly included in K_N.
Therefore, we have that $P(K_N|X) \geq r, \forall N$.
Hence: $\lim_{N\to\infty} P(K_N|X) \geq r$.
Thus, we get that $1 \leq 1 - r$ from the equalities above.
Contradiction. □

COROLLARY 4.10. *The family of contingent stable sets is closed under arbitrary intersections.*

PROOF. Let $\mathcal{G}$ be a non-empty family of contingent stable sets.
Then by Corollary 4.6, we know that sets in $\mathcal{G}$ are totally ordered by $\supseteq$.
By Proposition 4.9 there has to exist a smallest set in $\mathcal{G}$ (included in all others), otherwise we would obtain an infinite descending chain of sets in $\mathcal{G}$.
Then obviously the intersection of all the sets in $\mathcal{G}$ coincides with this smallest set of $\mathcal{G}$. □

Now we have two important propositions.

PROPOSITION 4.11. *If $S \in \mathcal{F}$ is stable and there exists an abnormal set $a \in \mathcal{F}$ with $a \neq \emptyset$, such that $a \subseteq S$, then S is a priori.*

PROOF. Take S an $r-$*stable* set and assume that there exists an abnormal set $a \in \mathcal{F}$ with $a \neq \emptyset$ and $a \subseteq S$.
Assume towards a contradiction that S is not a priori.
Then $W - S$ is normal.
Now take $E = a \cup (W - S)$. We have that $S \cap E = S \cap (a \cup (W - S)) = a \neq \emptyset$ and since S is r-stable we have that $P(S|E) \geq r$.
By Observation 3.7 this is equivalent to: $P(a|E) \geq r$.
Contradiction. □

COROLLARY 4.12. *Any non-empty subset of a contingent stable set is normal.*

PROPOSITION 4.13. *For $S \in \mathcal{F}$ an a priori set, if for some $A \in \mathcal{F}$ we have that $S \cap A$ is abnormal, then A is abnormal.*

PROOF. Take $S \in \mathcal{F}$ a priori set and assume that for some $A \in \mathcal{F}$, $S \cap A$ is abnormal.
We have that S^c is abnormal and therefore $A - S$ is abnormal as well as its subset.
Moreover, we have that $S \cap A$ is abnormal as well.
But $A = (S \cap A) \cup (A - S)$ and a countable union of abnormal sets is abnormal, by Property 3.13.
Hence A is abnormal. □

Propositions 4.11, 4.13 will give us a standard way to argue that a non-empty intersection of an r-stable set S with a normal set E is normal. This is because if we assume otherwise, i.e. that $\emptyset \neq S \cap E \subseteq S$ is abnormal with S r-stable and E normal, then by 4.11 S is a priori and then by 4.13 E is abnormal. Contradiction. This argument will be used extensively in what comes next.

The following Proposition is a generalization (to conditional probability spaces) and strengthening of a result by Leitgeb on characterizing stable sets in a quantitative manner (in the framework of classical probability). For its statement and its proof, we adopt the following *conventions*:
$\frac{1}{0} = \infty$, $\infty \cdot 0 = 1$, $\infty \cdot x = \infty$ for $x > 0$.
These conventions allow us to deal with the case $r = 1$, in such a way that we can state and prove the result below in a uniform manner, for all thresholds $r \in (\frac{1}{2}, 1]$.

Proposition 4.14. *Let $X \in \mathcal{F}$ be a contingent set, such that every non-empty subset of X is normal. Then the following are equivalent:*

(1) *X is stable.*
(2) *For all sets $Y, Z, T \in \mathcal{F}$ such that $\emptyset \neq Y \subseteq X$, $Z \subseteq (W - X)$, $Y \cup Z \subseteq T$, we have*
$$P(Y|T) \geq \frac{r}{1-r} P(Z|T).$$
(3) *For all sets $Y, Z \in \mathcal{F}$ such that $\emptyset \neq Y \subseteq X$, $Z \subseteq (W - X)$, we have*
$$P(Y|Y \cup Z) \geq \frac{r}{1-r} P(Z|Y \cup Z).$$

Proof. (1) $\Rightarrow$ (2). Let X be a contingent stable set, and $Y, Z, T \in \mathcal{F}$ be sets s.t. $\emptyset \neq Y \subseteq X$, $Z \subseteq (W - X)$, $Y \cup Z \subseteq T$.
Since X is stable and $X \cap (Y \cup Z) = Y \neq \emptyset$, we must have that
$$P(Y|Y \cup Z) = P(X \cap (Y \cup Z)|Y \cup Z) = P(X|Y \cup Z) \geq r.$$
Since Y is a non-empty subset of X it must be normal (by the assumption of this Proposition). So $Y \cup Z$ is also normal, and hence (given that Y and Z are disjoint, since $Y \subseteq X$ and $Z \subseteq (W - X)$) we have that
$$P(Z|Y \cup Z) = 1 - P(Y|Y \cup Z) \leq 1 - r.$$
By applying the Multiplication Axiom, we obtain that:
$$P(Y|T) = P(Y|Y \cup Z) \cdot P(Y \cup Z|T) \geq r \cdot P(Y \cup Z|T) = \frac{r}{1-r} \cdot (1-r) \cdot P(Y \cup Z|T) \geq$$
$$\geq \frac{r}{1-r} \cdot P(Z|Y \cup Z) \cdot P(Y \cup Z|T) = \frac{r}{1-r} \cdot P(Z|T).$$
(Note that, with the above conventions, this derivation works even for $r = 1$.)

(2) $\Rightarrow$ (3). Obvious (just take $T := Y \cup Z$).

(3) $\Rightarrow$ (1). Suppose towards a contradiction that (3) holds for X, but that X is *not* stable.
Then there exists some $E \in \mathcal{F}$ such that $E \cap X \neq \emptyset$ but $P(X|E) < r$.
Take $Y = E \cap X$, $Z = E \cap (W - X)$.

Since Y is a non-empty subset of X, it follows that Y is normal (by the assumption of this Proposition), and hence E is normal (since $Y \subseteq E$). Clearly, we have

$$P(Y|E) = P(E \cap X|E) = P(X|E) < r,$$

and from the normality of E (and since Y and Z are disjoint) we obtain that

$$P(Z|E) = 1 - P(Y|E) > 1 - r.$$

Note that the conditions of (3) are satisfied by this choice of Y and Z and that moreover $Y \cup Z = E$. By applying (3), we get that

$$P(Y|E) \geq \frac{r}{1-r} \cdot P(Z|E) \geq \frac{r}{1-r} \cdot (1-r) = r$$

(where again the derivation goes through even for $r = 1$, given our conventions).
From this, we get:

$$P(X|E) = P(E \cap X|E) = P(Y|E) \geq r,$$

which contradicts the choice of E (which was s.t. $P(X|E) < r$). □

COROLLARY 4.15 (Leitgeb's characterization). *Suppose we have a* classical *probability space* $(W, \mathcal{F}, P)$ *(i.e. in which all sets of measure* 0 *are abnormal), and let* $X \in \mathcal{F}$ *with* $P(X) < 1$. *Let* $r \in (\frac{1}{2}, 1)$. *Then the following are equivalent:*

(1′): *X is r-stable.*

(2′): *For all sets $Y, Z, T \in \mathcal{F}$ such that $\emptyset \neq Y \subseteq X$, $Z \subseteq (W - X)$, we have*

$$P(Y) \geq \frac{r}{1-r} \cdot P(Z).$$

PROOF. (1′) ⇒ (2′). Assume that X is r-stable.
Note that X is contingent, since it has probability different from 1, so its complement is normal (since it has non-zero probability and the space is classical).
By Corollary 4.12 from before, we know that every non-empty subset of a contingent stable set is normal.
Hence we can apply the implication (1) ⇒ (2) of the above Proposition (Proposition 4.14), with $T := W$.

(2′) ⇒ (1′). Assume (2′). We will prove first the assumption of the above Proposition 4.14: let $Y \in \mathcal{F}$ s.t. $\emptyset \neq Y \subseteq X$.
Take $Z := W - X$. Then $P(Z) = 1 - P(X) > 0$. By (2′), we have $P(Y) \geq \frac{r}{1-r} \cdot P(Z) > 0$ (since $r \neq 1$ and $P(Z) > 0$).
Hence Y is normal (since we are in a classical probability space), and we have proved the assumption of the above Proposition: every non-empty subset of X is normal.

We now prove condition (3) of the above Proposition 4.14: assume given $Y, Z \in \mathcal{F}$ such that $\emptyset \neq Y \subseteq X$, $Z \subseteq (W - X)$.
Using (2′) and the fact that $P(Y \cup Z) \neq 0$ (since $P(Z) > 0$), we obtain:

$$P(Y|Y \cup Z) = \frac{P(Y)}{P(Y \cup Z)} \geq \frac{r}{1-r} \cdot \frac{P(Z)}{P(Y \cup Z)} = \frac{r}{1-r} \cdot P(Z|Y \cup Z),$$

i.e. condition (3) of the above Proposition.
Applying now the implication (3) ⇒ (1) of the above Proposition, we obtain that X is r-stable. □

Example 2. Consider Example 1 again: $W = \{w_1, ..., w_8\}$ and P a classical probability measure on $\mathcal{P}(W)$ such that:
$P(\{w_1\}) = 0,54$, $P(\{w_2\}) = 0,342$, $P(\{w_3\} = 0,058$, $P(\{w_4\}) = 0,03994$, $P(\{w_5\}) = 0,018$, $P(\{w_6\}) = 0,002$, $P(\{w_7\}) = 0,00006$, $P(\{w_8\}) = 0$.
Now take $r = 3/4$ as our threshold and take the set $X = \{w_1, w_2, ..., w_5\}$.
Notice that then $X^c = \{w_6, w_7, w_8\}$.
Now we want to check whether X is $3/4 - stable$.
According to the previous Corollary we need to verify that

$$P(Y|W) \geq 3^5 P(Z|W)$$

for $Y \subseteq X$ with $P(Y) > 0$ and $Z \subseteq X^c$ with $P(Z) > 0$.
Take $Z = \{w_6, w_7, w_8\}$.
We have that $3P(Z) = 0,00618$.
And $P(\{w_1\}) = 0,54 > 0,00618$, $P(\{w_2\}) = 0,342 > 0,00618$, $P(\{w_3\}) = 0,058 > 0,00618$, $P(\{w_4\}) = 0,03994 > 0,00618$, $P(\{w_5\}) = 0,018 > 0,00618$.
Therefore, according to the previous Corollary X is indeed $3/4 - stable$.

Finally, we have the following definition:

Definition 4.16. *An* ω-stable conditional probability space $(W^\omega, \mathcal{F}^\omega, P^\omega)$ *is a conditional probability space that contains only countably many contingent stable sets.*

These spaces will be used in chapter 8.

[5] $3 = \frac{\frac{3}{4}}{1-\frac{3}{4}}$

CHAPTER 5

Conditional Belief

In this chapter we will define the notion of conditional belief on the grounds of r-stability, similarly to Leitgeb ([34, p.: 1381]). Leitgeb's theory of belief is based on what he calls the "Humean conception of belief". As mentioned above, for Leitgeb a proposition H is believed given E if there is some r-stable S set such that $S \cap E \neq \emptyset$ and $S \cap E \subseteq H$.

The intuition is that an agent would rationally believe a hypothesis H given some evidence E if one of his r-stable beliefs is consistent with the evidence E and furthermore H is "implied" by this r-stable belief and the evidence set. In other words, we demand that the evidence is first consistent with our agent's current beliefs and second that it is rational for the agent to conclude that the hypothesis H holds after observing (conditioning on) E. However, we should also consider something else that did non exist in Leitgeb's setting, the abnormal sets.

As mentioned in the introduction the abnormal sets are essentially our agent's "absurdities". When conditioning on an abnormal set a, everything gets probability 1. Something analogous will hold for our conditional belief as well, namely: when "conditioning" on an abnormal set, everything is believed. This captures the idea that if something the agent considered absurd occurs, then the agent is completely confused, not knowing what to believe and thus believing everything!

Notice that we will use the notion of conditional belief to derive full / absolute belief, by conditioning on the whole space of possible worlds.

We are also going to prove certain essential properties of our conditional beliefs, including closure under finite intersections, monotonicity and consistency w.r.t. normal sets. Notice that consistency of conditional beliefs w.r.t. all sets does not hold, because as written above everything is believed (event the empty set) when conditioning on an abnormal set.

Finally, we connect conditional belief directly with probability: H is believed given E implies $P(H|E) \geq r$, getting one direction of the Lockean thesis, exactly as Leitgeb does for his notion of belief.

We begin by considering $(W, \mathcal{F}, P)$ a conditional probability space with W a set of possible worlds.

Once again, consider a number $r \in (\frac{1}{2}, 1]$ to be our agent's threshold.

DEFINITION 5.1. *For $E, H \in \mathcal{F}$:*
$B^E H$ iff $B(H|E)$ iff either E is abnormal or $\exists S$: $r-$stable such that $S \cap E$ is normal and $S \cap E \subseteq H$

$B^E H$ should be read as: "H is believed *given* E". We introduced two notations for conditional belief: $B^E H$ and $B(H|E)$ and we will be switching to one another depending on the context (if we have a probability measure $P^E(\cdot)$ we will be using $B(H|E)$ and if we have a conditional probability $P(H|E)$ we will be using $B^E H$).

The definition above says that a hypothesis H is believed given some evidence E if and only if either E is abnormal or there exists an r-stable set S of which intersection with E is normal and is a subset of H. This definition captures the idea that a belief should be *justifiable* by some other *persistent* (stable) belief together with the evidence.

The first question is why (in the case that E is normal) do we demand that the intersection $S \cap E$ is normal and not just non-empty? This is the first time that we will use Propositions 4.11 and 4.13. Assume that $S \cap E$ is non empty. Then if it is abnormal, proposition 4.11 entails that S is a priori and proposition 4.13 entails that E is abnormal. Contradiction. On the other hand, if $S \cap E$ is normal, then it can not be the empty set (the empty set is by definition abnormal). Therefore, it appears that as long as E is normal $S \cap E$ being normal and $S \cap E$ being non empty are equivalent.

We also have the following corollary that expresses what we discussed about abnormal sets in the introduction of this chapter.

COROLLARY 5.2. *For $H, E \in \mathcal{F}$, if E is abnormal then $B(H|E)$ holds.*

PROOF. Follows directly from the definition of conditional beliefs. □

DEFINITION 5.3. *We will say that a hypothesis H is believed by our agent and write $B(H)$ iff $B(H|W)$ holds, with W the whole space of possible worlds.*

PROPERTY 5.4. *For $H, E \in \mathcal{F}$, if E is normal and $B(H|E)$ holds, then H is normal.*

PROOF. Assume otherwise, i.e. for $E, H \in \mathcal{F}$, $B(H|E)$ holds but H is abnormal.
Then $\exists S$: r-stable set such that: $\emptyset \neq S \cap E \subseteq H$.
Then by the properties of normal and abnormal sets we get that $S \cap E$ is abnormal as well (as a subset of an abnormal set), by Property 3.12.
Now $P(S|E) \geq r$ since S is r-stable and $S \cap E \neq \emptyset$.
However we also have that: $P(S|E) = P(S \cap E|E) = 0$, by Observation 3.7 and the fact that E is normal.
Contradiction.
Another way of proving this is by using propositions 4.11 and 4.13. These would give us that since $S \cap E$ is abnormal and non empty then S is a priori and then E is abnormal.
Contradiction. □

This property tells us that abnormal sets can not be believed when conditioning on normal sets. Or in other words, abnormal sets can be believed by the agent only when conditioning on abnormal sets.

PROPERTY 5.5. *For $H, E \in \mathcal{F}$, if H is a priori then $B(H|E)$.*

PROOF. Pick $H, E \in \mathcal{F}$ such that H is a priori. Since H is a priori we have that H^c is abnormal and also that H is $1-$stable.
Now if $H \cap E = \emptyset$, then $E \subseteq H^c$ and that would make E abnormal as well by Property 3.12, in which case we get that $B(H|E)$ holds by Corollary 5.2.
Therefore: $\emptyset \neq H \cap E \subseteq H$ and thus $B(H|E)$ holds, since H is $1-$stable. □

As Van Fraassen mentions ([25]) the idea of a priori is the opposite of the idea of abnormal sets. Therefore it makes sense to expect something analogous w.r.t. conditional belief. And we do have this analogy, since *everything is believed given an abnormal set* while *an a priori proposition is believed given anything.*

PROPERTY 5.6 (Closure of conditional belief under finite intersections). *For $E, H, H' \in \mathcal{F}$ if $B(H|E), B(H'|E)$ hold, then $B(H \cap H'|E)$ holds as well.*

PROOF. Take $E, H, H' \in \mathcal{F}$ and assume that $B(H|E)$ and $B(H'|E)$ hold.
We have two cases:

- Case 1. E is abnormal. Then by the definition of conditional belief, we immediately get: $B(H \cap H'|E)$, as desired.
- Case 2. E is normal. This, together with $B(H|E)$ gives us that $\exists S$: stable set such that $\emptyset \neq S \cap E \subseteq H$.
  Since $B(H'|E)$ holds, we get that $\exists S'$: stable set such that $\emptyset \neq S' \cap E \subseteq H'$.
  Now consider: $S'' = S' \cap S$. By Observation 4.7 S'' is stable as well and therefore $S'' \neq \emptyset$.
  We have that $S'' \cap E \neq \emptyset$, since $S \cap E \neq \emptyset$ and $S' \cap E \neq \emptyset$.
  Also $S'' \cap E \subseteq H$ and $S'' \cap E \subseteq H'$, since $S \cap E \subseteq H$ and $S' \cap E \subseteq H'$.
  Therefore: $\emptyset \neq S'' \cap E \subseteq H \cap H''$ and hence $B(H \cap H'|E)$ holds.

□

PROPERTY 5.7 (Consistency of conditional belief w.r.t. normal sets). *For $E \in \mathcal{F}$, if E is normal then $\neg B(\emptyset|E)$.*

PROOF. Assume that $B(\emptyset|E)$ holds for E normal set in $\mathcal{F}$.
Then $\exists S$: r-stable set such that: $\emptyset \neq S \cap E \subseteq \emptyset$.
But this is of course a contradiction. e □

PROPERTY 5.8 (Monotonicity of conditional belief). *Take $A, B \in \mathcal{F}$ such that $A \subseteq B$. If $B(A|E)$ holds, then $B(B|E)$ holds as well.*

Proof. We have two cases:

- Case 1. E is abnormal. Then $B(B|E)$ holds by definition, as desired.
- Case 2. E is normal. Then $\exists S$: r-stable set such that $S \cap E \subseteq A$.
  But since $A \subseteq B$, then $S \cap E \subseteq B$ as well, entailing that $B(B|E)$ holds as well.

$\square$

Lemma 5.9. *If for $E, H \in \mathcal{F}$ we have $B(H|E)$, then $P(H|E) \geq r$.*

Proof. Take $E, H \in \mathcal{F}$ such that $B(H|E)$ holds.
We have two cases:

- Case 1. E is abnormal. Then $P(H|E) = 1 \geq r$.

- Case 2. E is normal. Then $\exists S$: r-stable set such that $\emptyset \neq S \cap E \subseteq H$.
  Hence since E is normal we have that $P^E(\cdot)$ is a probability measure and therefore $P(H|E) \geq P(S \cap E|E) \geq r$, since S is r-stable and $S \cap E \neq \emptyset$.

$\square$

CHAPTER 6

Probabilistic frames

In this chapter we essentially switch from the single agent case we have been considering so far to the multi agent case. We do so by defining the structures called probabilistic frames. These structures will give us the semantics for the Logic of Conditional Belief we will present in the next chapter.

As the name implies, a probabilistic frame is based on the notion of a conditional probability space.

First, we consider a set of agents Ag and we define a function $\mathbf{r}$ that assigns to each $i \in Ag$, i.e. to each one of these agents a number $r_i \in (\frac{1}{2}, 1]$. This number r_i serves as the agent's threshold, representing her own degree of gullibility/mistrust.

Now the set of possible worlds is divided into partitions Π_i, one for each of our agents i. Moreover, we are not going to have a single two-place probability function. Instead, we will have a two-place probability function for each one of our agents at each one of our worlds. Or in other words, we will have a function P_i for each one of our agents that will assign a two-place probability function P_i^w at every possible world. Therefore, we will be talking about the two-place probability function *of agent i at world w*. This function P_i looks similar to B-S's function g_i in [12], [13].

Furthermore, we will define an operator $B_i(H|E)$ that will stand for "the set of possible worlds in which agent i believes H given E". This operator will depend on agent's i r_i-stable sets w.r.t. the two-place probability function P_i^w.

Finally, it is also important to note that we will impose certain constraints on the partition sets and on the two-place probability functions to ensure that the sets of the form $B_i(H|E)$ for $H, E \in \mathcal{F}$ are measurable.

We begin by taking W to be a set of possible worlds and $\mathcal{F}$ a $\sigma - algebra$ on W.

Moreover, consider a set of agents $Ag = \{1, ..., n\}$.

Now for each agent i we will consider a partition Π_i over W ([8]).

For each $w \in W$, we will use $w(i)$ to denote the information cell of w for agent i induced by the partition Π_i.

Moreover, define the following relation over W induced by Π_i.

Definition 6.1. *For $w, v \in W$ define the relation:*

$$w \sim_i v \text{ iff } w(i) = v(i).$$

Definition 6.2. *We will say that $Y \subseteq W$ is* closed under $\sim_i$ *iff*

$$\forall w, v(w \in Y, w \sim_i v \Rightarrow v \in Y).$$

Definition 6.3. *The structure: $M = (W, \mathcal{F}, \Pi_i, \mathbf{r}, P_i)$, such that:*

- *W is a set of possible worlds*
- *$\mathcal{F}$ a $\sigma-$algebra on W*
- *Π_i partitions of W for $i \in Ag$, such that $\forall i \in Ag$, $\forall Y \subseteq W$ such that Y is closed under $\sim_i$ we have: $Y \in \mathcal{F}$,*
- *$\mathbf{r} : Ag \to (\frac{1}{2}, 1]$ a function assigning a number $r_i \in (\frac{1}{2}, 1]$ to each $i \in Ag$,*
- *$P_i : W \to (\mathcal{F} \times \mathcal{F} \to [0, 1])$ a function that assigns to each world $w \in W$ a two-place probability function P_i^w over $\mathcal{F} \times \mathcal{F}$ as defined above such that the following conditions are satisfied:*
  - *(a) $(W, \mathcal{F}, P_i^w)$ is a non-trivial conditional probability space,*
  - *(b) $W - w(i)$ is abnormal,*
  - *(c) if $w' \in w(i)$, then $P_i^{w'} = P_i^w$*

will be called a probabilistic frame.

The restriction we imposed on the partition sets (3rd clause) is needed to ensure that our operator for conditional belief will be measurable.

Concerning the function P_i, the idea is that for each $w \in W$, P_i induces a non-trivial conditional probability space $(W, \mathcal{F}, P_i^w)$, with the two-place probability function P_i^w forcing all subsets of $w(i)^c$ to be abnormal. Notice that this in turn entails that the set $w(i)$ as well as all the sets $X \in \mathcal{F}$ such that: $w(i) \subseteq X$ are a priori sets w.r.t. P_i^w. Moreover, we also demanded that $P_i^{w'} = P_i^w$, for all $w' \in w(i)$. This entails that agent i assigns the same probabilities to all states in W, while at any state of the same information cell induced by the partition. We need this constraint in order to ensure that the conditional belief operator will behave as we desire and more precisely that it will be closed under $\sim_i$, which will in turn entail that it is measurable, by the way we defined the partitions.

Now we will define the following operator $B_i : \mathcal{F} \times \mathcal{F} \to \mathcal{F}$ for conditional belief:

Definition 6.4. *For $E, H \in \mathcal{F}$ define:*
$B_i(H|E) := \{w|\text{either } P_i^w(\emptyset|E) = 1 \text{ or } \exists S : r_i\text{−stable set w.r.t. } P_i^w : S \cap E \text{ is normal and } S \cap E \subseteq H\}$

Conditional belief now becomes dependent on the information that the agent possesses about the state. The interpretation is that agent i believes H given E at all states w in which either E is abnormal w.r.t. his probability function P_i^w, or he has some r_i-stable belief S w.r.t. P_i^w such that $\emptyset \neq S \cap E \subseteq H$. Therefore, we now need to specify a state w when talking about conditional belief: "H is believed given E by agent i *at* state w".

Finally we show that for two measurable sets E, H $B(H|E)$ is also measurable.

PROPOSITION 6.5. *If $E, H \in \mathcal{F}$ then $B_i(H|E) \in \mathcal{F}$.*

PROOF. Pick $E, H \in \mathcal{F}$.
We want to show that $B_i(H|E) \in \mathcal{F}$.
We will use the following property:

PROPERTY 6.6. *$B_i(H|E)$ is closed under $\sim_i$.*[6]

PROOF. Pick $w \in B_i(H|E)$ and $v \in W$ such that $w \sim_i v$.
I want to show that $v \in B_i(H|E)$.
Since $v \in w(i)$ we get that $P_i^v = P_i^w$, by the properties of P_i^w.
Therefore $v \in B_i(H|E)$.
Hence $B_i(H|E)$ is indeed closed under $\sim_i$. □

Now, the definition of the partition Π_i entails that $B_i(H|E) \in \mathcal{F}$. □

EXAMPLE 3. We will now present an example from game theory.
Consider the game $\mathcal{G} = (I, (S_i, \pi_i)_{i \in I})$:

	$\mathcal{A}$	$\mathcal{B}$
$\mathcal{C}$	2,2	0,0
$\mathcal{D}$	0,0	1,1

with $I = \{R, C\}$, the row player R and the column player C, $S_R = \{\mathcal{C}, \mathcal{D}\}$ the strategies of player R, $S_C = \{\mathcal{A}, \mathcal{B}\}$ the strategies of player C and the payoff functions $\pi_i : S_i \times S_{-i} \to \mathbb{R}$, for $i \in I$, as in the figure above.

Now consider the following belief matrix:

	$t_{\mathcal{A}}$	$t'_{\mathcal{A}}$	$t_{\mathcal{B}}$	$t'_{\mathcal{B}}$
$t_{\mathcal{C}}$	$0; 0$	$0; \frac{1}{12}$	$1; \frac{1}{12}$	$0; \frac{1}{12}$
$t'_{\mathcal{C}}$	$\frac{1}{12}; 0$	$\frac{3}{12}; \frac{3}{12}$	$\frac{4}{12}; \frac{3}{12}$	$\frac{4}{12}; \frac{3}{12}$
$t_{\mathcal{D}}$	$\frac{1}{12}; 1$	$\frac{3}{12}; \frac{4}{12}$	$\star\frac{4}{12}; \frac{4}{12}$	$\frac{4}{12}; \frac{4}{12}$
$t'_{\mathcal{D}}$	$\frac{1}{12}; 0$	$\frac{3}{12}; \frac{4}{12}$	$\frac{4}{12}; \frac{4}{12}$	$\frac{4}{12}; \frac{4}{12}$

with $t_{\mathcal{A}}$, $t'_{\mathcal{A}}$, $t_{\mathcal{B}}$, $t'_{\mathcal{B}}$ the types of player C and $t_{\mathcal{C}}$, $t'_{\mathcal{C}}$, $t_{\mathcal{D}}$, $t'_{\mathcal{D}}$ the types of player R.

Moreover, we have a function $P_i : T_i \to (T_{-i} \to [0, 1])$, for $i \in I$, a function assigning a classical probability measure to each type of player i over his opponent's types.
For example $P_R^{(t_{\mathcal{C}})}(t_{\mathcal{A}}) = 0$ and is the probability that player R's type $t_{\mathcal{C}}$ assigns to player C's type $t_{\mathcal{A}}$

[6]This is introduced as a proposition instead of a claim because we will later refer to it separately.

Now notice that the only irrational types are $t_{\mathcal{C}}$ and $t_{\mathcal{A}}$.
Therefore, if we examine the state $(t_{\mathcal{D}}, t_{\mathcal{B}})$, we get that:

$$P_R^{(t_{\mathcal{D}})}(\text{rationality}_C) = \frac{11}{12}$$
$$P_C^{(t_{\mathcal{B}})}(\text{rationality}_R) = \frac{11}{12}$$

with $P_R^{(t_{\mathcal{D}})}(\text{rationality}_C)$ being the probability that player R's type $t_{\mathcal{D}}$ assigns to player C's rationality (w.r.t. Definition 2.6) — and analogously for $P_C^{(t_{\mathcal{B}})}$.

These probabilities come up by adding the probabilities that each player's type in the state $(t_{\mathcal{D}}, t_{\mathcal{B}})$ assigns to the other player's *rational* types.
Take for example player R. As we said, the only irrational types here are $t_{\mathcal{C}}$ and $t_{\mathcal{A}}$. Now at state $(t_{\mathcal{D}}, t_{\mathcal{B}})$, player R's type $t_{\mathcal{D}}$ assigns probability $\frac{3}{12}$ to player C's type $t'_{\mathcal{A}}$, $\frac{4}{12}$ to player C's type $t_{\mathcal{B}}$ and $\frac{4}{12}$ to player C's type $t'_{\mathcal{B}}$. Adding these, the probability measure $P_R^{(t_{\mathcal{D}})}$ assigns probability $\frac{11}{12}$ to the set $\{t'_{\mathcal{A}}, t_{\mathcal{B}}, t'_{\mathcal{B}}\}$, which is the set of the rational types of agent C.
Analogously for player C's type $t_{\mathcal{B}}$ probability measure $P_C^{(t_{\mathcal{B}})}$.

Hence, we have that both players R and C assign probability $\frac{11}{12} < 1$ to each other's rationality at state $(t_{\mathcal{D}}, t_{\mathcal{B}})$. Now we will check that this belief in rationality is $\frac{3}{5}$-stable w.r.t. to $P_R^{(t_{\mathcal{D}})}$ and $P_C^{(t_{\mathcal{B}})}$. To do that, we will use Corollary 4.15.
Let's begin with player R.
According to this Corollary, we need to check that the probability of every subset of the set $\{t'_{\mathcal{A}}, t_{\mathcal{B}}, t'_{\mathcal{B}}\}$ has a probability bigger than $\frac{3}{2}$ [7] times the probability of $t_{\mathcal{A}}$ (w.r.t. $P_R^{(t_{\mathcal{D}})}$): the only irrational type of player C.
Therefore, we have:

- For the Row player R:
  $$P_R^{(t_{\mathcal{D}})}(t'_{\mathcal{A}}) = \frac{3}{12} > \frac{3}{2} P_R^{(t_{\mathcal{D}})}(\{t_{\mathcal{A}}, t'_{\mathcal{A}}, t_{\mathcal{B}}, t'_{\mathcal{B}}\} \backslash \{t'_{\mathcal{A}}, t_{\mathcal{B}}, t'_{\mathcal{B}}\}) = \frac{3}{2} P_R^{(t_{\mathcal{D}})}(t_{\mathcal{A}}) = \frac{3}{24}$$
  (since $t_{\mathcal{A}}$ is the only irrational type of player C);
  likewise for $P_R^{(t_{\mathcal{D}})}(t_{\mathcal{B}}), P_R^{(t_{\mathcal{D}})}(t'_{\mathcal{B}})$.
- Analogous for player C, where the only irrational type of player R is $t_{\mathcal{C}}$.

Therefore, we have that the belief in rationality is $\frac{3}{5}$-stable w.r.t. $P_R^{(t_{\mathcal{D}})}$ and $P_C^{(t_{\mathcal{B}})}$.

Finally, we would like to make some comments concerning B-S's strong and conditional belief. In 2.7 we have that:

$$B_{i,h}(E) = \{(s, \theta, t) \in \Omega : g_{i,h}(t_i)(proj_{\Omega_{-i}} E) = 1\}, \text{ for } h \in \mathcal{F},\ E \in \mathcal{A}_{-i}.$$

The analogous definition of B-S's conditional belief in our framework is:

$$B_{i,E}(H) = \{w \in W : P_i^w(H|E) = 1\}, \text{ for } E, H \in \mathcal{F}.$$

Moreover, in 2.8 we have that:

$$SB_i(E) = \bigcap_{h \in \mathcal{H}: E \cap [h] \neq \emptyset} B_{i,h}(E).$$

[7] bigger than $\frac{3}{2}$, because $\frac{r}{1-r} = \frac{\frac{3}{5}}{1-\frac{3}{5}} = \frac{15}{10} = \frac{3}{2}$

Therefore, we have that in our framework:

$$SB_i(E) = \bigcap_{\{H \in \mathcal{F} : E \cap H \neq \emptyset\}} B_{i,H}(E) \text{ and equivalently:}$$

$$SB_i(E) = \bigcap_{\{H \in \mathcal{F} : E \cap H \neq \emptyset\}} \{w \in W : P_i^w(E|H) = 1\} \text{ and equivalently:}$$

$$w \in SB_i(E) \text{ iff } P_i^w(E|H) = 1 \text{ for all } H \in \mathcal{F} : E \cap H \neq \emptyset \text{ and equivalently:}$$

"E is strongly-believed at w iff E is 1-stable w.r.t. P_i^w"

Therefore, B-S's strong belief is our notion of 1-stable belief.

Now for their notion of conditional belief, it is interesting to consider the following.

In Example 3 from before, we showed that player R's belief in player C's rationality at state $(t_{\mathcal{D}}, t_{\mathcal{B}})$ is $\frac{3}{5}$-stable. Therefore, this means that player R believes that player C is rational at state $(t_{\mathcal{D}}, t_{\mathcal{B}})$ with our notion of conditional belief. However, $P_R^{(t_{\mathcal{D}})}(\text{rationality}_C) = \frac{11}{12} < 1$. Therefore player R does not believe that player C is rational in B-S's sense.

CHAPTER 7

Logic of r-stable conditional beliefs

In this chapter we introduce a formal language for the notion of conditional belief. We present our logic of r-stable conditional beliefs, to which we will be referring as **rCBL**.

In section 7.1 we introduce the language and the semantics of **rCBL**. The language is the standard language of epistemic logic along with a new modality $B_i^{\phi}\psi$ standing for "agent i's belief in ψ conditioning on ϕ". The semantics are given by the structures called probabilistic models, which are essentially the frames we defined in the previous chapter enhanced with a valuation.

In section 7.2 we present our axiom system, which is almost the same with Board's axiom system in his paper [18]. Board also includes the operator $B_i\phi$ in his language (as a primitive symbol) and the axiom he calls **Triv**: $B_i\phi \Leftrightarrow B_i^{true}\phi$. His operator $B_i\phi$ stands for the notion of absolute belief. We decided to define absolute belief in terms of the conditional belief operator, to simplify our language (having only one primitive symbol) and our axiom system. Board argues ([18, p. 55]) that having absolute belief as a primitive symbol might be useful in game theoretic applications as it could encode the beliefs of the players prior to the game. This is a fair point, however we decided in favor of keeping our language — and axioms — as simple as possible. Furthermore, we included the axiom **D**: $\neg B_i\bot$ (Consistency of Belief) which will express that our space is not trivial in the sense of Definition 3.6.

In section 7.3 we prove the soundness of our axiom system w.r.t. our probabilistic models. We provide a separate proof that each axiom is sound.

Finally, in section 7.4 we prove completeness of our axioms w.r.t. our probabilistic semantics. The completeness proof proceeds in the following steps. First, we present Board's semantics, consisting of *belief revision structures*, which look like a generalized version of Stalnaker's structures ([18, p. 52]). We also add one more condition in his structures, referring to our axiom **D**. Second, we present Board's completeness proof, in which we have added a final clause in the proof that the canonical model is a belief revision structure — that the canonical model satisfies our extra condition. Finally, we prove a truth preserving lemma between Board's semantics and ours. We show that for any of his belief revision structures, we can construct a probabilistic model with the same set of worlds and same valuation and such that the same sentences of our language are satisfied at the same worlds in the two models. Then, completeness of our axiom system w.r.t. our probabilistic semantics follows from completeness of our axiom system w.r.t. Board's semantics.

1. Syntax and Semantics

The language of **rCBL** is defined as follows:

DEFINITION 7.1. *Consider a set of atomic sentences At and a set of agents: $Ag = \{1, .., n\}$.*
Our language $\mathcal{L}$ is a set of formulas ϕ of **rCBL** *and is defined recursively:*

$$\phi ::= p|\neg\phi|\phi \wedge \phi|B_i^{\phi}\phi$$

for $p \in At$, $i \in Ag$. Call our language $\mathcal{L}$.
Our language is that of epistemic logic augmented by adding the operators B_i^{ϕ}, for $i \in Ag$ that tell us what agent i believes after learning (observing) ϕ.

Moreover, we use the standard abbreviations: $\phi \vee \psi$ for $\neg(\neg\phi \wedge \neg\psi)$, $\phi \to \psi$ for $\neg\phi \vee \psi$ and $\phi \leftrightarrow \psi$ for $(\phi \to \psi) \wedge (\psi \to \phi)$. Furthermore define: $\bot = p \wedge \neg p$ for $p \in At$ and $\top = \neg\bot$. Finally, we introduce the following abbreviation: $B_i\phi := B_i^{\top}\phi$.

Now for our semantics, consider the following structure:

DEFINITION 7.2. *A* probabilistic model *M is a tuple: $M = (F, ||\cdot||)$, such that $F = (W, \mathcal{F}, \Pi_i, \mathbf{r}, P_i)$ is a probabilistic frame and $||\cdot|| : \mathcal{L} \to \mathcal{F}$ is a function assigning a set of worlds to each atomic proposition.*
Therefore, M is of the form:

$$M = (W, \mathcal{F}, \Pi_i, \mathbf{r}, P_i, ||\cdot||).$$

Finally, we require that for all $p \in At$, we have that $||p|| \in \mathcal{F}$.

Truth in probabilistic models is defined as follows:

DEFINITION 7.3. *Let $M = (W, \mathcal{F}, \Pi_i, \mathbf{r}, P_i, ||\cdot||)$ a probabilistic model, $w \in W$, $i \in Ag$ and $p \in At$. The relation $\models$ between pairs (M, w) and formulas $\phi \in \mathcal{L}$ is defined as follows:*

- $(M, w) \models p$ *iff* $w \in ||p||$,
- $(M, w) \models \neg\phi$ *iff* $w \in ||\phi||^c$,
- $(M, w) \models \phi \wedge \psi$ *iff* $w \in ||\phi|| \cap ||\psi||$,
- $(M, w) \models B_i^{\theta}\phi$ *iff* $w \in B_i^{||\theta||}||\phi||$

Finally we have the following important proposition:

PROPOSITION 7.4. *$||\cdot||$ is a well-defined function from $\mathcal{L}$ to $\mathcal{F}$.*

PROOF. We proceed by induction on the length of ϕ.
We have that $||p||$ is well defined and that $||p|| \in \mathcal{F}$.
We have that if $||\phi||$ is well defined and $||\phi|| \in \mathcal{F}$, then $||\phi||^c$ is well defined and $||\phi||^c \in \mathcal{F}$ by the properties of a $\sigma - algebra$.
If $||\phi||, ||\psi||$ are well-defined and $||\phi||, ||\psi|| \in \mathcal{F}$ we have that $||\phi|| \cap ||\psi||$ is well-defined and that $||\phi|| \cap ||\psi|| \in \mathcal{F}$, once again by the properties of a $\sigma - algebra$.

Assume that $||\theta||, ||\phi||$ are well-defined and $||\theta||, ||\phi|| \in \mathcal{F}$.
Then using Proposition 6.5 and the fact that $||\theta||$ is well-defined we get that $B_i^{||\theta||}||\phi||$ is well-defined and also in $\mathcal{F}$. $\square$

2. Axiom System

Consider $i \in Ag$ and ϕ, ψ formulas in $\mathcal{L}$.
Our axioms and inference rules are the following, along with **MP**:

Taut	$true$
Distr	$B_i^{\phi}\psi \wedge B_i^{\phi}(\psi \to \chi) \Rightarrow B_i^{\phi}\chi$
Succ	$B_i^{\phi}\phi$
IE(a)	$B_i^{\phi}\psi \Rightarrow (B_i^{\phi\wedge\psi}\chi \Leftrightarrow B_i^{\phi}\chi)$
IE(b)	$\neg B_i^{\phi}\neg\psi \Rightarrow (B_i^{\phi\wedge\psi}\chi \Leftrightarrow B_i^{\phi}(\psi \to \chi))$
RE	from ψ infer $B_i^{\phi}\psi$
LE	from $\phi \Leftrightarrow \psi$ infer $B_i^{\phi}\chi \Leftrightarrow B_i^{\psi}\chi$
PI	$B_i^{\phi}\psi \Rightarrow B_i^{\chi}B_i^{\phi}\psi$
NI	$\neg B_i^{\phi}\psi \Rightarrow B_i^{\chi}\neg B_i^{\phi}\psi$
D	$\neg B_i^{\top}\bot$

Call the axiomatic system above **BRSID**.

As mentioned above, we have adopted Board's axiomatic system ([18, p. 54]). However, note that as we mentioned in the introduction of this chapter, Board also has the following axiom **Triv** : $B_i\phi \Leftrightarrow B_i^{true}\phi$, since for him $B_i\phi$ is a primitive symbol in the language. Moreover, our axiom **D** is a weakening of Board's axiom **WCon** : $\phi \Rightarrow \neg B_i^{\phi}\bot$ ([18, p. 62]).

Now $true$ stands for any propositional tautology, while $false$ stands for $\neg true$. As Board argues ([18, p. 55]) this system is close to the system K of epistemic logic, corresponding roughly to the AGM axioms of belief revision. Now **Taut, Dist, RE** and **D** are already familiar from epistemic logic (tautologies, **K**, Necessitation and Consistency of Belief). **IE(a)** states that an agent does not revise his beliefs if he learns something that he already believed. **IE(b)** states that if the agent learns something consistent with her original beliefs, then she adds the new information to her existing beliefs, closes under **MP** and thus forms her revised beliefs ([18, p. 55]). Moreover, our axioms **PI** and **NI** are Board's axioms **TPI** and **TNI** ([18, p. 60]) and are the axioms for total positive and negative introspection, stating that an agent has complete introspective access to his own beliefs.

Finally, **Distr** and **RE** jointly correspond to the AGM axiom (K^*1),[8] **Succ** is the analogue of (K^*2), **IE(a)** is implied by (K^*7) and (K^*8) in the presence of (K^*5), **IE(b)** corresponds to (K^*7) and (K^*8) and **LE** to (K^*6). Also, in Board's setting, **IE(b)** also corresponds to (K^*3) and (K^*4) in the presence of **Triv** and Board argues that **Triv** itself is implied by the AGM axioms ([18, p. 55]).

3. Soundness

In this section we will prove that each axiom of **BRSID** is sound.
Consider therefore a set of agents $Ag = \{1, ..., n\}$ and a probabilistic model $M = (W, \mathcal{F}, \Pi_i, \mathbf{r}, P_i, || \cdot ||)$.
Here are the proofs that the axioms above are sound:

Distr:$B_i^{\phi}\psi \wedge B_i^{\phi}(\neg\psi \vee \chi) \Rightarrow B_i^{\phi}\chi$

Proof. Pick $w \in W$.
Assume that $B_i^{\phi}\psi$ and $B_i^{\phi}(\neg\psi \vee \chi)$ hold at w.

- Case 1. $||\phi||$ is abnormal w.r.t. P_i^w. Then $B_i^{\phi}\chi$ holds, by Corollary 5.2.
- Case 2. $||\phi||$ is normal w.r.t. P_i^w. Then since $w \in B_i^{\phi}\psi$ we derive that:
  $\exists S$: r_i-stable set w.r.t. P_i^w such that $\emptyset \neq S \cap ||\phi|| \subseteq ||\psi||$.
  Now since $w \in B_i^{\phi}(\neg\psi \vee \chi)$, we get that:
  $\exists S'$: r_i-stable set w.r.t. P_i^w such that: $\emptyset \neq S' \cap ||\phi|| \subseteq ||\neg\psi|| \cup ||\chi||$.
  Now consider the set $S'' = S' \cap S$.
  By Observation 4.7 in chapter 4, S'' is r_i-stable, w.r.t. P_i^w.
  Now $S'' \cap ||\phi|| = S \cap S' \cap ||\phi|| \neq \emptyset$ since $S \cap ||\phi|| \neq \emptyset$, $S' \cap ||\phi|| \neq \emptyset$ and $S'' \neq \emptyset$.
  Now pick $x \in S'' \cap ||\phi||$.
  Assume that $x \in ||\neg\psi||$.
  Then since $x \in S'' \cap ||\phi||$, we get that $x \in S \cap ||\phi||$ and since $S \cap ||\phi|| \subseteq ||\psi||$, we get that $x \in ||\psi||$. Contradiction.
  Therefore if $x \in S'' \cap ||\phi||$, we have that $x \notin ||\neg\psi||$. However if $x \in S'' \cap ||\phi||$, then $x \in S' \cap ||\phi||$ and $S' \cap ||\phi|| \subseteq ||\neg\psi|| \cup ||\chi||$. Hence $x \in ||\chi||$. Therefore $\emptyset \neq S'' \cap ||\phi|| \subseteq ||\chi||$.
  This shows that $B_i^{\phi}\chi$ holds at w and our proof is complete.

□

[8]AGM axioms (Under the numbering system of [27], as presented in Appendix A of [18]. For a more detailed account look at [27, 2].)

(K^*1) K_{ϕ}^* is a belief set
(K^*2) $\phi \in K_{\phi}^*$
(K^*3) $K_{\phi}^* \subseteq K_{\phi}^+$
(K^*4) if $\neg\phi \notin K$ then $K_{\phi}^+ \subseteq K_{\phi}^*$
(K^*5) $K_{\phi}^* = K_{false}$ iff ϕ is logically inconsistent
(K^*6) if $\phi \leftrightarrow \psi$ then $K_{\phi}^* = K_{\psi}^*$
(K^*7) $K_{\phi\wedge\psi}^* \subseteq (K_{\phi}^*)_{\psi}^+$
(K^*8) if $\neg\psi \notin K_{\phi}^*$ then $(K_{\phi}^*)_{\psi}^+ \subseteq K_{\phi\wedge\psi}^*$

Unless otherwise specified, for the rest of this section all the sets are abnormal, normal, r_i-stable, contingent or a priori w.r.t. P_i^w.

Succ:$B_i^{\phi}\phi$

PROOF. Pick a world $w \in W$.

- Case 1. $||\phi||$ is abnormal. Then $B_i^{\phi}\phi$ holds at w, by Corollary 5.2.
- Case 2. $||\phi||$ is normal. Take W to be the whole space of possible worlds.
  Now W is normal and $1-\mathit{stable}$.
  Moreover, we have that: $W \cap ||\phi|| = ||\phi||$.
  Therefore $B_i^{\phi}\phi$ holds at w.

□

IE(a):$B_i^{\phi}\psi \Rightarrow (B_i^{\phi\wedge\psi}\chi \Leftrightarrow B_i^{\phi}\chi)$.

PROOF. Pick $w \in W$ and assume that $B_i^{\phi}\psi$ holds at w.

- Case 1. $||\phi||$ is abnormal. Then $||\phi|| \cap ||\psi|| \subseteq ||\phi||$ and therefore $||\phi|| \cap ||\psi||$ is abnormal as well.
  Hence $B_i^{\phi\wedge\psi}\chi \Leftrightarrow B_i^{\phi}\chi$ holds at w, by Corollary 5.2.
- Case 2. $||\phi||$ is normal. Assume that $B_i^{\phi\wedge\psi}\chi$ holds at w.
  Then by $B_i^{\phi}\psi$, we have that $\exists S$: r_i-stable set such that: $\emptyset \neq S \cap ||\phi|| \subseteq ||\psi||$.
  Moreover, by $B_i^{\phi\wedge\psi}\chi$, we have that $\exists S'$: r_i-stable set such that $\emptyset \neq S' \cap ||\phi|| \cap ||\psi|| \subseteq ||\chi||$.
  Now consider the set: $S'' = S' \cap S$.
  By Observation 4.7, S'' is r_i-stable as well.
  Now take the set $S'' \cap ||\phi||$.
  Since $S \cap ||\phi|| \neq \emptyset$, $S' \cap ||\phi|| \neq \emptyset$ and $S'' \neq \emptyset$, we have that $S'' \cap ||\phi|| \neq \emptyset$.
  Now pick $x \in S'' \cap ||\phi||$.
  Then $x \in S \cap ||\phi||$ and therefore $x \in ||\psi||$.
  Now $x \in S' \cap ||\phi|| \cap ||\psi||$ as well. Therefore $x \in ||\chi||$.
  Hence: $\emptyset \neq S'' \cap ||\phi|| \subseteq ||\chi||$. Thus $B_i^{\phi}\chi$ holds at w.

Now for the other direction, assume that $B_i^{\phi}\chi$ holds at w.

- Case 1. $||\phi|| \cap ||\psi||$ is abnormal. Then $B_i^{\phi\wedge\psi}\chi$ holds at w as well.
- Case 2. $||\phi|| \cap ||\psi||$ is normal. Then:
  by $B_i^{\phi}\psi$, we have that $\exists S$: r_i-stable set , such that: $\emptyset \neq S \cap ||\phi|| \subseteq ||\psi||$.
  Also, by $B_i^{\phi}\chi$, we have that $\exists S'$: r_i-stable set, such that: $\emptyset \neq S' \cap ||\phi|| \subseteq ||\chi||$.
  Once again consider the set $S'' = S' \cap S$. We know that S'' is r_i-stable as well, by Property 4.7.
  Take the set $S'' \cap ||\phi|| \cap ||\psi||$.
  Now $S'' \cap ||\phi||$ is non-empty.
  Pick $x \in S'' \cap ||\phi||$.
  Then $x \in S \cap ||\phi||$. Therefore $x \in ||\psi||$.
  Hence $S'' \cap ||\phi|| \cap ||\psi|| \neq \emptyset$.
  Now if $x \in S'' \cap ||\phi|| \cap ||\psi||$, then $x \in S'' \cap ||\phi||$.

Therefore $x \in S' \cap ||\phi||$. Hence $x \in ||\chi||$.
Thus $\emptyset \neq S'' \cap ||\phi|| \cap ||\psi|| \subseteq ||\chi||$ and hence $B_i^{\phi\wedge\psi}\chi$ holds at w.

□

IE(b):$\neg B_i^{\phi}\neg\psi \Rightarrow (B_i^{\phi\wedge\psi}\chi \Leftrightarrow B_i^{\phi}(\neg\psi \vee \chi))$.

PROOF. Pick a world $w \in W$.
Assume that $\neg B_i^{\phi}\neg\psi$ holds at world w. If $||\phi||$ is abnormal, then $\neg B_i^{\phi}\neg\psi$ can not hold since by Corollary 5.2 everything is believed when conditioning on abnormal sets.

Therefore assume that $||\phi||$ is normal and that $B_i^{\phi\wedge\psi}\chi$ holds at w.
Now we need to consider the following cases:

- Case 1. $||\psi||$ is abnormal. Then $||\psi||^c$ is a priori. Now if $||\phi|| \cap ||\psi||^c$ is abnormal, then by 4.13 we have that $||\phi||$ is abnormal, which contradicts our assumptions.
  Hence $||\phi|| \cap ||\psi||^c$ is normal and $||\psi||^c$ is r_i-stable as a priori.
  Moreover, $||\phi|| \cap ||\psi||^c \subseteq ||\psi||^c$, hence $B_i^{\phi}\neg\psi$ holds. Contradiction.
- Case 2. $||\psi||$ is normal. Then we have two more cases:
  - First case: $||\phi \cap \psi||$ is normal. Then since $\neg B_i^{\phi}\neg\psi$ holds at w, we get that $\nexists S$: r_i-stable set such that $\emptyset \neq S \cap ||\phi|| \subseteq ||\neg\psi||$.
    Therefore, for all S : r_i-stable sets, we have that if $S \cap ||\phi||$ is normal, then $S \cap ||\phi|| \cap ||\psi|| \neq \emptyset$.
    Now we also have that $\exists S'$: r_i-stable set such that: $\emptyset \neq S' \cap ||\phi|| \cap ||\psi|| \subseteq ||\chi||$.
    Now pick $x \in S' \cap ||\phi||$.
    Then either $x \in ||\psi||$ or $x \in ||\psi||^c$.
    If $x \in ||\psi||$ then $x \in ||\chi||$.
    Hence $\emptyset \neq S \cap ||\phi|| \subseteq ||\neg\psi|| \cup ||\chi||$ and $B_i^{\phi}(\neg\psi \vee \chi)$ holds at w.
  - Second case: $||\phi \cap \psi||$ is abnormal.
    Then $(||\phi|| \cap ||\psi||)^c = (||\phi||^c \cup ||\psi||^c) = K$ is a priori and $1 - stable$.
    Now $K \cap ||\phi|| = ||\phi|| - ||\psi||$. If $||\phi|| - ||\psi|| = \emptyset$, then $||\phi|| \cap ||\psi|| = ||\phi||$ and hence $||\phi||$ is abnormal, contradicting our assumption.
    Therefore $||\phi|| - ||\psi|| = K \cap ||\phi|| \neq \emptyset$.
    But now we have that $\neg B_i^{\phi}\neg\psi$ holds at w.
    Hence $K \cap ||\phi|| \cap (||\neg\psi||)^c = K \cap ||\phi|| \cap ||\psi|| \neq \emptyset$.
    But this means that $(||\phi|| \cap ||\psi||)^c \cap (||\phi|| \cap ||\psi||) \neq \emptyset$, which is a contradiction.

Now for the other direction assume that $B_i^{\phi}(\neg\psi \vee \chi)$ holds at w.
Since $\neg B_i^{\phi}\neg\psi$ holds at w, we have that: $\forall S$: r_i-stable sets if $S \cap ||\phi|| \neq \emptyset$ then $S \cap ||\phi|| \cap ||\psi|| \neq \emptyset$.
Now since $B_i^{\phi}(\neg\psi \vee \chi)$ holds at w, we have that: $\exists S'$: r_i-stable set such that: $\emptyset \neq S' \cap ||\phi|| \subseteq ||\neg\psi|| \cup ||\chi||$.
Since $S' \cap ||\phi|| \neq \emptyset$, then $S' \cap ||\phi|| \cap ||\psi|| \neq \emptyset$.
Now: $S' \cap ||\phi|| \cap ||\psi|| \subseteq S' \cap ||\phi||$ and we have that $S' \cap ||\phi|| \subseteq ||\neg\psi|| \cup ||\chi||$. Therefore: $S' \cap ||\phi|| \cap ||\psi|| \subseteq ||\chi||$.
Hence: $\emptyset \neq S' \cap ||\phi|| \cap ||\psi|| \subseteq ||\chi||$ and $B_i^{\phi\wedge\psi}\chi$ holds at w. □

RE: from ψ infer $B_i^\phi\psi$.

Proof. Pick $w \in W$.
If ψ holds, then $||\psi||$ is an 1-stable set.

- Case 1. $||\phi||$ is abnormal. Then $B_i^\phi\phi$ holds at w, by Corollary 5.2.
- Case 2. $||\phi||$ is normal. Then $||\phi|| \cap ||\psi|| \neq \emptyset$.
  For assume otherwise. Then $\exists x \in ||\phi||$ such that $x \notin ||\psi||$.
  But ψ holds everywhere in W and therefore this can not be the case.
  Therefore: $\emptyset \neq ||\phi|| \cap ||\psi|| \subseteq ||\psi||$.
  And hence $B_i^\phi\phi$ holds at w.

□

LE: from $\phi \Leftrightarrow \psi$ infer $B_i^\phi\chi \Leftrightarrow B_i^\psi\chi$.

Proof. Pick $w \in W$.
Assume that $\phi \Leftrightarrow \psi$ holds.
Assume that $B_i^\phi\chi$ holds at w.

- Case 1. $||\phi||$ is abnormal. Then so is $||\psi||$ and hence $B_i^\psi\chi$ holds at w, by Corollary 5.2.
- Case 2. $||\phi||$ is normal. Then $\exists S : r_i$-stable set such that: $\emptyset \neq S \cap ||\phi|| \subseteq ||\chi||$.
  But since $\phi \Leftrightarrow \psi$ holds, then $||\phi|| = ||\psi||$.
  Therefore $\emptyset \neq S \cap ||\psi|| \subseteq ||\chi||$ and $B_i^\psi\chi$ holds at w.

Assume that $B_i^\psi\chi$ holds at w.
Analogous. □

PI:$B_i^\phi\psi \Rightarrow B_i^\chi B_i^\phi\psi$

Proof. Pick $w \in W$.
Assume that $B_i^\phi\psi$ holds in w.

- Case 1. $||\chi||$ is abnormal. Then $B_i^\chi B_i^\phi\psi$ holds at w, by Corollary 5.2.
- Case 2. $||\chi||$ is normal.
  Consider the set $w(i)$, which is $1 - \textit{stable}$.
  Then $\emptyset \neq w(i) \cap ||\chi|| \subseteq ||B_i^\phi\psi||$, since $B_i^\phi\psi$ holds at w.
  Therefore $B_i^\chi B_i^\phi\psi$ holds at w.

□

NI:$\neg B_i^\phi\psi \Rightarrow B_i^\chi\neg B_i^\phi\psi$

Proof. Pick $w \in W$.
Assume that $\neg B_i^\phi\psi$ holds in w.

- Case 1. $||\chi||$ is abnormal. Then $B_i^{\chi}\neg B_i^{\phi}\psi$ holds, by Corollary 5.2.
- Case 2. $||\chi||$ is normal. Then consider the set $w(i)$ which is $1-stable$.
  Then $\emptyset \neq w(i) \cap ||\chi|| \subseteq ||\neg B_i^{\phi}\psi||$, since $\neg B_i^{\phi}\psi$ holds at w.
  Therefore $B_i^{\chi}\neg B_i^{\phi}\psi$ holds at w.

□

D:$\neg B_i^{\top}\bot$

PROOF. Follows directly from the condition of probabilistic frames that $(W, \mathcal{F}, P_i^w)$ is a non-trivial conditional probability space. □

4. Completeness

In this section we prove that **BRSID** is a complete axiomatization w.r.t. our probabilistic models.
The idea is to prove a truth preserving lemma that will connect our probabilistic models with Board's *belief revision structures* and use Board's completeness result.

So we will begin by presenting Board's semantics ([18]).
Consider a set of agents: $Ag = \{1, ..., n\}$ and a set of atomic propositions At.

DEFINITION 7.5. *The structure*

$$M = \langle W, \preccurlyeq, ||\cdot||\rangle, \textit{ such that:}$$

- W *a set of possible worlds,*
- $\preccurlyeq$ *a vector of binary relations over* W*,*
- $||\cdot||$ *is a valuation function, assigning sets of worlds to each atomic proposition*

will be called a belief revision structure.

$\preccurlyeq_i^w$ is the plausibility ordering of agent i at world w.
$x \preccurlyeq_i^w y$ denotes that world y is at least as possible as world x for agent i while at world w.

DEFINITION 7.6. *Define:* $W_i^w := \{x | y \preccurlyeq_i^w x \textit{ for some } y\}$*, the set of all the conceivable worlds for agent* i *at world* w*.*

We assume moreover that:

R1 $\forall i, w : \preccurlyeq_i^w$ is complete and transitive on W_i^w
R2 $\forall i, w : \preccurlyeq_i^w$ is well-founded

R3 $\forall i, w, x, y, z$: if $x \in W_i^w$, then: $y \preccurlyeq_i^x z$ iff $y \preccurlyeq_i^w z$
R4 $\forall i, w : W_i^w \neq \emptyset$

R1 ensures that each plausibility ordering divides all the worlds into ordered equivalence classes. Note that the inconceivable worlds, i.e. those not in W_i^w, are a class unto themselves and are to be considered least plausible ([18, p. 56]).
R2 ensures that our relation is well founded.
R3 says that an agent has the same plausibility ordering in every world that is conceivable to her ([18, p. 61]).
R4 says that for every agent i and every world w, there exists some conceivable world.

Definition 7.7. *For $w \in W, X \subseteq W, i \in Ag$, define:*

$$best_i^w(X) := \{x \in X | \forall y \in X : y \preccurlyeq_i^w x\}.$$

So for some set $X \subseteq W$, $i \in Ag$ and $w \in W$, $best_i^w(X)$ is the set of the most plausible worlds for agent i at state w.

Now for the definition of $\models$ relation between pairs (M, w) and formulas of $\mathcal{L}$:

Definition 7.8. *Take M a belief revision structure.*
For $w \in W$ and $i \in Ag$, p an atomic proposition:

- $(M, w) \models p$ *iff* $w \in ||p||$
- $(M, w) \models \phi \wedge \psi$ *iff* $(M, w) \models \phi$ *and* $(M, w) \models \psi$
- $(M, w) \models \neg\phi$ *iff* $(M, w) \not\models \phi$
- $(M, w) \models B_i^\phi \psi$ *iff* $(M, x) \models \psi$, $\forall x \in best_i^w(||\phi|| \cap W_i^w)$.

No surprises here, the first three clauses are straightforward and the truth conditions for the conditional belief operator are as usual in plausibility models.

Proposition 7.9. **BRSID** *is a sound and complete axiomatization w.r.t. finite belief revision structures satisfying* **R1**, **R2**, **R3**, **R4**.

Proof. Board proves that **BRSI**, which is **BRSID** without our last axiom **D** is a sound and complete axiomatization w.r.t. the set of finite belief revision structures that satisfy **R1**, **R2**, **R3** ([18, p. 72-77]). This means that we need to prove the soundness and completeness of the axiom **D** w.r.t. finite belief revision structures that satisfy **R1**, **R2**, **R3** and the extra property **R4** that we added. Let's call this class of belief revision structures $\mathcal{M}$.

For the soundness proofs check [18, p. 72-77]. Notice that the soundness of the axiom **D** w.r.t. $\mathcal{M}$ is trivial due to our semantic assumption that W_i^w is non-empty (**R4**).

Now for the completeness proof, we decided to present Board's proof, along with the part of the proof that we added concerning the axiom **D** and the property **R4** that we added.

So we will begin the proof with a round of definitions.

Definition 7.10. *We say that a formula ϕ is* **BRSID**-consistent *if $\neg\phi$ is not provable in* **BRSID**.

Definition 7.11. *A finite set of formulas $\{\phi_1, ..., \phi_k\}$ is* **BRSID**-consistent *exactly if $\phi_1 \wedge ... \wedge \phi_k$ is* **BRSID**-consistent *and an infinite set of formulas is* **BRSID**-*consistent if all its finite subsets are* **BRSID**-*consistent.*

Definition 7.12. *Given two sets of formulas $S \subseteq T \subseteq \mathcal{L}$ we say that S is a* maximal **BRSID**-consistent subset of T *if:*

- *it is* **BRSID**-*consistent and*
- *for all $\phi \in T$ but not in S the set $S \cup \{\phi\}$ is not* **BRSID**-*consistent*

Now to prove completeness, Board argues ([18, p. 73]) that it suffices to show that

$(\star)$ Every **BRSID**-consistent formula in $\mathcal{L}$ is satisfiable w.r.t. $\mathcal{M}$

For if $(\star)$ holds and ϕ is a valid formula in $\mathcal{L}$, then if ϕ is not provable in **BRSID** then neither is $\neg\neg\phi$, so by definition $\neg\phi$ is **BRSID**-consistent. Then by $(\star)$ $\neg\phi$ is satisfiable w.r.t. $\mathcal{M}$, contradicting the validity of ϕ w.r.t. $\mathcal{M}$.

Now we proceed with another round of definitions.

Definition 7.13. *Let ϕ be a formula in $\mathcal{L}$. Then define $Sub(\phi)$ to be the set of all the subformulas of ϕ, i.e.: $\psi \in Sub(\phi)$ if:*

- *$\psi = \phi$, or*
- *ϕ is of the form $\neg\phi', \phi' \wedge \phi'', B_i^{\phi'}\phi''$ and $\psi \in Sub(\phi')$ or $\psi \in Sub(\phi'')$*

Definition 7.14. *Let $Sub^+(\phi)$ consist of all the formulas in $Sub(\phi)$ and their negations and conjunctions, i.e. $Sub^+\phi$ is the* smallest *set such that:*

- *if $\psi \in Sub(\phi)$, then $\psi \in Sub^+(\phi)$,*
- *if $\psi, \chi \in Sub^+(\phi)$, then $\neg\psi, \psi \wedge \chi \in Sub^+(\phi)$*

Definition 7.15. *Let $Sub^{++}(\phi)$ consist of*

- *of all formulas of $Sub^+(\phi)$, together with all formulas of the form $B_i^\chi\psi$, where $\psi, \chi \in Sub^+(\phi)$ and*
- *if $\xi \in Sub^+(\phi)$ and $B_i^\chi\psi \in Sub^{++}(\phi)$, then $B_i^\xi B_i^\chi\psi \in Sub^{++}(\phi)$ and $B_i^\xi\neg B_i^\chi\psi \in Sub^{++}(\phi)$*

Definition 7.16. *Let $Sub_{neg}^{++}(\phi)$ consist of all the formulas in $Sub^{++}(\phi)$ and their negation.*

Finally, let $Con(\phi)$ be the set of maximal **BRSID**-consistent subsets of $Sub^{++}_{neg}(\phi)$. We know that every **BRSID**-consistent subset of $Sub^{++}_{neg}(\phi)$ can be extended to an element of $Con(\phi)$ ([15, p. 199]).

Finally, any S member of $Con(\phi)$ must satisfy the following properties:

- for every $\psi \in Sub^{++}(\phi)$, either ψ or $\neg\psi$ is in S,
- if $\psi \wedge \chi \in S$ then $\psi \in S$ and $\chi \in S$,
- if $\psi \vee \chi \in S$ then $\psi \in S$ or $\chi \in S$,
- if $\psi \in S$ and $\psi \to \chi \in S$ then $\chi \in S$,
- if $\psi \leftrightarrow \chi$ then $\psi \in S$ iff $\chi \in S$,
- if $\psi \in Sub^{++}_{neg}(\phi)$ and **BRSID** $\vdash \psi$ then $\psi \in S$.

Finally, we introduce the notation: $S/B^{\phi}_i = \{\psi | B^{\phi}_i \psi \in S\}$ to refer to the set of formulas believed by agent i when learning that ϕ.

Now we proceed by constructing the canonical model.
We use maximally **BRSID**-consistent sets as building blocks for the canonical model:

Definition 7.17. *Define the following structure:*

$$M_\phi = \langle W, \preccurlyeq, ||\cdot|| \rangle \textit{ where}$$

- $W = \{w_s | S \in Con(\phi)\}$,
- $w_U \preccurlyeq^{w_S}_i w_T$ *if there is some* $\psi \in Sub^{+}(\phi) \cap T \cap U$ *such that* $S/B^{\psi}_i \subseteq T$,
- $||p|| = \{w_S \in W | p \in S\}$

The canonical model M_ϕ has a world w_S that corresponds to every $S \in Con(\phi)$.
The canonical relation tells us that w_T is at least as possible as w_U for agent i at world w_S, if there is some formula $\psi \in Sub^{+}(\phi) \cap T \cap U$ such that the set of the formulas that agent i believes at world w_S when learning that ψ are a subset of T.

Now in order to prove $(\star)$ we will prove the following lemma (truth lemma).

Lemma 7.18. *We have that:*

$$\textit{For every } \psi \in Sub(\phi)\textit{: } (M_\phi, w_S) \vDash \psi \textit{ if and only if } \psi \in S.$$

Now suppose that we have proved this lemma. Then we have that if ϕ is a **BRSID**-consistent formula, then it is contained in some set $S \in Con(\phi)$. Then by the truth lemma above we have that $(M_\phi, w_S) \vDash \phi$. Now if we also establish that $M_\phi \in \mathcal{M}$ then we will get that ϕ is satisfiable w.r.t. $\mathcal{M}$ as required. And this (proving that $M_\phi \in \mathcal{M}$) will be the next important step of the proof. But for now, we will prove the truth lemma.

Proof. As always, we proceed by induction on the structure of formulas.
The Induction Hypothesis is that the truth lemma holds for all subformulas $\psi \in Sub(\phi)$ and we show that it holds for ψ.
We will only deal with the case that ψ is of the form $B^{\chi}_i \zeta$.

($\Leftarrow$)
Assume therefore that $B_i^{\chi}\zeta \in S$. We want to show that $(M_\phi, w_S) \vDash B_i^{\chi}\zeta$.

Consider the set $best_i^{ws}(||\chi|| \cap W_i^{ws})$ and assume that $w_T \in best_i^{ws}(||\chi|| \cap W_i^{ws})$, since otherwise if $best_i^{ws}(||\chi|| \cap W_i^{ws}) = \emptyset$ then $(M_\phi, w_S) \vDash B_i^{\chi}\zeta$.

Since $w_T \in best_i^{ws}(||\chi|| \cap W_i^{ws})$ we get that $w_U \preccurlyeq_i^{ws} w_T$ for all $w_U \in \{||\chi|| \cap W_i^{ws}\}$. Then by the definition of $\preccurlyeq_i^{ws}$ from the canonical model (Definition 7.17), we get that $\exists \xi \in Sub^+(\phi) \cap T$ such that $S/B_i^{\xi} \subseteq T$.

Now we want to show that $\zeta \in T$ to use the Induction Hypothesis and get that $(M_\phi, w_T) \vDash \zeta$. Then, we would have that the most plausible worlds w.r.t. $\preccurlyeq_i^{ws}$ that satisfies χ also satisfies ζ. Hence, we will have that $(M_\phi, w_S) \vDash B_i^{\chi}\zeta$ and establish the right-to-left direction. So lets proceed.

We have that $S/B_i^{\xi} \subseteq T$. This means that S/B_i^{ξ} is a **BRSID**-consistent set, since otherwise T would be inconsistent, but $T \in Con(\phi)$ and $Con(\phi)$ is the set of maximal **BRSID**-consistent subsets of $Sub_{neg}^{++}(\phi)$.

LEMMA 7.19. *S/B_i^{χ} is a* **BRSID**-*consistent set as well.*

PROOF. Suppose not towards a contradiction. Then there is a finite set of formulas $\{\phi_1, \phi_2, ..., \phi_k\} \subseteq S/B_i^{\chi}$ such that

$$\mathbf{BRSID} \vdash \neg(\phi_1 \wedge ... \wedge \phi_k).$$

Now Board shows that ([18, p. 74])

$$\mathbf{BRSID} \vdash (B_i^{\chi}\phi_1 \wedge ... \wedge B_i^{\chi}\phi_k \wedge \neg B_i^{\xi}\neg\chi) \rightarrow (B_i^{\xi}\phi_1 \wedge ... \wedge B_i^{\xi}\phi_k).$$

Since $w_T \in ||\chi||$, i.e. $(M_\phi, w_T) \vDash \chi$, using the induction hypothesis we get that $\chi \in T$. Now $B_i^{\xi}\neg\chi \in Sub^{++}(\phi)$. If $B_i^{\xi}\neg\chi \in S$, then since $S/B_i^{\xi} \subseteq T$, we get that $\neg\chi \in T$. But T is a member of $Con(\phi)$ and hence this contradicts that $\chi \in T$. Therefore $B_i^{\xi}\neg\chi \notin S$ and thus $\neg B_i^{\xi}\neg\chi \in S$ by the properties of the members of $Con(\phi)$.

Claim: $B_i^{\xi}\phi_1, ..., B_i^{\xi}\phi_k \in S$.

PROOF. Since $\{\phi_1, \phi_2, ..., \phi_k\} \subseteq S/B_i^{\chi}$, we have that $B_i^{\chi}\phi_1, ..., B_i^{\chi}\phi_k \in S$ (remember that $S/B_i^{\chi} = \{\psi | B_i^{\chi}\psi \in S\}$).
Moreover, we just argued that $\neg B_i^{\xi}\neg\chi \in S$.
Hence we get that $B_i^{\chi}\phi_1 \wedge ... \wedge B_i^{\chi}\phi_k \wedge \neg B_i^{\xi}\neg\chi \in S$.

Then since

$$\mathbf{BRSID} \vdash (B_i^{\chi}\phi_1 \wedge ... \wedge B_i^{\chi}\phi_k \wedge \neg B_i^{\xi}\neg\chi) \rightarrow (B_i^{\xi}\phi_1 \wedge ... \wedge B_i^{\xi}\phi_k)$$

holds, using the properties of the members of $Con(\phi)$, we get that

$$(B_i^{\xi}\phi_1 \wedge ... \wedge B_i^{\xi}\phi_k) \in S.$$

But this entails that $B_i^{\xi}\phi_1, ..., B_i^{\xi}\phi_k \in S$, once again by the properties of the members of $Con(\phi)$. ☐

Therefore $B_i^{\xi}\phi_1, ..., B_i^{\xi}\phi_k \in S$. Hence $\{\phi_1, \phi_2, ..., \phi_k\} \subseteq S/B_i^{\xi}$, therefore S/B_i^{ξ} is not a **BRSID**-consistent set. Contradiction. ☐

Now that we have established that S/B_i^χ is a **BRSID**-consistent set, consider its maximal extension U.

By (**Succ**) we get that $B_i^\chi \chi \in S$. Hence $\chi \in (S/B_i^\chi) \subseteq U$. Therefore $w_T \preccurlyeq_i^{ws} w_U$.

On the other hand, we have that $w_T \in best_i^{ws}(||\chi|| \cap W_i^{ws})$. Hence $w_U \preccurlyeq_i^{ws} w_T$. Thus $\exists \lambda \in Sub^+(\phi) \cap T \cap U$ such that $S/B_i^\lambda \subseteq T$ (Definition 7.17).

Now $\lambda \in U$ and $S/B_i^\chi \subseteq U \Leftrightarrow \{\psi | B_i^\chi \psi \in S\} \subseteq U$. Now if $B_i^\chi \neg\lambda \in S$, then $\neg\lambda \in S/B_i^\chi$, thus $\neg\lambda \in U$. Contradiction. Therefore $\neg B_i^\chi \neg\lambda \in S$.

Moreover, $\chi \in T$ and $S/B_i^\lambda \subseteq T$. Hence with the same argument $\neg B_i^\lambda \neg\chi \in S$.

Furthermore, Board establishes that ([18, p. 75])

$$\mathbf{BRSID} \vdash (\neg B_i^\chi \neg\lambda \wedge B_i^\chi \zeta \wedge \neg B_i^\lambda \neg\chi) \to (B_i^\lambda \zeta \vee B_i^\lambda(\chi \to \zeta)).$$

Now since $\chi, \zeta, \lambda \in Sub^+(\phi)$, by the definition of $Sub^{++}(\phi)$, we have that $B_i^\lambda \zeta$, $B_i^\lambda(\chi \to \zeta) \in Sub^{++}(\phi)$.

Moreover, we have assumed that $B_i^\chi \zeta \in S$.

Therefore, we get that $(B_i^\lambda \zeta \vee B_i^\lambda(\chi \to \zeta)) \in S$.

Now we also have that $S/B_i^\lambda \subseteq T$. Hence we get that $\zeta \in T$ or that $(\chi \to \zeta) \in T$. However, if $(\chi \to \zeta) \in T$, then combined with $\chi \in T$ we derive that $\zeta \in T$.

Therefore $\forall w_T \in best_i^{ws}(||\chi|| \cap W_i^{ws})$ we have $\zeta \in T$ and using the induction hypothesis we get that $(M_\phi, w_T) \vDash \zeta$, which in turn entails that $(M_\phi, w_S) \vDash B_i^\chi \zeta$.

This ends the $(\Leftarrow)$ direction of the truth lemma.

$(\Rightarrow)$
Assume that $(M_\phi, w_S) \vDash B_i^\chi \zeta$.

Claim: The set $(S/B_i^\chi) \cup \{\neg\zeta\}$ is not **BRSID**-consistent.

Proof. Suppose otherwise towards a contradiction. Then it would have a maximal **BRSID**-consistent extension T. By (**Succ**) we get that $B_i^\chi \chi \in S$. Therefore $\chi \in (S/B_i^\chi) \subseteq T$.

Hence this means that $w_U \preccurlyeq_i^{ws} w_T$ for all U such that $\chi \in U$, thus $w_T \in best_i^{ws}(||\chi|| \cap W_i^{ws})$.
Now $\neg\zeta \in T$ (since T is a maximal **BRSID**-consistent extension of $(S/B_i^\chi) \cup \{\neg\zeta\}$) and using the induction hypothesis, we get that $(M_\phi, w_T) \vDash \neg\zeta$. Therefore, we have a world w_T in $best_i^{ws}(||\chi|| \cap W_i^{ws})$ such that $(M_\phi, w_T) \vDash \neg\zeta$.

This entails that $(M_\phi, w_T) \vDash \neg B_i^\chi \zeta$. Contradiction. □

Now since $(S/B_i^\chi) \cup \{\neg\zeta\}$ is not **BRSID**-consistent, there must be some finite subset $\{\phi_1, ..., \phi_k, \neg\zeta\}$ that is not **BRSID**-consistent.

Now

$$\mathbf{BRSID} \vdash \phi_1 \to (\phi_2 \to (... \to (\phi_k \to \zeta)...)).$$

By applying (**RE**), (**Distr**), (**MP**) we derive that: ([23])

$$\mathbf{BRSID} \vdash B_i^\chi \phi_1 \to (B_i^\chi \phi_2 \to (... \to (B_i^\chi \phi_k \to B_i^\chi \zeta)...))$$

and therefore

$$(B_i^\chi \phi_1 \to (B_i^\chi \phi_2 \to (... \to (B_i^\chi \phi_k \to B_i^\chi \zeta)...))) \in S.$$

Now $\phi_1, \phi_2, ..., \phi_k \in (S/B_i^\chi)$, thus $B_i^\chi \phi_1, ..., B_i^\chi \phi_k \in S$ (since $S/B_i^\chi = \{\psi | B_i^\chi \psi \in S\}$). Therefore $B_i^\chi \zeta \in S$.

And this ends the proof of the ($\Rightarrow$) direction and the proof of the truth lemma. $\square$

Hence, we have established that if ϕ is a **BRSID**-consistent formula contained in some set $S \in Con(\phi)$, then $(M_\phi, w_S) \vDash \phi$.

The next and final step of the proof is to show that $M_\phi \in \mathcal{M}$. That way, we will have that every **BRSID**-consistent formula is satisfiable w.r.t. $\mathcal{M}$, thus we will prove ($\star$) and in turn completeness w.r.t. $\mathcal{M}$.

Proposition 7.20. $M_\phi \in \mathcal{M}$

Proof. We will show that $\preccurlyeq_i^{w_S}$ is complete and transitive on $W_i^{w_S}$, well-founded, absolute and that $W_i^{w_S} \neq \emptyset$.[9]

So let us begin.

- $\preccurlyeq_i^{w_S}$ is complete on $W_i^{w_S}$.

  Proof. Pick $w_T, w_U \in W_i^{w_S}$.
  We w.t.s. that either $w_T \preccurlyeq_i^{w_S} w_U$ or $w_U \preccurlyeq_i^{w_S} w_T$.
  From the definition of W_i^w and $\preccurlyeq_i^{w_S}$, we have that there is some $\psi \in Sub^+(\phi) \cap T$, such that $S/B_i^\psi \subseteq T$ and some $\chi \in Sub^+(\phi) \cap U$ such that $S/B_i^\chi \subseteq U$.
  Now S is a maximal **BRSID**-consistent set.
  This entails that either $B_i^{\psi\vee\chi} \neg\psi \in S$, or that $\neg B_i^{\phi\vee\chi} \neg\psi \in S$.
  - Case 1. $B_i^{\psi\vee\chi} \neg\psi \in S$.
    By (**Succ**)
    $$B_i^{\psi\vee\chi}(\psi \vee \chi) \in S. \ (\mathbf{1})$$
    By (**Distr**)
    $$((B_i^{\psi\vee\chi} \neg\psi \wedge B_i^{\psi\vee\chi}(\psi \vee \chi)) \to B_i^{\psi\vee\chi} \chi) \in S. \ (\mathbf{2})$$
    By (**IE(a)**)
    $$(B_i^{\psi\vee\chi} \chi \to (B_i^{(\psi\vee\chi)\wedge\chi} \zeta \leftrightarrow B_i^{\psi\vee\chi} \zeta)) \in S. \ (\mathbf{3})$$
    By (**Taut**)
    $$(((\psi \vee \chi) \wedge \chi) \leftrightarrow \chi) \in S. \ (\mathbf{4})$$
    Therefore by (**4**) and (**LE**)
    $$B_i^{(\psi\vee\chi)\wedge\chi} \zeta \leftrightarrow B_i^\chi \zeta \in S. \ (\mathbf{5})$$
    Now by (**2**), (**1**) and the fact that $B_i^{\psi\vee\chi} \neg\psi \in S$, we get that $B_i^{\psi\vee\chi} \chi \in S$.

[9] This is our addition: **R4**.

Hence (**3**) gives us that $B_i^{(\psi\vee\chi)\wedge\chi}\zeta \in S$ if and only if $B_i^{\psi\vee\chi}\zeta \in S$, i.e.: $S/B_i^{(\psi\vee\chi)\wedge\chi} = S/B_i^{\psi\vee\chi}$.
Now (**5**) tells us that $S/B_i^{(\psi\vee\chi)\wedge\chi} = S/B_i^{\chi}$.
Combining the last two sentences we get that $S/B_i^{\chi} = S/B_i^{\psi\vee\chi} \subseteq U$ (since we argued above that $S/B_i^{\chi} \subseteq U$).
However $\psi \vee \chi \in T \cap U$, since $\psi \in T$ and $\chi \in U$, thus $w_T \preccurlyeq_i^{ws} w_U$.

- Case 2. $\neg B_i^{\psi\vee\chi}\neg\psi \in S$.
By (**IE(b)**)
$$(\neg B_i^{\psi\vee\chi}\neg\psi \to (B_i^{(\psi\vee\chi)\wedge\psi}\zeta \leftrightarrow B_i^{\psi\vee\chi}(\chi \to \zeta))) \in S. \ (\mathbf{6})$$
By (**RE**), (**Dist**), (**Taut**), (**MP**)
$$(B_i^{\psi\vee\chi}\zeta \to B_i^{\psi\vee\chi}(\chi \to \zeta)) \in S. \ (\mathbf{7})$$
Now (**7**) entails that if $B_i^{\psi\vee\chi}\zeta \in S$, then $B_i^{\psi\vee\chi}(\chi \to \zeta) \in S$.
And in that case since $\neg B_i^{\psi\vee\chi}\neg\psi \in S$, (**6**) gives us that $B_i^{(\psi\vee\chi)\wedge\psi}\zeta \in S$.
Thus $S/B_i^{\psi\vee\chi} \subseteq S/B_i^{(\psi\vee\chi)\wedge\psi} = S/B_i^{\psi} \subseteq T$.
But $\psi \vee \chi \in T \cap U$ as before and thus $w_U \preccurlyeq_i^{ws} w_T$.

Hence completeness of $\preccurlyeq_i^{ws}$ on W_i^{ws} has been established. □

- $\preccurlyeq_i^{ws}$ is transitive on W_i^{ws}

Proof. Let $w_T, w_U, w_V \in W_i^{ws}$ such that $w_V \preccurlyeq_i^{ws} w_U$ and $w_U \preccurlyeq_i^{ws} w_T$. We want to show that $w_V \preccurlyeq_i^{ws} w_T$.
By the definition of $\preccurlyeq_i^{ws}$, we have that $\exists\chi \in Sub^+(\phi)\cap U \cap V$ such that $S/B_i^{\chi} \subseteq U$ and $\exists\psi \in Sub^+(\phi) \cap T \cap U$ such that $S/B_i^{\psi} \subseteq T$.

Now S is a maximal **BRSID**-consistent set.
Hence either $B_i^{\psi\vee\chi}\neg\psi \in S$ or $\neg B_i^{\psi\vee\chi}\neg\psi \in S$.

- Case 1. $B_i^{\psi\vee\chi}\neg\psi \in S$.
By the first part of the proof that $\preccurlyeq_i^{ws}$ is complete, we have that $S/B_i^{\chi} = S/B_i^{\psi\vee\chi} \subseteq U$.
But $\psi \in U$ contradicting that $B_i^{\psi\vee\chi}\neg\psi \in S$.
- Case 2. $\neg B_i^{\psi\vee\chi}\neg\psi \in S$.
By the second part of the proof that $\preccurlyeq_i^{ws}$ is complete, we have that $S/B_i^{\psi\vee\chi} \subseteq S/B_i^{\psi} \subseteq T$.
But $\psi \vee \chi \in T \cap U \cap V$, therefore $w_V \preccurlyeq_i^{ws} w_T$.

□

- $\preccurlyeq_i^{ws}$ is well-founded.

Proof. We will show that W is finite. We will do so, by showing that $Con(\phi)$ is finite.

Now $Sub(\phi)$ is a finite set therefore it has a finite number of maximal **BRSID**-consistent subsets.

Claim: Each maximal **BRSID**-consistent subset of $Sub(\phi)$, has a unique extension to a maximal **BRSID**-consistent subset of $Sub^+(\phi)$.

Proof. Let S be a maximal **BRSID**-consistent subset of $Sub(\phi)$ and S^+ a maximal **BRSID**-consistent subset of $Sub^+(\phi)$ such that $S \subseteq S^+$. Then $\psi \in S^+$ only if either $\psi \in S$, or ψ is of the form $\neg\chi$ and $\chi \notin S^+$, or ψ is of the form $\chi \wedge \zeta$ and $\chi, \zeta \in S^+$.

Now if there is no maximal **BRSID**-consistent subset S of $Sub(\phi)$ such that $S \subseteq S^+$, S^+ can not be a maximal **BRSID**-consistent subset of $Sub^+(\phi)$. Hence, there is a one-to-one correspondence between the maximal **BRSID**-consistent subsets of $Sub(\phi)$ and the maximal **BRSID**-consistent subsets of $Sub^+(\phi)$. $\square$

Now there are at most $2^{||Sub(\phi)||}$ logically distinct formulas in $Sub^+(\phi)$. If **BRSID** $\vdash \psi \leftrightarrow \chi$ and S^{++}_{neg} is a maximal **BRSID**-consistent subset of $Sub^{++}_{neg}(\phi)$, then $B_i^{\psi}\zeta \in S^{++}_{neg}$ iff $B_i^{\chi}\zeta \in S^{++}_{neg}$. Moreover, for every formula of the form $B_i^{\zeta} B_i^{\xi} \cdots B_i^{\chi}\psi \in S^{++}_{neg}$ iff $B_i^{\xi} \cdots B_i^{\chi}\psi \in S^{++}_{neg}$, by (**PI**) and (**NI**).

Hence, each **BRSID**-consistent subset of $Sub^+(\phi)$ has a finite number of extensions to maximal **BRSID**-consistent subsets of $Sub^{++}_{neg}(\phi)$. Hence $Con(\phi)$ is a finite set. Thus W is finite and well-foundedness follows. $\square$

- $\preccurlyeq_i^{w_S}$ is absolute.

Proof. Take $w_T \in W_i^{w_S}$. Then $\exists\psi \in Sub^+(\phi) \cap T$ such that $S/B_i^{\psi} \subseteq T$. We must show that $w_V \preccurlyeq_i^{w_S} w_U$ iff $w_V \preccurlyeq_i^{w_T} w_U$.

($\Rightarrow$)
Assume $w_V \preccurlyeq_i^{w_S} w_U$. Therefore we get that $\exists\chi \in Sub^+(\phi) \cap U \cap V$ such that $S/B_i^{\chi} \subseteq U$.

Claim: $T/B_i^{\chi} \subseteq S/B_i^{\chi}$

Proof. Suppose that $B_i^{\chi}\zeta \notin S$. Then $\neg B_i^{\chi}\zeta \in S$.
By **(NI)**

$$(\neg B_i^{\chi}\zeta \to B_i^{\psi}\neg B_i^{\chi}\zeta) \in S.$$

Hence $B_i^{\psi}\neg B_i^{\chi}\zeta \in S$. But $S/B_i^{\psi} \subseteq T$. Hence $\neg B_i^{\chi}\zeta \in T$. Therefore $B_i^{\chi}\zeta \notin T$. Hence indeed $T/B_i^{\chi} \subseteq S/B_i^{\chi} \subseteq U$. $\square$

Thus $w_V \preccurlyeq_i^{w_T} w_U$.

($\Leftarrow$)
Assume $w_V \preccurlyeq_i^{w_T} w_U$. Then there is some $\chi \in Sub^+(\phi) \cap U \cap V$ such that $T/B_i^{\chi} \subseteq U$.

Claim: $S/B_i^{\chi} \subseteq T/B_i^{\chi}$

PROOF. Suppose that $B_i^{\chi}\zeta \in S$.
By (**PI**)

$$(B_i^{\chi}\zeta \rightarrow B_i^{\psi}B_i^{\chi}\zeta) \in S.$$

Hence $B_i^{\psi}B_i^{\chi}\zeta \in S$. But $S/B_i^{\psi} \subseteq T$, thus $B_i^{\chi}\zeta \in T$. Therefore $S/B_i^{\chi} \subseteq T/B_i^{\chi} \subseteq U$. □

Thus $w_V \preccurlyeq_i^{w_S} w_U$. □

- For all $w_S \in M_\phi$ we have that: $W_i^{w_S} \neq \emptyset$.

PROOF. We begin with the following claim:

Claim: The set $S/B_i^{\top}$ is **BRSID**-consistent.

PROOF. Assume not towards a contradiction. Then $\exists\{\phi_1, ...\phi_k\} \subseteq S/B_i^{\top}$ such that

$$\textbf{BRSID} \vdash ((\phi_1 \wedge ... \wedge \phi_k) \rightarrow \bot).\ (\mathbf{1'})$$

Now since $\phi_1, ..., \phi_k \in S/B_i^{\top}$, we get that $B_i^{\top}\phi_1, ..., B_i^{\top}\phi_k \in S$. Therefore

$$(B_i^{\top}\phi_1 \wedge ... \wedge B_i^{\top}\phi_k) \in S.\ (\mathbf{2'})$$

We also have that

$$\textbf{BRSID} \vdash ((B_i^{\top}\phi_1 \wedge ... \wedge B_i^{\top}\phi_k) \rightarrow B_i^{\top}(\phi_1 \wedge ... \wedge \phi_k)).\ (\mathbf{3'})$$

By ($\mathbf{2'}$), ($\mathbf{3'}$) and the fact that S is a member of $Con(\phi)$ we get that

$$B_i^{\top}(\phi_1 \wedge ... \wedge \phi_k) \in S.\ (\mathbf{4'})$$

By ($\mathbf{1'}$) and **(RE)**

$$(B_i^{\top}((\phi_1 \wedge ... \wedge \phi_k) \rightarrow \bot)) \in S.\ (\mathbf{5'})$$

By **(Distr)**

$$(B_i^{\top}(\phi_1 \wedge ... \wedge \phi_k) \wedge B_i^{\top}((\phi_1 \wedge ... \wedge \phi_k) \rightarrow \bot) \rightarrow B_i^{\top}\bot) \in S.\ (\mathbf{6'})$$

By ($\mathbf{4'}$), ($\mathbf{5'}$), ($\mathbf{6'}$)

$$B_i^{\top}\bot \in S.\ (\mathbf{7'})$$

But this (($\mathbf{7'}$)) contradicts axiom **(D)**.
Hence we have proved the claim. □

Therefore, the set $S/B_i^{\top}$ can be extended to a maximal **BRSID**-consistent subset U of $Sub_{neg}^{++}(\phi)$.
Now the definition of $\preccurlyeq_i^{w_S}$ from the canonical model entails that $w_U \preccurlyeq_i^{w_S} w_U$ and hence $W_i^{w_S} \neq \emptyset$. □

Hence we have shown that $W_i^{w_S} \neq \emptyset$.

□

Hence, we have established that $(M_\phi, w_S) \vDash \psi$ iff $\psi \in S$ and that $M_\phi \in \mathcal{M}$.
Now we go back to what we set out to prove:

($\star$) Every **BRSID**-consistent formula in $\mathcal{L}$ is satisfiable w.r.t. $\mathcal{M}$

Suppose now that ϕ is **BRSID**-consistent. Then it is contained in some set $S \in Con(\phi)$. Then it follows from our truth lemma that $(M_\phi, w_S) \vDash \phi$. Since we proved that $M_\phi \in \mathcal{M}$ we get that ϕ is satisfiable w.r.t. $\mathcal{M}$. Therefore this proves $(\star)$ and we have shown that **BRSID** is a complete axiomatization w.r.t. $\mathcal{M}$.

□

Now we move on to the next step.

The completeness proof w.r.t. our probabilistic semantics, will be a combination of the previous (7.9) and the following (7.21) proposition.

PROPOSITION 7.21. *For every finite belief revision structure $M = \langle W, \preccurlyeq, ||\cdot||\rangle$ satisfying* **R1**, **R2**, **R3**, **R4**, *there exists a finite (non-trivial) probabilistic model M', having the same set of worlds and same valuation as M and such that the same sentences of our language are satisfied at the same worlds in the two models, i.e.:*

$$\forall w \in W : (M, w) \vDash \phi \textit{ iff } (M', w) \vDash \phi.$$

PROOF. We begin by fixing some threshold r such that $r \in (\frac{1}{2}, 1)$.
We will prove that our logic is complete for non-trivial probabilistic models in which all agents' thresholds are equal to r. This in turn implies completeness for arbitrary non-trivial probabilistic models.

Suppose now that $M = \langle W, \preccurlyeq, ||\cdot||\rangle$ is a finite belief revision structure, satisfying **R1**, **R2**, **R3**, **R4**.
For $w, v \in W$, we put:

$$w \sim_i v \text{ iff } W_i^w = W_i^v.$$

This gives us the equivalence classes $w(i)$ and the partitions Π_i.

Now define the following family of subsets of $w(i)$:

$$\mathcal{G}^{w,i} := \{S \subseteq w(i) | S \neq \emptyset \text{ and } \forall x \in S \forall y (x \preccurlyeq_i^w y \Rightarrow y \in S)\}.^{10}$$

LEMMA 7.22. *The family $\mathcal{G}^{w,i}$ is nested.*

PROOF. Pick $S', S'' \in \mathcal{G}^{w,i}$.
Assume towards a contradiction that $S' \nsubseteq S''$ and that $S'' \nsubseteq S'$.
Now pick $x \in S' - S''$ and $y \in S'' - S'$.
Now since $\preccurlyeq_i^w$ is complete on $w(i)$, we get that either $x \preccurlyeq_i^w y$ or $y \preccurlyeq_i^w x$.
If $x \preccurlyeq_i^w y$ then since $x \in S'$ we get that $y \in S'$, by the definition of $\mathcal{G}^{w,i}$ which is a contradiction.
If $y \preccurlyeq_i^w x$ we get a contradiction with an analogous argument. □

[10] $\mathcal{G}^{w,i}$ depends on both i and w

Therefore, we have that $\mathcal{G}^{w,i}$ is a family of nested and well-founded spheres that are upwards-closed w.r.t. the plausibility relation $\preccurlyeq_i^w$.

Consider now $S_0^{w,i} \subset S_1^{w,i} \subset ... \subset S_{n(w,i)}^{w,i}$ [11] to be an enumeration of all those spheres in $\mathcal{G}^{w,i}$.

Definition 7.23. *Define* $N_0^{w,i} := |S_0^{w,i}|$ *to be the* cardinality *of the sphere* $S_0^{w,i}$ *and define* $N_j^{w,i} := |S_j^{w,i}| - |S_{j-1}^{w,i}|, \forall j = 1, ..., n(w,i)$*.*

Now, we will proceed with the following steps:

Step 1 We will define the numbers $x_j^{w,i}$ for $j = 0, ..., n(w,i)$ that we will later assign to all the elements of the spheres in $\mathcal{G}^{w,i}$. In particular, we will later assign (when we will define our probability function) the same probability $x_0^{w,i}$ to all the worlds in $S_0^{w,i}$ and similarly the same probability $x_j^{w,i}$ to all the worlds in $S_j^{w,i} - S_{j-1}^{w,i}$ for all $j = 1, ..., n(w,i)$. This is the key idea behind the construction of our probabilistic model.

Step 2 We will show that $x_j^{w,i} \geq \frac{r}{1-r}(1 - \sum_{0 \leq m \leq j} N_m^{w,i} x_m^{w,i})$, for all $j \in \{0, ..., n(w,i) - 1\}$.

This combined with Corollary 4.15 will later entail that all $S_j^{w,i}$ in $\mathcal{G}^{w,i}$ are r-stable sets w.r.t. the probability function we will define.

Step 3 We will define the probability function P_i^w.

Step 4 We will show (using Step 1) that every set $S' \supseteq S_{n(w,i)}^{w,i}$ is an a priori set w.r.t. P_i^w (with $S_{n(w,i)}^{w,i} \subseteq w(i)$ the largest sphere in $\mathcal{G}^{w,i}$). And of course, since $S_{n(w,i)}^{w,i} \subseteq w(i)$, we will have that $w(i)$ and all $A \supset w(i)$ will be a priori sets w.r.t. P_i^w as well. Moreover, all $A \subseteq W - w(i)$ will be abnormal w.r.t. P_i^w.

Step 5 We will prove that the family of r-stable sets w.r.t. P_i^w is the union of $\mathcal{G}^{w,i}$ and of the a priori sets. This will be done by first using Step 2 and Corollary 4.15 to show that all the spheres in $\mathcal{G}^{w,i}$ are r-stable.

Step 6 We will define our non-trivial probabilistic model M'.

Step 7 Finally, we will prove the truth preserving lemma:

$$\forall w \in W : (M, w) \vDash \phi \text{ iff } (M', w) \vDash \phi.$$

So let's begin.

Step 1

Define the following numbers:

$$k = \frac{r}{1-r}.$$

And:

- For $j = 0, ..., n(w,i) - 1$, we put:

$$x_j^{w,i} = \frac{k}{\Pi_{0 \leq m \leq j}(1 + kN_m^{w,i})} \text{ for every agent } i.$$

[11]Notice that n depends on w, i, since we may have a different number of spheres for different agents and different worlds w. This is why we use $n(w,i)$.

- For $j = n(w,i)$, we put:

$$x^{w,i}_{n(w,i)} = \frac{1 - \sum\limits_{j=0}^{n(w,i)-1} N^{w,i}_j x^{w,i}_j}{N^{w,i}_{n(w,i)}} \text{ for every agent } i.$$

Now we will prove the following lemmas:

LEMMA 7.24. *For all $j = 0, ..., n(w,i) - 1$ we have that $x^{w,i}_j \in [0,1]$*

PROOF.

$$\begin{aligned}
x^{w,i}_j &= \frac{k}{\Pi_{0 \le m \le j}(1 + kN^{w,i}_m)} \\
&= \frac{k}{(1 + kN^{w,i}_0)(1 + kN^{w,i}_1) \cdots (1 + kN^{w,i}_j)} \\
&= \frac{k}{k(\frac{1}{k} + N^{w,i}_0)(1 + kN^{w,i}_1) \cdots (1 + kN^{w,i}_j)} \\
&= \frac{1}{(\frac{1}{k} + N^{w,i}_0)(1 + kN^{w,i}_1) \cdots (1 + kN^{w,i}_j)}
\end{aligned}$$

Now since $r \in (\frac{1}{2}, 1)$, we have that $k \in (1, \infty)$.
Therefore $x^{w,i}_j \in (0,1)$, for all $j = 0, ..., n(w,i) - 1$. □

LEMMA 7.25. $\sum\limits_{j=0}^{n(w,i)} N^{w,i}_j x^{w,i}_j = 1$

PROOF.

$$\begin{aligned}
\sum_{j=0}^{n(w,i)} N^{w,i}_j x^{w,i}_j &= \sum_{j=0}^{n(w,i)-1} N^{w,i}_j x^{w,i}_j + N^{w,i}_{n(w,i)} x^{w,i}_{n(w,i)} \\
&= \sum_{j=0}^{n(w,i)-1} N^{w,i}_j x^{w,i}_j + 1 - \sum_{j=0}^{n(w,i)-1} N^{w,i}_j x^{w,i}_j \\
&= 1
\end{aligned}$$

□

These numbers that we defined here will be the probabilities of the elements of the spheres of $\mathcal{G}^{w,i}$. In Step 3, we will define our probability function in a way that all the worlds in $S^{w,i}_0$ have the same probability $x^{w,i}_0$ and similarly for all $j = 1, ..., n(w,i)$ that all the worlds in $S^{w,i}_j - S^{w,i}_{j-1}$ have the same probability $x^{w,i}_j$.

Step 2
We will now prove the following Lemma that will be used later to establish that the sets in $\mathcal{G}^{w,i}$ are r-stable.

LEMMA 7.26. *For all* $j \in \{0, ..., n(w,i)-1\}$*:*

$$x_j^{w,i} \geq \frac{r}{1-r}(1 - \sum_{0 \leq m \leq j} N_m^{w,i} x_m^{w,i}).$$

PROOF. By Induction.
Base cases:
For $j = 0$:

$$\begin{aligned}
x_0^{w,i} &= \frac{k}{1 + kN_0^{w,i}} \\
&= \frac{\frac{r}{1-r}}{1 + \frac{r}{1-r}N_0^{w,i}} \\
&= \frac{r}{1 - r + rN_0^{w,i}} \\
&= \frac{r}{1-r}(1 - N_0^{w,i}x_0^{w,i})
\end{aligned}$$

Therefore indeed we have that: $x_0^{w,i} \geq \frac{r}{1-r}(1 - N_0^{w,i}x_0^{w,i})$.

Now we will do one more case, for $j = 1$:

$$\begin{aligned}
x_1^{w,i} &= \frac{k}{(1 + kN_0^{w,i})(1 + kN_1^{w,i})} \\
&= \frac{\frac{k(1+kN_0^{w,i}) - k^2N_0^{w,i}}{1+kN_0^{w,i}}}{1 + kN_1^{w,i}} \\
&= \frac{k - \frac{k^2N_0^{w,i}}{1+kN_0^{w,i}}}{1 + kN_1^{w,i}} \\
&= \frac{\frac{r}{1-r} - \frac{rN_0^{w,i}k}{(1-r)(1+kN_0^{w,i})}}{1 + \frac{rN_1^{w,i}}{1-r}} \\
&= \frac{r - rN_0^{w,i}\frac{k}{1+kN_0^{w,i}}}{1 - r + rN_1^{w,i}} \\
&= \frac{r - rN_0^{w,i}x_0^{w,i}}{1 - r + rN_1^{w,i}} \\
(1 - r + rN_1^{w,i})x_1^{w,i} &= r - rN_0^{w,i}x_0^{w,i} \\
(1-r)x_1^{w,i} + x_1^{w,i}rN_1^{w,i} &= r - rN_0^{w,i}x_0^{w,i} \\
x_1^{w,i} &= \frac{r}{1-r}(1 - N_0^{w,i}x_0^{w,i} - N_1^{w,i}x_1^{w,i})
\end{aligned}$$

Therefore indeed $x_1^{w,i} \geq \frac{r}{1-r}(1 - N_0^{w,i}x_0^{w,i} - N_1^{w,i}x_1^{w,i})$.

Induction Hypothesis: For $j = 1, ..., n(w,i) - 2$:
$x_j^{w,i} = \frac{r}{1-r}(1 - N_0^{w,i}x_0^{w,i} - N_1^{w,i}x_1^{w,i} - \cdots - N_j^{w,i}x_j^{w,i})$.

Inductive Step: We will show that:

$x_{j+1}^{w,i} = \frac{r}{1-r}(1 - N_0^{w,i}x_0^{w,i} - N_1^{w,i}x_1^{w,i} - \cdots - N_j^{w,i}x_{j+1}^{w,i})$.
We have that:

$$x_{j+1}^{w,i} = \frac{k}{(1+kN_0^{w,i})(1+kN_1^{w,i})\cdots(1+kN_{j+1}^{w,i})}$$

$$x_{j+1}^{w,i} = \frac{\frac{k(1+kN_0^{w,i})(1+kN_1^{w,i})\cdots(1+kN_j^{w,i}) - k^2N_0^{w,i}(1+kN_1^{w,i})(1+kN_2^{w,i})\cdots(1+kN_j^{w,i}) - \cdots - k^2N_j^{w,i}}{(1+kN_0^{w,i})(1+kN_1^{w,i})\cdots(1+kN_j^{w,i})}}{1+kN_{j+1}^{w,i}} \text{ [12]}$$

$$x_{j+1}^{w,i} = \frac{k - \frac{k^2N_0^{w,i}}{1+kN_0^{w,i}} - \frac{k^2N_1^{w,i}}{(1+kN_0^{w,i})(1+kN_1^{w,i})} - \cdots - \frac{k^2N_j^{w,i}}{(1+kN_0^{w,i})(1+kN_1^{w,i})\cdots(1+kN_j^{w,i})}}{1+kN_{j+1}^{w,i}}$$

$$x_{j+1}^{w,i} = \frac{k - kN_0^{w,i}x_0^{w,i} - kN_1^{w,i}x_1^{w,i} - \cdots kN_j^{w,i}x_j^{w,i}}{1+kN_{j+1}^{w,i}}$$

$$x_{j+1}^{w,i}(1 - r + rN_{j+1}^{w,i}) = r - rN_0^{w,i}x_0^{w,i} \cdots - rN_j^{w,i}x_j^{w,i}$$

$$x_{j+1}^{w,i} = \frac{r}{r-1}(1 - N_0^{w,i}x_0^{w,i} - N_1^{w,i}x_1^{w,i} - N_2^{w,i}x_2^{w,i} - \cdots - N_j^{w,i}x_j^{w,i} - N_{j+1}^{w,i}x_{j+1}^{w,i})$$

Therefore indeed
$x_{j+1}^{w,i} \geq \frac{r}{r-1}(1 - N_0^{w,i}x_0^{w,i} - N_1^{w,i}x_1^{w,i} - N_2^{w,i}x_2^{w,i} - \cdots - N_j^{w,i}x_j^{w,i} - N_{j+1}^{w,i}x_{j+1}^{w,i})$.

□

Step 3

Now we are going to define the probability function P_i^w.
First define the function $\mu_i^w : W \to [0,1]$ such that:

$$\mu_i^w(v) := \begin{cases} x_0^{w,i} & \text{if } v \in S_0^{w,i} \\ x_j^{w,i} & \text{if } v \in S_j^{w,i} - S_{j-1}^{w,i}, \text{ with } j \geq 1 \\ 0 & \text{else} \end{cases}$$

And now define the classical probability function $P_i^w : \mathcal{F} \to [0,1]$:

$$P_i^w(A) := \sum\{\mu_i^w(v) | v \in A\}$$

[12] Here we use that:
$k(1+kN_0^{w,i})(1+kN_1^{w,i})\cdots(1+kN_j^{w,i}) - k^2N_0^{w,i}(1+kN_1^{w,i})(1+kN_2^{w,i})\cdots(1+kN_j^{w,i}) - k^2N_1^{w,i}(1+kN_2^{w,i})(1+kN_3^{w,i})\cdots(1+kN_j^{w,i}) - \cdots - k^2N_j^{w,i} =$
$(k+k^2N_0^{w,i})(1+kN_1^{w,i})\cdots(1+kN_j^{w,i}) - k^2N_0^{w,i}(1+kN_1^{w,i})(1+kN_2^{w,i})\cdots(1+kN_j^{w,i}) - k^2N_1^{w,i}(1+kN_2^{w,i})(1+kN_3^{w,i})\cdots(1+kN_j^{w,i}) - \cdots - k^2N_j^{w,i}$
$k(1+kN_1^{w,i})\cdots(1+kN_j^{w,i}) + k^2N_0^{w,i}(1+kN_1^{w,i})\cdots(1+kN_j^{w,i}) - k^2N_0^{w,i}(1+kN_1^{w,i})(1+kN_2^{w,i})\cdots(1+kN_j^{w,i}) - k^2N_1^{w,i}(1+kN_2^{w,i})(1+kN_3^{w,i})\cdots(1+kN_j^{w,i}) - \cdots - k^2N_j^{w,i} =$
$k(1+kN_1^{w,i})\cdots(1+kN_j^{w,i}) - k^2N_1^{w,i}(1+kN_2^{w,i})(1+kN_3^{w,i})\cdots(1+kN_j^{w,i}) - \cdots - k^2N_j^{w,i} =$
$(k+k^2N_1^{w,i})(1+kN_2^{w,i})(1+kN_3^{w,i})\cdots(1+kN_j^{w,i}) - k^2N_1^{w,i}(1+kN_2^{w,i})(1+kN_3^{w,i})\cdots(1+kN_j^{w,i}) - \cdots - k^2N_j^{w,i} =$
$k(1+kN_2^{w,i})(1+kN_3^{w,i})\cdots(1+kN_j^{w,i}) + k^2N_1^{w,i}(1+kN_2^{w,i})(1+kN_3^{w,i})\cdots(1+kN_j^{w,i}) - k^2N_1^{w,i}(1+kN_2^{w,i})(1+kN_3^{w,i})\cdots(1+kN_j^{w,i}) - \cdots - k^2N_j^{w,i} =$
$k(1+kN_2^{w,i})(1+kN_3^{w,i})\cdots(1+kN_j^{w,i}) - k^2N_2^{w,i}(1+kN_3^{w,i})\cdots(1+kN_j^{w,i}) - \cdots - k^2N_j^{w,i} = k$

Now for $A, B \in \mathcal{F}$, conditional probabilities will be given by:

- $P_i^w(A|B) = \frac{P_i^w(A\cap B)}{P_i^w(B)}$, if $P_i^w(B) > 0$ and
- $P_i^w(A|B) = 1$ if $P_i^w(B) = 0$.

Step 4

We will show (using Step 1) that every set $S' \supseteq S^{w,i}_{n(w,i)}$ is an a priori set w.r.t. P_i^w (with $S^{w,i}_{n(w,i)} \subseteq w(i)$ the largest sphere in $\mathcal{G}^{w,i}$).

LEMMA 7.27. *Every $X \subseteq W$ such that $X \supseteq S^{w,i}_{n(w,i)}$ is a priori w.r.t. P_i^w.*

PROOF. Pick $X \subseteq W$ such that $X \supseteq S^{w,i}_{n(w,i)}$.
Then $X^c \subseteq W - S^{w,i}_{n(w,i)}$. However, everything in $W - S^{w,i}_{n(w,i)}$ has probability 0 w.r.t. P_i^w (Lemma 7.25) and since $(W, \mathcal{F}, P_i^w)$ is a classical probability space (Step 3), Observation 3.9 entails that X^c is abnormal, meaning that X is a priori w.r.t. P_i^w. □

Step 5

We will first prove that the sets in $\mathcal{G}^{w,i}$ are r-stable.

LEMMA 7.28. *Each $S^{w,i}_j$ is r-stable w.r.t. P_i^w, $\forall j \in \{1, ...n(w,i)\}$.*

PROOF. Proved directly by combining Lemma 7.26 (Step 2), Corollary 4.15 and Lemma 7.27 for $S^{w,i}_{n(w,i)}$. □

Now we will show that the family of r-stable sets w.r.t. P_i^w is the union of $\mathcal{G}^{w,i}$ and of the a priori sets.

LEMMA 7.29. *The only r-stable sets w.r.t. P_i^w are the ones in $\mathcal{G}^{w,i}$ or the a priori sets.*

PROOF. Fix i and w, and suppose that K is r-stable w.r.t. P_i^w, but is not in $\mathcal{G}^{w,i}$ or a priori.
Hence, K is a contingent stable set. By Corollary 4.12, it follows that every non-empty subset of K is normal.
This implies that $K \cap (W - S^{w,i}_{n(w,i)}) = \emptyset$ (since all subsets of $W - S^{w,i}_{n(w,i)}$ are abnormal, so the intersection $K \cap (W - S^{w,i}_{n(w,i)})$ is an abnormal subset of K, hence it must be empty).
Thus $K \subseteq S^{w,i}_{n(w,i)}$.

Let now j be the *smallest* index such that $K \subseteq S^{w,i}_j$. (We know that j's with this property exist, since already $n(w,i)$ has this property.)
This means that $K \subseteq S^{w,i}_j$, but that $K \not\subseteq S^{w,i}_m$ for any $m < j$. But both K and all these spheres (having $m < n(w,i)$) are contingent stable sets, hence they must be comparable by inclusion.

Therefore, we must have that
$S_m^{w,i} \subseteq K$ for all $m < j$, i.e. $\bigcup_{m<j} S_m^{w,i} \subseteq K$.
Moreover, this union $\bigcup_{m<j} S_m^{w,i}$ is either equal to $S_{j-1}^{w,i}$ (in case that $j > 0$) or equal to $\emptyset$ (in case $j = 0$), so in both cases K differs from this union (since K is stable thus non-empty, and not covered by case 1, so not equal to $S_{j-1}^{w,i}$). So we have a *strict* inclusion $\bigcup_{m<j} S_m^{w,i} \subset K$.

Let now w be some world $w \in K - (\bigcup_{m<j} S_m^{w,i}) \subseteq S_j^{w,i} - (\bigcup_{m<j} S_m^{w,i})$.
It is easy to see (by our construction of the probabilities for $S_j^{w,i}$), that we have $P_i^w(w) = x_j^{w,i}$.
By the same construction, we have $P_i^w(v) = x_j^{w,i}$ for all worlds $v \in S_j^{w,i} - K \subseteq S_j^{w,i} - (\bigcup_{m<j} S_m^{w,i})$ (since all the worlds in $S_j^{w,i} - (\bigcup_{m<j} S_m^{w,i})$ received probability $x_j^{w,i}$).

Finally, note that the inclusion $K \subseteq S_j^{w,i}$ is also strict (since K is not in $\mathcal{G}^{w,i}$). So, if we put $N_K := |S_j^{w,i} - K|$, then we have $N_K \geq 1$.
Putting all these together, we obtain:

$$P_i^w(K|\{w\} \cup (S_j^{w,i} - K)) = \frac{P(\{w\})}{P(\{w\} \cup (S_j^{w,i} - K))} = \frac{x_j^{w,i}}{x_j^{w,i} + N_K \cdot x_j^{w,i}} \leq \frac{x_j^{w,i}}{x_j^{w,i} + x_j^{w,i}} = \frac{1}{2},$$

which contradicts the r-stability of K. □

Step 6
Now define the following structure:

$$M' = (W, \mathcal{F}, \Pi_i, \mathbf{r}, P_i, ||\cdot||) \text{ where:}$$

- W as in the belief revision structure M,
- $\mathcal{F} = \mathcal{P}(W)$,
- Π_i partitions of W for $i \in Ag$, given by the $\sim_i$ defined above ($w \sim_i v$ iff $W_i^w = W_i^v$),
- $\mathbf{r}$ assigns the same number $r \in (\frac{1}{2}, 1)$ to all $i \in Ag$,
- the conditional probabilities $P_i^w(A|B)$ is given by applying the usual formula for conditional probabilities to the above-defined classical probabilities P_i^w (which are defined in terms of μ_i^w):
  $P_i^w(A|B) := \frac{P_i^w(A \cap B)}{P_i^w(B)}$,
- $||\cdot||$ is the same as in the belief revision structure M

We have that $\mathcal{F}$ is a $\sigma - algebra$ on W. Moreover, for some $i \in Ag$ and $Y \subseteq W$ such that Y is closed under $\sim_i$, we have that $Y \in \mathcal{F}$, by the definition of $\sim_i$. Finally, for each $i \in Ag$ and $w \in W$ we have that P_i^w is a two-place probability function over $\mathcal{F} \times \mathcal{F}$ (Observation 3.4) such that the space $(W, \mathcal{F}, P_i^w)$ is not trivial, the set $W - w(i)$ is abnormal (Lemma 7.27) and if $w' \in w(i)$ then $P_i^{w'} = P_i^w$.
Therefore M' is a non-trivial probabilistic model.

Step 7
And now we are ready for our final proposition:

PROPOSITION 7.30. $\forall w \in W : (M, w) \models \phi$ *iff* $(M', w) \models \phi$

PROOF. We will proceed by induction on the structure of ϕ.
The only interesting case is that of $B_i^{\phi}\psi$.

$(\Rightarrow)$
Assume that for $w \in W$ we have that: $w \models_M B_i^{\phi}\psi$.
Then $best_i^w(||\phi|| \cap W_i^w) \subseteq ||\psi||$.
Now we move to our probabilistic model M'.
If $||\phi|| \cap w(i) = \emptyset$, then $||\phi||$ is abnormal w.r.t. P_i^w, in which case $w \models_{M'} B_i^{\phi}\psi$.
If $||\phi|| \cap w(i) \neq \emptyset$, consider $S \subseteq w(i)$ the least r-stable set w.r.t. P_i^w such that $S \cap ||\phi|| \neq \emptyset$, i.e. $S \subseteq S'$, $\forall S'$: r-stable sets w.r.t. P_i^w such that $S' \cap ||\phi|| \neq \emptyset$.
Pick now $x \in S \cap ||\phi||$.
By the definition of P_i^w, we have that:
$S \cap ||\phi|| = best_i^w(||\phi|| \cap w(i)) \subseteq ||\psi||$.
Therefore $\emptyset \neq S \cap ||\phi|| \subseteq ||\psi||$, with S : r-stable set w.r.t. P_i^w.
Hence $w \models_{M'} B_i^{\phi}\psi$.

$(\Leftarrow)$
Assume that for some $w \in W$ we have that: $w \models_{M'} B_i^{\phi}\psi$.
Now if $||\phi||$ is abnormal w.r.t. P_i^w, then by the definition of P_i^w, we have that: $||\phi|| \cap S = \emptyset$ for all S r-stable sets w.r.t. P_i^w. This in turn implies that $best_i^w(||\phi||) = \emptyset$ and hence $w \models_M B_i^{\phi}\psi$.
Now if $||\phi||$ is normal, then we have that $\exists$ an r-stable set S w.r.t. P_i^w such that: $\emptyset \neq S \cap ||\phi|| \subseteq ||\psi||$.
Pick now the least r-stable set S' w.r.t. P_i^w such that $\emptyset \neq S' \cap ||\phi||$.
Then $S' \subseteq S$ hence $S' \cap ||\phi|| \subseteq S \cap ||\phi|| \subseteq ||\psi||$.
Now, by the construction of the model M' we get that: if $x \in S' \cap ||\phi||$, then $x \in best_i^w(||\phi|| \cap W_i^w)$.
Therefore $best_i^w(||\phi|| \cap W_i^w) \subseteq ||\psi||$ and thus $w \models_M B_i^{\phi}\psi$. □

And we have now completed the proof of Proposition 7.21. □

And finally:

PROPOSITION 7.31. **BRSID** *is a complete axiomatization w.r.t. non-trivial probabilistic frames.*

PROOF. Combine Propositions 7.21 and 7.9. □

CHAPTER 8

Safe Belief, Certainty operators and the notion of *quasi*-stability

Now that we have established a logic for conditional belief based on the notion of r-stability we are ready to proceed with our next goal: a logic with a language able to express statements such as "agent i has an r_i-stable belief that ϕ".

The straightforward way of doing that, would be to introduce an operator Sb_i^r. However, such an operator would not satisfy the **K** axiom. Therefore, we need another way of defining r-stable beliefs. To do so, we adopted the following more fundamental modalities: $\Box$: "safe belief" ([4]) and C: *certainty.* The idea comes from Baltag and Smets' papers, in which they provide a definition of conditional belief in terms of safe belief ($\Box$) and their notion of knowledge (K): $B_i^P Q = \tilde{K}_i P \to \tilde{K}_i(P \wedge \Box_i(P \to Q))$. In their setting, $\Box$ is an **S4** modality and K an **S5**. Now C for us would imply something stronger than probability 1 but nevertheless weaker than absolute truth. For a set A in our algebra CA would imply that A is a priori. Therefore, while A^c would be abnormal and hence $P(A^c|W) = 0$, that would not necessarily mean that $A^c = \emptyset$. On the other hand, $\Box$ will be neither truthful nor negatively introspective in our probabilistic setting.

We will follow Baltag and Smets' work in [6] and we will define conditional belief in terms of $\Box$ and C. We will do so in ω-stable conditional probability spaces (Definition 4.16), when we have countably many stable sets. Sadly, we will still be unable to define the notion of r-stability syntactically in terms of these operators. In fact, as we will discuss at the end of this chapter, after we have presented a formal account of our theory, in order to define the operator $Sb(\phi)$ (ϕ is an r-stable belief), we need the stronger **S5** modality: K, instead of our C. We decided to tackle this problem by defining a more general notion of stability, that of *quasi*-stability. This notion is based on the same idea the *quasi*-subset relation was based on in chapter 4.

This chapter will be divided in 3 sections.

First, we will introduce the $\Box$ and C operators. We will characterize sets of the form $\Box A$ for $A \in \mathcal{F}$ and prove some important properties about the $\Box$ operator.

Next, we will define the notion of *quasi*-stability. We will express it in terms of the $\Box$ and C operators and we will also show that in an ω-stable conditional probability space, a set Σ is *quasi*-stable if it is the union of an r-stable set S and an abnormal set a: $\Sigma = S \cup a$.

Finally, we will move on to conditional belief. Once again we will work in ω-stable conditional probability spaces and define conditional belief in terms of *quasi*-stable sets, proving the most important result of this chapter, namely that conditional belief can be fully expressed in terms of the $\Box$ and C operators, as long as we have countably many r-stable sets.

1. $\Box$ and C operators

Consider W a set of possible worlds, $(W, \mathcal{F}, P)$ a conditional probability space and fix some threshold $r \in (\frac{1}{2}, 1]$.
Now we will define two new operators:

DEFINITION 8.1. *For $A \in \mathcal{F}$:*

$$x \in \Box A \textit{ iff } \forall E \in \mathcal{F}(x \in E \Rightarrow B^E A).$$

We read $w \in \Box A$ as saying: "at state w our agent safely believes that A".

DEFINITION 8.2. *For $A \in \mathcal{F}$ we have that:*

$$CA := B^{\neg A}\bot$$

We read CA as "A is an a priori set". It is easy to see that the C operator is global in the sense that either A is a priori on every state $w \in W$ or A is not a priori anywhere in W. However, C is unlike knowledge in the sense that it is not necessarily truthful: $C\phi \to \phi$ is not sound in our probabilistic semantics. As discussed above for a formula ϕ, $C\phi$ implies that $P(||\phi|||W) = 1$ and $P(||\phi||^c|W) = 0$ but it is important to note that this does not imply that $||\phi||^c = \emptyset$, i.e. that $||\phi|| = W$.

Now we have the two following lemmas:

LEMMA 8.3 (Monotonicity of the $\Box$ operator). *For $A, B \in \mathcal{F}$, if $A \subseteq B$, then $\Box A \subseteq \Box B$.*

PROOF. Take $A, B \in \mathcal{F}$ such that $A \subseteq B$.
Let $x \in \Box A$.
Then we have that for all $E \in \mathcal{F}$, if $x \in E$ then $B^E A$ holds.
However, we have shown in chapter 5 (5.8) that if $A \subseteq B$ and $B^E A$ holds we have that $B^E B$ holds as well.
Therefore we get that $\forall E \in \mathcal{F}$ if $x \in E$ then $B^E B$ holds.
Therefore $x \in \Box B$. $\Box$

LEMMA 8.4. *For a set $A \in \mathcal{F}$ we have that either $\Box A \subseteq A$ or that A is a priori.*

PROOF. Consider $A \in \mathcal{F}$ and assume that $\Box A \not\subseteq A$.
Then $\exists x \in \Box A$ such that $x \in A^c$.
But then we get that $B^{A^c} A$, which entails that A is a priori. $\Box$

Finally, the above lemma gives us the following interesting corollary:

COROLLARY 8.5. *For all sets $A \in \mathcal{F}$, we have that:*

$$\Box A \subseteq_q A$$

Now we will define r-stable beliefs in terms of the $\Box$ operator:

PROPOSITION 8.6. *$S \in \mathcal{F}$ is an r-stable set iff $S \neq \emptyset$ and $S \subseteq \Box S$.*

PROOF. $(\Rightarrow)$
Consider $S \in \mathcal{F}$ an r-stable set.
Then $S \neq \emptyset$ by definition.
Now pick $x \in S$.
We need to show that $x \in \Box S$, i.e. that $B^E S$ holds for all $E \in \mathcal{F}$ such that $x \in E$.
Pick $E \in \mathcal{F}$ such that $x \in E$.
Then we need to show that $B^E S$ holds.

- Case 1. E is abnormal. Then $B^E S$ holds.
- Case 2. E is normal. Then we have that $x \in S \cap E$.
  Therefore $S \cap E \neq \emptyset$.
  And we also have that $S \cap E \subseteq S$ with S an r-stable set, so indeed $B^E S$ holds.

$(\Leftarrow)$
Take $S \in \mathcal{F}$ such that $S \neq \emptyset$ with $S \subseteq \Box S$.
We need to show that S is an r-stable set.
Pick $E \in \mathcal{F}$ such that $S \cap E \neq \emptyset$.
Then since $S \subseteq \Box S$ we have that $B^E S$ holds.
This entails that $\exists S' \in \mathcal{F}$ such that S' is r-stable with $S' \cap E$ a normal set and $S' \cap E \subseteq S$.

- Case 1. E is abnormal. Then $P(S|E) = 1 \geq r$.
- Case 2. E is normal. Then $P(S|E) \geq P(S' \cap E|E) = P(S'|E) \geq r$, since S' is r-stable and $S' \cap E \neq \emptyset$.

Therefore S is r-stable and our proof is complete. $\Box$

The above is an important result that connects the $\Box$ operator directly with r-stability.

The following propositions characterize sets of the form $\Box A$.

PROPOSITION 8.7. *Let $A \in \mathcal{F}$. Then $w \in \Box A$ holds iff we have either*

- *A is a priori, or*
- *$\exists S$: r-stable set in $\mathcal{F}$ such that $w \in S$ and $S \subseteq A$.*

PROOF. $(\Rightarrow)$
Suppose that $w \in \Box A$ but $\nexists S$: r-stable set such that $w \in S$ and $S \subseteq A$.
Given this we want to show that A is a priori.
Call $\mathcal{H}$ the family $\mathcal{H} := \{S \in \mathcal{F} | S : r\text{−stable and } S \subseteq A\}$.
We have that $\mathcal{H}$ is non-empty, since we have that $w \in W$ and $w \in \Box A$, hence $B^W A$ holds.
Therefore $\exists S$: r-stable set such that $S \cap W = S$ is normal and $S \cap W = S \subseteq A$.
Hence this particular S is in $\mathcal{H}$.

Now define $D := \bigcup \mathcal{H}$, the union of the elements of $\mathcal{H}$.
Now $\forall H \in \mathcal{H} : H \subseteq A$, hence $D \subseteq A$, therefore $A^c \subseteq D^c$.
Now we have assumed that w does not belong to any of the sets in $\mathcal{H}$, since we have assumed that $\nexists S$: r-stable set such that: $w \in S$ and $S \subseteq A$.
Hence w does not belong in D as well. Therefore $w \in D^c$.
Now we want to show that A^c is abnormal.
To do this, it is enough to show that D^c is abnormal, by Property 3.12.

Claim: D^c *is abnormal*

PROOF. Suppose D^c is normal. Then since $w \in \Box A$ and $w \in D^c$ we get $B^{D^c} A$.
Therefore $\exists S'$: r-stable set such that $S' \cap D^c$ is normal and $S' \cap D^c = S' - D \subseteq A$.
However, we also have that $S' \cap D \subseteq A$, since $S' \cap D \subseteq D$ and $D \subseteq A$.
Therefore: $S' = (S' - D) \cup (S' \cap D) \subseteq A$.
But then $S' \in \mathcal{H}$. Hence $S' \subseteq D$, meaning that $S' - D = \emptyset$.
But this contradicts that $S' - D$ is normal.
Hence D^c is indeed abnormal. $\square$

This entails that A^c is abnormal as well, as its subset and this gives us that A is a priori, as desired.

($\Leftarrow$)
Consider $A \in \mathcal{F}$ and $x \in W$.
We split our premise into two subcases:

- Case 1. A is a priori.
  Then $\forall E \in \mathcal{F}$ we have: $B^E A$.
  Therefore: $\Box A = W$.
- Case 2. A is contingent and $\exists S$: r-stable set such that $x \in S$ and $S \subseteq A$.
  Let $E \in \mathcal{F}$ be any set in the algebra such that $x \in E$.
  Then $E \cap S \subseteq A$.
  Moreover, A is contingent and $S \subseteq A$. Therefore S is contingent.
  Also, $E \cap S \neq \emptyset$, since $x \in E$.
  Now by Corollary 4.12 $S \cap E$ is normal.
  Therefore $B^E A$ holds.
  Hence $x \in \Box A$.

$\square$

COROLLARY 8.8. *We have that*

(a) *if* $A \in \mathcal{F}$ *is a priori, then* $\Box A = W$.
(b) *if* $A \in \mathcal{F}$ *is contingent, then* $\Box A$ *is the union of all the stable subsets of* A.

PROPOSITION 8.9. *For* $A \in \mathcal{F}$, *we have that if* $\Box A$ *is a priori then* $\Box A = W$.

PROOF. Let $A \in \mathcal{F}$ such that $\Box A$ is a priori.
Then we have two cases:

- Case 1: A is a priori. Then by Corollary 8.8 we have that $\Box A = W$.
- Case 2: A is contingent. Then Lemma 8.4 tells us that $\Box A \subseteq A$.
  But this together with $\Box A$ being a priori, implies that A is a priori. Contradiction.

$\square$

PROPOSITION 8.10. *For all $A \in \mathcal{F}$, we have that: $\Box A = \Box\Box A$.*

PROOF. Pick $A \in \mathcal{F}$. We have the two following cases:

- Case 1. A is a priori.
  Then by Corollary 8.8 we have that $\Box A = W$, which is also a priori, hence $\Box\Box A = \Box W = W = \Box A$.
- Case 2. A is contingent.
  Then by Corollary 8.8 $\Box A$ is the union of all stable subsets of A. Also, $\Box A \subseteq A$ in this case, thus $\Box A$ is contingent (since A is), hence $\Box\Box A$ is the union of all stable subsets of [the union of all stable subsets of A].
  But (using nestedness of contingent stable sets (Corollary 4.6)), this is just the union of all stable subsets of A, i.e. $\Box A$.

$\Box$

Finally, we have a result for sets of the form $\Box A$.

PROPOSITION 8.11. *Consider $(W^\omega, \mathcal{F}^\omega, P^\omega)$ to be an $\omega-$stable conditional probability space as in* Definition 4.16.
Then the sets of the form $\Box A$ for $A \in \mathcal{F}^\omega$ are stable.

PROOF. Pick $A \in \mathcal{F}^\omega$. We have two cases:

- Case 1. If A is a priori, then $\Box A = W^\omega$, in which case it is stable.
- Case 2. If A is contingent, by Corollary 8.8 we have that $\Box A$ is the union of all the stable subsets of A, i.e. $\Box A = \bigcup\limits_k S_k$ with S_k contingent (as subsets of A) stable sets. By the definition of the $\omega - \textit{stable}$ conditional probability space, we get that these stable sets S_k are countably many. Now by Property 4.8 we have that $\bigcup\limits_k S_k$ is a stable set. Therefore $\Box A$ is indeed a stable set.

$\Box$

Notice that the set $\bigcup\limits_k S_k$ is stable, only because we have countably many stable sets. In more general (not necessarily ω-stable) conditional probability spaces, Proposition 8.11 does not hold.

2. *quasi*-stable sets

In this section we will define a new notion of stability, called: *quasi*-stability.

Let $(W, \mathcal{F}, P)$ be a conditional probability space.

DEFINITION 8.12. *We will say that a set $\Sigma \in \mathcal{F}$ is* quasi-stable *and write: $QSb(\Sigma)$ iff $\Sigma \neq \emptyset$ and $\exists a$ abnormal set such that $\Sigma \subseteq \Box\Sigma \cup a$.*

Notice that the above definition also entails that a set $\Sigma \in \mathcal{F}$ is *quasi*-stable if and only if $\Sigma \neq \emptyset$ and $\Sigma \subseteq_q \Box\Sigma$, with $\subseteq_q$ the *quasi*-subset relation defined in chapter 4 (Definition 4.3).

The following proposition connects *quasi*-stable and r-stable sets.

PROPOSITION 8.13. *Let* $(W^\omega, \mathcal{F}^\omega, P^\omega)$ *be an* ω*-stable conditional probability space as in* Definition 4.16.

Now let $\Sigma \in \mathcal{F}^\omega$*. Then* Σ *is quasi-stable iff there exists a stable set* $S \in \mathcal{F}^\omega$ *and an abnormal set* $a \in \mathcal{F}^\omega$ *such that* $\Sigma = S \cup a$.

PROOF. ($\Rightarrow$) Let $\Sigma \in \mathcal{F}^\omega$ a *quasi*-stable set.

We have the following cases:

- Case 1. Σ is a priori. Then $\Sigma = \Sigma \cup \emptyset$, with Σ stable (as a priori) and $\emptyset$ abnormal.
- Case 2. Σ is contingent. Then by Corollary 8.8 $\Box\Sigma$ is the union of all the stable subsets of Σ: $\Box\Sigma = \bigcup_n S_n$ with S_n stable sets. Since we are in an ω-stable space, we have that these S_n sets are countable. Now by Property 4.8 we have that $\bigcup_n S_n$ is stable, thus $\Box\Sigma$ is stable.

  Thus we take $S = \Box\Sigma$ and this direction is complete.

($\Leftarrow$)
Let $S \in \mathcal{F}$ stable, i.e. $S \neq \emptyset$ and $S \subseteq \Box S$.
Let $a \in \mathcal{F}$ abnormal. We have $S \cup a \subseteq \Box S \cup a$ and $S \cup a \neq \emptyset$.
Hence by Definition 8.12 $S \cup a$ is *quasi*-stable. $\Box$

3. Conditional Belief: *quasi*-stability and $\Box$, C operators

In the final section of this chapter, we are going to define conditional belief in terms of the $\Box$ and C operators, analogously to what Baltag and Smets did in [6].
In this section we will be working in ω*-stable conditional probability spaces as in* Definition 4.16*. In these spaces we have countably many contingent stable sets.*

For notational simplicity we take $W^\omega = W$, $\mathcal{F}^\omega = \mathcal{F}$ and $P^\omega = P$.

First we define conditional beliefs in terms of *quasi*-stable sets.

PROPOSITION 8.14. *For* $E, H \in \mathcal{F}$*, we have:* $B^E H$ *iff either* E *is abnormal or* $\exists\Sigma$: $quasi-stable$ *set such that* $\Sigma \cap E$ *is normal and* $\Sigma \cap E \subseteq H$.

PROOF. ($\Rightarrow$) Assume that $B^E H$ holds for $E, H \in \mathcal{F}$.
Then $\exists S$ stable set such that $S \cap E$ is normal and $S \cap E \subseteq H$. Now since $\emptyset$ is an abnormal set, we take $\Sigma = S \cup \emptyset = S$ and we are done.

($\Leftarrow$) Take $E, H \in \mathcal{F}$ such that E is normal (if E is abnormal then it is trivial) and assume that there exists a set $\Sigma \in \mathcal{F}$ such that Σ is a *quasi*-stable set with $\Sigma \cap E$ a

normal set and $\Sigma \cap E \subseteq H$.
Then by Proposition 8.13 $\exists S \in \mathcal{F}$ such that S is stable and $\Sigma = S \cup a$ with $a \in \mathcal{F}$ an abnormal set. We have the following two cases:

- Case 1. S is a priori.
  Then if E is normal we get that $S \cap E$ is normal.
  We also have that $S \cap E \subseteq \Sigma \cap E \subseteq H$ and we are done.
- Case 2. S is contingent.
  Then $\Sigma \cap E = (S \cap E) \cup (a \cap E)$.
  But $\Sigma \cap E$ is a normal set and $a \cap E$ an abnormal set, therefore $S \cap E$ should be a normal set as well, since otherwise we would have that $\Sigma \cap E$ is abnormal by Property 3.13.
  Having also that $S \cap E \subseteq H$ we derive that $B^H E$ holds.

$\Box$

Hence we have established a new definition for conditional belief in terms of *quasi*-stable sets. However, this result holds only in ω-stable spaces.

And now we will move on to the central point of this section, which is the definition of conditional belief in terms of the $\Box$ and C operators:
(use $\tilde{C}(E)$ for $\neg C(\neg E) = \neg C(E^c)$)

PROPOSITION 8.15. *$B^E H$ iff $\tilde{C}E \Rightarrow (\tilde{C}(E \wedge \Box(E \to H)))$*

PROOF. First notice that $\tilde{C}E = \neg C \neg E$ essentially means that E is normal.

($\Rightarrow$)
Assume that $B^E H$ holds for some $E, H \in \mathcal{F}$.
We have two cases:

- Case 1. E is abnormal. Then the right hand side implication holds trivially.
- Case 2. E is normal.
  We want to show that $E \cap \Box(E \to H)$ is normal as well.
  Since $B^E H$ holds, we get that:
  $\exists S$: r-stable set such that $S \cap E$ is normal and $S \cap E \subseteq H$.
  Since S is stable we have that $S \subseteq \Box S$.
  Moreover, we have that $S \subseteq (S \cap E) \cup (S \cap E^c)$.
  Now $S \cap E \subseteq H \subseteq E^c \cup H$.
  Also $S \cap E^c \subseteq E^c$.
  Therefore $S \subseteq (S \cap E) \cup (S \cap E^c) \subseteq (E^c \cup H) = (E \to H)$.
  Therefore $S \subseteq E \to H$.
  Now by 8.3 we have that $\Box$ is a monotonic operator.
  Therefore $\Box S \subseteq \Box(E \to H)$.
  Hence $E \cap S \subseteq E \cap \Box S \subseteq E \cap \Box(E \to H)$.
  Now $E \cap S$ is a normal set, otherwise by 4.11 we would have that S is a priori, in which case we would have by 4.13 that E itself is abnormal, contradicting our assumption.
  Therefore $E \cap \Box(E \to H)$ is a superset of a normal set and therefore by Property 3.11 we get that $E \cap \Box(E \to H)$ is a normal set as well.

($\Leftarrow$)
Assume that for some sets $E, H \in \mathcal{F}$ we have that $\tilde{C}E \Rightarrow \tilde{C}(E \cap \Box(E \to H))$ holds.

As written above this means that if E is normal then $E \cap \Box(E \to H)$ is normal as well. We want to show that $B^E H$ holds.
We have two cases:

- Case 1. E is abnormal. Then by Corollary 5.2 $B^E H$ holds.

- Case 2. E is normal.
  Then we have the two following subcases:
  - Subcase 1. $(E \to H)$ is not a priori.
    Then $\Box(E \to H) \subseteq (E \to H)$, by Lemma 8.4.
    Therefore $E \cap \Box(E \to H) \subseteq H$, since $(E \to H) = E^c \cup H$.
    Now by Proposition 8.11, the set $\Box(E \to H)$ is stable.
    Moreover we know that $E \cap \Box(E \to H)$ is normal by our assumptions.
    Hence combining these two we get that $B^E H$ holds.

  - Subcase 2. $(E \to H)$ is a priori.
    Then $(E \to H)^c = E - H$ is abnormal.
    However we have that E is normal.
    Hence $E \cap H$ is normal, otherwise $E = (E - H) \cup (E \cap H)$ would be abnormal contradicting our assumption.
    Now since $(E \to H)$ is a priori, it is also stable.
    Therefore we have that: $E \cap (E \to H) = E \cap H \subseteq H$, with $(E \to H)$ stable and $E \cap (E \to H)$ a normal set, since $E \cap H$ is normal.
    Therefore $B^E H$ holds.

$\Box$

We saw that S is stable if and only if $S \neq \emptyset$ and $S \subseteq \Box S$.
This is equivalent to saying that S is stable if and only if $S \neq \emptyset$ and $(S \cap (\Box S)^c) \neq \emptyset$. To express this in our language, we need an operator K such that $K\phi \to \phi$ holds. In that case we could define an operator Sb for stable beliefs as:

$$Sb(S) \Leftrightarrow \neg K \neg S \wedge K(S \to \Box S).$$

Then $\neg K \neg S$ tells us that $S \neq \emptyset$ and that $S^c \cup \Box S$ holds, i.e. that $(S \cap (\Box S)^c) \neq \emptyset$, i.e. that $S \subseteq \Box S$.
However, the closest we can get to this K operator is C. CA now tells us that A is a priori. Now if we try to define stable sets in terms of the C operator, we run into problems:

$$Sb(S) \Leftrightarrow \neg C \neg S \wedge C(S \to \Box S).$$

Now the first part: $\neg C \neg S$ tells us that S is normal. As we explained in chapter 4 this is in fact equivalent to saying that S is non-empty when it comes to r-stability. The problem however comes in the second part of the conjunction. This is because $C(S \to \Box S)$ tells us that the set $S^c \cup \Box S$ is a priori, i.e. that the set $S \cap (\Box S)^c$ is abnormal, i.e. that there exists an abnormal set a such that $S \subseteq \Box S \cup a$, i.e. that $S \subseteq_q \Box S$. But this is not enough to conclude that S is stable. This observation motivates the concept of *quasi*-stability. As we saw in Proposition 8.14, when in ω-stable conditional probability spaces defining conditional beliefs in terms of *quasi*-stable sets is equivalent to our old definition that was in terms of stable sets. Therefore we can express conditional beliefs using *quasi*-stable sets. However, this is possible only in spaces that we have countably many stable sets.

CHAPTER 9

Conclusion

1. Overview

The goal of this thesis was twofold. First, to extend Leitgeb's theory of stable beliefs into non-classical probability spaces, where conditioning on events with measure 0 is defined. Furthermore, to develop a formal language in order to express the notion of conditional belief.

We first gave a brief presentation of Leitgeb's stability theory as found in his papers [35], [34]. We argued that his theory is a path between the materially wrong proposal that equates belief with probability 1 and the logically wrong Lockean Thesis (or any version of it). We believe that Leitgeb's formalization of belief is more intuitive than the current alternatives and at the same time it comes without a cost, since we maintain the logical closure of belief without running into the lottery paradox. However, his classical probability setting is not enough to do belief revision, because conditioning on events of measure 0 can not be defined.

Therefore, our first goal was to extend Leitgeb's ideas into a non-classical probability space. According to Halpern ([29]) the three most popular approaches of dealing with conditioning on sets of measure 0 are: conditional probability spaces, lexicographic probability spaces and nonstandard probability spaces. We chose to develop our notion of r-stable sets using conditional probability spaces ([12], [13], [29]) and more specifically in Van Fraassen's setting ([25], [3], [21]).

We used our two-place probability functions, to define the notion of r-stability similarly to Leitgeb. A set is r-stable if its probability remains above the threshold r given the occurrence of any event consistent with it. We also proved that r-stable sets are nested w.r.t. the $\subseteq_q$ relation and that there is no infinitely descending chain of *contingent* r-stable sets.

The next step was to define conditional belief in terms of r-stable sets: agent i believes H given E if and only if either E is abnormal, or there is an r-stable set S such that $S \cap E \neq \emptyset$ and $S \cap E \subseteq H$. This definition comes in accordance with what Leitgeb calls the "Humean conception of belief", ([35, p. 33]) that "it is rational to believe a proposition in case it is rational to have a stably high degree of belief in it". Moreover, if E (the evidence set) is abnormal, we followed Van Fraassen's idea that the agent is so confused that does not know what to believe anymore and hence he believes everything. Finally, we showed that our conditional belief is closed under conjunction and is consistent w.r.t. normal sets.

Our next goal was to develop a formal language to express statements such as "agent i believes H given E".

In chapter 6 we defined the structures called probabilistic frames; the structures that gave us the semantics for our logic of conditional belief in chapter 7. They are essentially the multi-agent version of the conditional probability spaces we defined in chapter 3 for Ag a finite set of agents. Now a probabilistic frame is a structure in which the set of possible worlds W is divided into partitions Π_i: one for each agent $i \in Ag$. Therefore, for each world $w \in W$ we have an information cell $w(i)$ for each i, induced by the partition. Moreover, we have a function $\mathbf{r}$ that assigns a number $r_i \in (\frac{1}{2}, 1]$ to each of our agents; their threshold. Finally, we defined a function P_i that assigns a two-place probability function P_i^w to each agent i at each state w. Therefore, we are now talking about r_i-stable sets w.r.t. P_i^w. At the end of chapter 6, we made a comparison of our notion of r-stable and conditional belief with Battigalli and Siniscalchi's. It appeared that B-S's "strong belief" can be expressed in our terms and is essentially the 1-stable belief (r-stable belief for threshold $r = 1$).

In chapter 7, we presented our logic of r-stable conditional beliefs. In the first section we presented our language, that of epistemic logic with the addition of the $B_i^{\phi}\psi$ operator expressing "agent's i belief in ψ given ϕ" and our semantics, the probabilistic models. In section 2, we presented our axiom system, which is similar to Board's in his logic of conditional beliefs ([18]). In section 3 we proved that our axioms are sound in our probabilistic semantics and in section 4 we proved the completeness of our axiom system w.r.t. our probabilistic models.

Our next goal was to introduce a formal language that would allow us to express statements such as "agent i has a stable belief in agent's j rationality". The straightforward way to achieve this, would be to introduce an operator $Sb_i^r(\phi)$ stating: "ϕ is an r-stable belief for agent i". However, such an operator would not satisfy the **K** axiom. Therefore, we had to go another way: express r-stable beliefs in terms of other operators that do satisfy the **K** axiom. We chose Baltag and Smets' modality $\Box$ ([4]) for *safe belief* and the modality C for *certainty*.

However, these operators were still unable to capture our notion of r-stable sets. This is because (as we argue in the last part of chapter 8) $\Box$ and C are not *truthful* in our setting. Therefore, we defined a new notion of stability: that of *quasi*-stability. This notion is minimally different from r-stability in the sense that it is more general. We showed that in a conditional probability space with countably many stable sets, a *quasi*-stable set is the union of an r-stable set with an abnormal set. Therefore, a *quasi*-stable belief has an abnormal part. We also showed that in such spaces, our notion of conditional belief can be equally expressed in terms of *quasi*-stable sets.

2. Future Work

The first step after this thesis, is to develop the logic of certainty and safe belief based on the modalities: $\Box$ and C. Let us call this logic **QSBL**.

Below, we will present our ideas about the language, semantics and a possible axiomatization of **QSBL**.

The language of **QSBL**, $\mathcal{L}_{\mathbf{QSBL}}$ can be defined as follows:

Definition 9.1. *Consider a set of atomic sentences At. Our language $\mathcal{L}_{\mathbf{QSBL}}$ is a set of formulas ϕ of* **QSBL** *and is defined recursively:*

$$\phi ::= p|\neg\phi|\phi \wedge \phi|\Box\phi|C\phi$$

for $p \in At$.

Once again, we have the language of epistemic logic augmented by adding $\Box$ and C operators. As it has been mentioned above $\Box\phi$ is read as: "agent safely believes that ϕ" and $C\phi$ is read as: "agent is certain of ϕ".

We use the standard abbreviations: $\phi \vee \psi$ for $\neg(\neg\phi \wedge \neg\psi)$, $\phi \rightarrow \psi$ for $\neg\phi \vee \psi$ and $\phi \leftrightarrow \psi$ for $(\phi \rightarrow \psi) \wedge (\psi \rightarrow \phi)$. Moreover, define $\bot = p \wedge \neg p$ for $p \in At$ and $\top = \neg\bot$. Finally abbreviate $\neg C \neg\phi$ as $\tilde{C}\phi$.

Now define the following operators:

Definition 9.2. *For ϕ, ψ formulas in $\mathcal{L}_{\mathbf{QSB}}$, we define:*

$$B^{\phi}\psi = \tilde{C}\phi \rightarrow \tilde{C}(\phi \wedge \Box(\phi \rightarrow \psi))$$

As mentioned above conditional belief is defined in terms of C and $\Box$. For the validity of this definition, look at Proposition 8.15.

Definition 9.3. *For ϕ a formula in $\mathcal{L}_{\mathbf{QSB}}$, we define:*

$$QSb(\phi) = \tilde{C}\phi \wedge C(\phi \rightarrow \Box\phi)$$

Here we defined the notion of *quasi*-stability in terms of C and $\Box$. This definition is motivated by our discussion at the last part of chapter 8.

Now for our semantics first consider a number $r \in (\frac{1}{2}, 1]$.
Now one should take care to note that the above definitions only hold in ω-stable conditional probability spaces. Therefore our probabilistic frames and models should be in ω-stable spaces, in which we have countably many stable sets.

With this in mind, our semantics are given by the structure

$$M = (W, \mathcal{F}, P, ||\cdot||)$$

with

- $(W, \mathcal{F}, P)$ is a non-trivial ω-stable conditional probability space and
- $||\cdot|| : \mathcal{L}_{\mathbf{QSBL}} \rightarrow \mathcal{F}$ a valuation.

M will be called a *single agent ω-probabilistic model.*[13]

As before, we require that for $p \in At$: $||p|| \in \mathcal{F}$.
We define truth as follows:

Definition 9.4. *Consider $M = (W, \mathcal{F}, P, ||\cdot||)$ a single agent probabilistic model, $w \in W$ and $p \in At$. The relation $\models$ between pairs (M, w) and formulas $\phi \in \mathcal{L}_{\mathbf{QSBL}}$ is defined as follows:*

- $(M, w) \models \phi$ *iff* $w \in ||p||$,
- $(M, w) \models \neg\phi$ *iff* $w \in ||\phi||^c$,
- $(M, w) \models \phi \wedge \psi$ *iff* $w \in ||\phi|| \cap ||\psi||$,
- $(M, w) \models \Box\phi$ *iff* $\forall E \in \mathcal{F}(w \in E \Rightarrow B^E||\phi||)$,
- $(M, w) \models C\phi$ *iff* $||\phi|| \in \mathcal{F}$ *is a priori.*

[13]Notice that the *single agent probabilistic model* is essentially the single-agent restriction of our probabilistic model.

Finally, one should take care to show that $||\cdot||$ is well-defined and in $\mathcal{F}$.

Notice that since we are in ω-stable conditional probability spaces, sets of the form $\Box A$ for some $A \in \mathcal{F}$ will be measurable. This is because if $A \in \mathcal{F}$ is a priori, then by Corollary 8.8 $\Box A = W$ and therefore $\Box A \in \mathcal{F}$. If A is contingent, then by Corollary 8.8 we have that $\Box A$ is the union of all the stable subsets of A. However, these are countably many (ω-stable conditional probability spaces) and therefore their union is in the algebra. Hence $\Box A$ is measurable.

Now here is an axiom system for this logic. Consider $\phi, \psi \in \mathcal{L}_{\mathbf{QSBL}}$.
Our axioms and inference rules are the following, along with **MP**:

Taut	true
Nec	from ϕ infer $C\phi$ and $\Box\phi$
K for C	$C(\phi \to \psi) \Rightarrow (C\phi \to C\psi)$
PI for C	$C\phi \to CC\phi$
NI for C	$\neg C\phi \to C\neg C\phi$
K for $\Box$	$\Box(\phi \to \psi) \Rightarrow (\Box\phi \to \Box\psi)$
PI for $\Box$	$\Box\phi \to \Box\Box\phi$
(a)	$C\phi \to \Box\phi$
(b)	$\neg\Box\bot$
(c)	$\Box\phi \wedge \neg C\phi \to \phi$
(d)	$C(\Box\phi \to \psi) \vee C(\Box\psi \to \phi)$

The first important observation is that neither of $\Box$ and C is truthful. The axioms $\Box\phi \to \phi$ and $C\phi \to \phi$ can not be sound in our probabilistic semantics.

Furthermore, $\Box$ is not negatively introspective. Now $C\phi \to \Box\phi$ says that if ϕ is a priori, then ϕ is a safe belief. In other words, this axiom tells us that the notion of a priori is stronger than that of safe belief. $\neg\Box\bot$ says that contradiction can not be safely believed.

Finally, $\Box\phi \wedge \neg C\phi \to \phi$ is as close we can get to factivity. This axiom says that if ϕ is safely believed, then either ϕ is the case, or ϕ is a priori. Intuitively, this axiom says that an agent can be "wrong" about his safe beliefs, only if what he safely believes is an a priori set. In other words, safe belief looses its "credibility" when it coincides with one of our agent's a priori convictions (certainties).

Now the next step is proving soundness and completeness w.r.t. the single agent probabilistic models.

At this point, we would like to write some comments about the completeness proof.

Completeness could be established in a way similar to what we did in section 7. First, we could use canonical models as in [15] and prove completeness w.r.t. finite partial plausibility models:

DEFINITION 9.5. *The structure* $(W, S, \leq, ||\cdot||)$ *such that:*

- W *is a finite set of possible worlds,*
- $S \subseteq W$*,*
- $\leq$ *a plausibility ordering on* S *such that:*

– $\leq$ *is complete and transitive on* S,
– $\leq$ *is well-founded*

- $||\cdot||$ *a function assigning formulas of our language to subsets of* W,

will be called a finite partial plausibility model.

Afterwards, we can "probabilize" a finite partial plausibility model in a way similar to what we did in section 7 and obtain a single agent probabilistic model.

Finally, we prove a truth preserving lemma between finite partial plausibility models and single agent ω-probabilistic models and we derive completeness w.r.t. our probabilistic semantics.

Now the next step is the multi-agent version of the logic of certainty and safe belief (**QSBL**).

The language of this logic can be

$$\phi ::= p|\neg\phi|\phi \wedge \phi|\Box_i^{\phi}\phi|C_i\phi$$

for $i \in Ag$ with Ag a finite set of agents.

The semantics will be given by our probabilistic frames as defined in chapter 6. However, we will be referring to ω-stable conditional probability spaces.

Once we have the multi-agent logic of certainty and safe beliefs, the next step is to introduce the notions of Mutual and Common *quasi*-stable beliefs in ϕ: $EQSb(\phi)$ and $CQSb(\phi)$ respectively.

Now $EQSb(\phi)$ is defined as usual:

$$EQSb(\phi) := \bigwedge_{i \in Ag} QSb_i(\phi)$$

and the usual axioms for $CQSb(\phi)$ are:

- $CQSb(\phi) \Rightarrow EQSb(\phi \wedge CQSb(\phi))$
- $CQSb(\phi \to EQSb(\phi)) \Rightarrow (EQSb(\phi) \to CQSb(\phi))$

At this point there are two possible directions one could take. Either introduce a logic of Common *quasi*-stable beliefs along with the necessary soundness and completeness proofs, or simply define these notions and apply them.

One possible application would be the characterization of the epistemic conditions of backward/forward induction.

We have already mentioned that Battigalli and Siniscalchi use their notion of strong belief to characterize backward induction.

Baltag, Smets and Zvesper do something analogous in their paper [10]. They introduce a new notion of rationality: *dynamic rationality* and use the notion of "stable belief", belief that is preserved during the play of the game, to provide a characterization of the epistemic conditions for backward induction: dynamic rationality and common knowledge of stable belief in rationality.

Here we will give a very brief and informal presentation of their work, only to show that there is a direct connection to our theory.

First, concerning their notion of dynamic rationality.

On one hand, this notion assesses the rationality of a player's move at a node w.r.t. the beliefs held when this node is reached ([10, p. 304]). On the other hand, this notion of rationality also incorporates the epistemic limitation to rationality: the rationality of an agent's move only makes sense when that move is not already known to her ([10, p. 304]).

They argue that their notion of rationality is future-oriented in the sense that at any stage of the game, the "dynamic rationality" of an agent depends only on her current and future moves. Therefore, a player can be rational *now*, even if he has made irrational moves before ([10, p. 305]). Suppose now that player i behaves irrationally at some node. Then the rest of the players do learn that i is irrational *at this moment*, but this may as well be forgotten as the game continues. This means that a previously irrational player can become rational after her wrong move. The intuition is that she might choose the right moves for all the decisions that she can still make ([10, p. 305]).

Hence, the meaning of "rationality" changes in time, due to the change of beliefs and of the known set of options ([10, p. 305]).

Their main argument is that the rationality of a player is an epistemic-doxastic concept, so it should be affected by any changes of the information of the player.

The main theorem of their paper is the following:

Common knowledge of (the game structure, of "open future" and of) stable common belief [14] in dynamic rationality entails common belief in the backward induction outcome ([10, p. 306]).

Now the "open future" assumption essentially is that players have no non-trivial "hard"[15] information about the outcomes of the game ([10, p. 329]).

The notion of "stable belief" corresponds to Baltag and Smets' notion of safe belief ([4]): an analogy of the operator $\Box$ we used in our chapters 8, 9. However, as it has been mentioned quite a few times before, there is a big difference (amongst others): our version is *not* truthful.

The first question that comes up at this point, is what would happen if we used our $quasi - stale$ belief in the place of their "stable belief". So suppose that we assumed that we have common knowledge of the game, of "open future" and of *quasi*-stable common belief in dynamic rationality. Would we get common belief in the backward induction outcome? Moreover, to what extent would the result depend on the selection of the threshold r?

The next point, is that B-S's work in [13] (briefly presented in chapter 2) is really close to what Baltag, Smets and Zvesper did. However, B-S's notion of rationality is only "partially-dynamic": it requires the agents to make rational choices at *all the nodes*, including the ones that have already been bypassed ([10, p. 329]). Therefore, this implies that *one irrational move* is enough to break the common belief in rationality.

This is why B-S use the assumption of the complete-type structure (Definition 2.4) in order to characterize forward induction by assuming common strong belief in rationality (with their notion of strong belief as presented above). A complete-type structure contains every possible epistemic-doxastic type of each player. This assumption demands that players consider all probabilistic assignments as epistemically possible ([10, p. 330]).

[14]As they state: "common" here is not necessary, since common knowledge that everybody has a stable belief in P is the same as common knowledge of common safe belief in P ([10, p. 306]).

[15]"hard" as absolutely "indefeasible", unrevisable information as in [41]

Therefore, the next interesting question is whether we can assume common *quasi*-stable belief in *dynamic rationality* and characterize the epistemic conditions of forward induction in a way similar to B-S, only by "loosening" their really strict assumption of complete-type structures. Once again, how would the choice of the threshold r be relevant?

3. Conclusion

Our venture in this thesis was mainly — if not exclusively — technical. With many results in probability spaces and one long completeness proof in chapter 7, one could accuse us of lacking convincing arguments as to why the theory of stable beliefs is worth dealing with in the first place.

However, arguing in favor of Leitgeb's theory was outside of the scope of our work. Instead, being convinced that Leitgeb's theory offers us both a philosophically intuitive and logically acceptable (or better yet not counter-intuitive and not unacceptable) quantitative probabilistic representation of belief, we aimed to provide a common conceptual framework that unifies Leitgeb's notion of stable belief, B-S's notion of strong belief and the notions of a priori, abnormal and conditional belief studied by Van Fraassen and Arló-Costa.

As we mentioned in the beginning of this thesis, there are turbulent waters between logic and probability. In [34, p. 1388] Leitgeb writes: "For now, we hope to have laid the foundations of what will hopefully become a valuable contribution to the "peace project" between logic and probability theory".

In turn, we hope that this thesis has strengthened these foundations.

Bibliography

[1] E. Adams. "Probability and the logic of conditionals". In: *Aspects of inductive logic* [1966], pp. 265–316.

[2] C. E. Alchourrón, P. Gärdenfors, and D. Makinson. "On the logic of theory change: Partial meet contraction and revision functions". In: *The journal of symbolic logic* 50.02 [1985], pp. 510–530.

[3] H. Arló-Costa. *Full Belief, Supposition, and Personal Probablility.* 1998.

[4] A. Baltag and S. Smets. "A qualitative theory of dynamic interactive belief revision". In: *Texts in logic and games* 3 [2008], pp. 9–58.

[5] A. Baltag and S. Smets. "Conditional doxastic models: A qualitative approach to dynamic belief revision". In: *Electronic Notes in Theoretical Computer Science* 165 [2006], pp. 5–21.

[6] A. Baltag and S. Smets. "Dynamic belief revision over multi-agent plausibility models". In: *Proceedings of LOFT*. Vol. 6. 2006, pp. 11–24.

[7] A. Baltag and S. Smets. "From conditional probability to the logic of doxastic actions". In: *Proceedings of the 11th conference on Theoretical aspects of rationality and knowledge*. ACM. 2007, pp. 52–61.

[8] A. Baltag and S. Smets. "Probabilistic dynamic belief revision". In: *Synthese* 165.2 [2008], pp. 179–202.

[9] A. Baltag and S. Smets. "The logic of conditional doxastic actions: A theory of dynamic multi-agent belief revision". In: *Proceedings of the Workshop on Rationality and Knowledge, ESSLLI, Malaga*. Citeseer. 2006.

[10] A. Baltag, S. Smets, and J. A. Zvesper. "Keep hoping for rationality: a solution to the backward induction paradox". In: *Synthese* 169.2 [2009], pp. 301–333.

[11] P. Battigalli and G. Bonanno. "Recent results on belief, knowledge and the epistemic foundations of game theory". In: *Research in Economics* 53.2 [1999], pp. 149–225.

[12] P. Battigalli and M. Siniscalchi. "Hierarchies of conditional beliefs and interactive epistemology in dynamic games". In: *Journal of Economic Theory* 88.1 [1999], pp. 188–230.

[13] P. Battigalli and M. Siniscalchi. "Strong belief and forward induction reasoning". In: *Journal of Economic Theory* 106.2 [2002], pp. 356–391.

[14] E. Ben-Porath. "Rationality, Nash equilibrium and backwards induction in perfect-information games". In: *The Review of Economic Studies* 64.1 [1997], pp. 23–46.

[15] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 2001.

[16] L Blume, A Brandenburger, and E Dekel. "Lexicographic probabilities and choice under uncertainty". In: *Econometrica* 59.1 [1991], pp. 61–79.

[17] L. Blume, A. Brandenburger, and E. Dekel. "Lexicographic probabilities and equilibrium refinements". In: *Econometrica: Journal of the Econometric Society* [1991], pp. 81–98.

[18] O. Board. "Dynamic interactive epistemology". In: *Games and Economic Behavior* 49.1 [2004], pp. 49–80.

[19] A. Brandenburger and E. Dekel. “Hierarchies of beliefs and common knowledge”. In: *Journal of Economic Theory* 59.1 [1993], pp. 189–198.

[20] A. Brandenburger, A. Friedenberg, and H. J. Keisler. “Admissibility in Games”. In: *Econometrica* 76.2 [2008], pp. 307–352.

[21] H. A. Costa and R. Parikh. “Conditional probability and defeasible inference”. In: *Journal of Philosophical Logic* 34.1 [2005], pp. 97–119.

[22] B. De Finetti. *Les probabilités nulles*. Gauthier-Villars, 1936.

[23] R. Fagin, J. Y. Halpern, M. Y. Vardi, and Y. Moses. *Reasoning about knowledge*. 1995.

[24] R. Foley. *Working without a net*. 1993.

[25] B. C. Fraassen. “Fine-grained opinion, probability, and the logic of full belief”. In: *Journal of Philosophical logic* 24.4 [1995], pp. 349–377.

[26] B. C. Fraassen. “Representational of conditional probabilities”. In: *Journal of Philosophical Logic* 5.3 [1976], pp. 417–430.

[27] P. Gärdenfors. *Knowledge in flux: Modeling the dynamics of epistemic states*. The MIT press, 1988.

[28] R. Goldbach. “Modelling Democratic Deliberation”. PhD thesis. Universiteit van Amsterdam, 2015.

[29] J. Halpern. “Lexicographic probability, conditional probability, and nonstandard probability”. In: *Games and Economic Behavior* 68.1 [2010], pp. 155–179.

[30] P. J. Hammond. “Consequentialism, Non-Archimedean Probabilities, and Lexicographic Expected Utility”. In: *The Logic of Strategy* [1999], p. 39.

[31] P. J. Hammond. “Elementary non-Archimedean representations of probability for decision theory and games”. In: *Patrick Suppes: scientific philosopher*. Springer, 1994, pp. 25–61.

[32] A. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitrechnung, Ergebnisse Der Mathematik (1933); translated as Foundations of Probability*. 1950.

[33] D. Lehmann and M. Magidor. “What does a conditional knowledge base entail?” In: *Artificial intelligence* 55.1 [1992], pp. 1–60.

[34] H. Leitgeb. “Reducing belief simpliciter to degrees of belief”. In: *Annals of Pure and Applied Logic* 164.12 [2013], pp. 1338–1389.

[35] H. Leitgeb. “The stability theory of belief”. In: *Philosophical Review* 123.2 [2014], pp. 131–171.

[36] L. E. Loeb. *Stability and justification in Hume’s treatise*. 2002.

[37] V. McGet. “Learning the impossible”. In: *Probability and conditionals: Belief revision and rational decision* [1994], p. 179.

[38] K. R. Popper. *Logik der Forschung (1934)*. 1968.

[39] A. Rényi. “On a new axiomatic theory of probability”. In: *Acta Mathematica Hungarica* 6.3-4 [1955], pp. 285–335.

[40] A. Robinson. “Function theory on some nonarchimedean fields”. In: *American Mathematical Monthly* [1973], pp. 87–109.

[41] J. Van Benthem. “Dynamic logic for belief revision”. In: *Journal of applied non-classical logics* 17.2 [2007], pp. 129–155.

[42] J. Van Benthem, J. Gerbrandy, and B. Kooi. “Dynamic update with probabilities”. In: *Studia Logica* 93.1 [2009], pp. 67–96.