KU Leuven
Humanities and Social Sciences Group
Institute of Philosophy

KU LEUVEN
DOCTORAL SCHOOL
HUMANITIES AND SOCIAL SCIENCES

BELIEVING IN LOGIC AND PHILOSOPHY

Lorenz DEMEY

Supervisor:
Prof. S. Cuypers

Dissertation presented in partial
fulfilment of the requirements for
the degree of Doctor in Philosophy

February 2014

Acknowledgments

Having arrived at the end of my PhD, I would like to take the time to thank several people. Many of them have directly helped me to improve the quality of the material presented here, while others have helped me or influenced my research in more indirect ways. I am almost certain that there are people whose name I have forgotten to mention here; I apologize to them in advance.

First of all, I would like to thank my supervisor, Stefaan Cuypers. He agreed to become my supervisor in quite unusual circumstances, and has supported me in every way I can imagine. I am particularly grateful for his sustained encouragement to explore topics that do not exactly fall under the classical conceptions of philosophy and philosophical logic, such as artificial intelligence and logical geometry. I would also like to thank the other members of the thesis committee: Jan Heylen, Leon Horsten, Chris Kelp and Bart Raymaekers. I am particularly happy to meet professor Horsten again, who was the supervisor of my master's thesis in philosophy back in 2008, and has now agreed to be the external member of my thesis committee.

I would like to thank the following professors, postdoctoral researchers and PhD students for making the KU Leuven's Institute of Philosophy such an interesting place to study and do research: Filip Buekens, Jake Chandler, Stefaan Cuypers, Andreas De Block, Helen De Cruz, Richard Dietz, Igor Douven, David Etlin, Russell Friedman, Chris Kelp, Olivier Lemeire, Christian Maes, Sebastian Sequoiah Grayson, Stijn Van Tongerloo and Roger Vergauwen. In particular, I would like to thank my office mates, Harmen Ghijsen and Jan Heylen, for all the lively discussions we have had over the years (philosophical and otherwise).

In the past five years, I have had the privilege of regularly visiting the Institute for Logic, Language and Computation in Amsterdam, first as a Master of Logic student, and then to give and attend talks in the LIRa seminar. In particular, I would like to thank the following people (a quick glance at the table of contents and bibliography of this thesis will immediately reveal the enormous influence they have had on my research): Alexandru Baltag, Johan van Benthem, Cédric Dégremont, Jan van Eijck, Dick de Jongh, Kohei Kishida, Benedikt Löwe, Ştefan Minică, Eric Pacuit, Alessandra Palmigiano, Joshua Sack, Sonja Smets, Sara Uckelman, Jouko Väänänen, Frank Veltman and Jonathan Zvesper.

Ever since I got interested in logical geometry, I have had the pleasure of discussing my ideas on this topic with the following people: Jean-Yves Béziau, Saloua Chatti, Dany Jaspers, Thierry Libert, Sebastian Löbner, Ingolf Max, Baptiste Mélès, Alessio Moretti, Henri Prade, Frédéric Sart, Fabien Schang, Hans Smessaert and Corina Strössner. In particular, I am grateful to Alessio Moretti for the numerous interesting discussions we have had over the past few years; although these discussions have rarely resulted in complete (or even partial) agreement between us, they have always been carried out in an atmosphere of friendship and mutual intellectual respect.

I would like to dedicate some special words of thanks to Hans Smessaert. In 2010, Hans and I discovered—almost by accident—that we share an interest in logical geometry. Hans taught me a lot about this topic, thereby deepening my interest in it. We found out that our academic backgrounds (linguistics/logic) are highly complementary and ideally suited for work in logical geometry, and therefore started working together. Over the years, this collaboration has become very intense (sometimes, we have multiple day-long meetings in a single week), and has turned out to be very fruitful (Chapter 9 presents a small fraction of our joint results). Finally, I would like to thank Hans for skillfully producing the figures that can be found in Chapters 8 and 9 of this thesis.

Over the years, I have presented my work at several conferences. I would like to thank the audiences of these conferences for their questions and suggestions. In particular, I have benefited a lot from discussions with Diderik Batens, Roberto Ciuni, Hans van Ditmarsch, Catarina Dutilh Novaes, Branden Fitelson, the late Paul Gochet, Valentin Goranko, Andreas Herzig, Wiebe van der Hoek, Andrea Kruse, Emiliano Lorini, Reinhard Muskens, Andrés Perea, Ahti-Veikko Pietarinen, Olivier Roy, Allard Tamminga, Jean Paul Van Bendegem, Heinrich Wansing, Sylvia Wenmackers and Stefan Wintein.

I would like to thank the members of the reading group on modal logic: Patrick Allo, Mathieu Beirlaen, Max Cresswell, Jonas De Vuyst, Giuseppe Primiero, Christian Strasser, Rafal Urbaniak, Frederik Van De Putte and Peter Verdée.

I would like to thank Alexandru Marcoci for co-organizing the Amsterdam reading group on dynamic epistemic logic. I would like to thank Jonas De Vuyst for co-organizing the PhDs in Logic III conference in 2011, and later co-editing its proceedings. I have learned a lot from all of them, not just academically speaking, but also with respect to practical skills.

Over the years, I have co-authored papers with Barteld Kooi, Joshua Sack and Hans Smessaert. These collaborations have been very diverse in format (from strictly Skype-based to 'sitting together at the keyboard'), and each of them has taught me a great deal about the process of writing a paper.

I would like to thank Erwin Blendeman, Stein De Cuyper, Ingrid Dierickx, Tanja Kassenaar, Sofie Keyaerts, Filip Ons, Dries Simons and Ines Van Houtte for their help in dealing with the administrative aspects of my PhD project.

I gratefully acknowledge financial support from the following projects and organizations: the NUFFIC Huygens Scholarship Programme (2009–2010), the Odysseus project on formal epistemology, led by Igor Douven (2009–2010), and a PhD fellowship ('aspirantschap') from the Fund for Scientific Research–Flanders (FWO) (2010–2014).

Ik zou mijn ouders, Alfons Demey en Simone Pittomvils, willen bedanken voor alles wat ze voor mij gedaan hebben, niet alleen tijdens mijn doctoraat, maar al mijn hele leven. Ik besef dat het niet evident was om te horen te krijgen dat ik filosofie wou gaan studeren (want wat kan je *daar* mee doen?), en later dat ik voor twee jaar naar het buitenland wou vertrekken voor een tweede studie (Amsterdam: zo dichtbij, maar toch ook zo veraf). Toch hebben zij mij in al deze beslissingen onvoorwaardelijk gesteund, in een periode die voor henzelf zeker niet de makkelijkste is geweest. Mijn dankbaarheid hiervoor is groter dan ik hen ooit duidelijk zal kunnen maken.

Finally, I would like to thank Margaux Smets. The ten years we have spent together are the most wonderful gift I have ever received from anyone. She has helped me throughout my PhD in countless ways, from proofreading several drafts of this thesis to dragging me through some of the most difficult moments in my life. I never cease to be amazed by her ability to persevere and stay positive, no matter what the circumstances. Most importantly, however, her endless love makes her the sole person in the world with whom I can truly be myself. I cannot hope to ever thank her enough for that.

Sources

Most of the material in this thesis has already been published, or is currently (December 2013) under review for publication. The list below states for each chapter the original publication(s) it is based on, and in the case of co-authored papers provides more details about authorship:

- Chapter 1 is based on Demey (2012c) and Demey (2013a).
- Chapter 2 is based on Demey et al. (2013) and Demey and Sack (forthcoming). The entire chapter was written primarily by myself, except for Section 2.5, which was written primarily by Kooi.
- Chapter 3 is based on Demey and Kooi (forthcoming). The entire chapter was written primarily by myself.
- Chapter 4 is based on Demey (2011a) and Demey (2012d).
- Chapter 5 is based on Demey (2011c) and Demey (forthcoming d).
- Chapter 6 is based on Demey (forthcoming a).
- Chapter 7 is based on Demey (2013b) and Demey (forthcoming b).
- Chapter 8 is based on Demey (2012c).
- Chapter 9 is based on Smessaert and Demey (2013b). The entire chapter was written jointly by Smessaert and myself; it is included in this thesis with his explicit consent.

Overview

Detailed Contents

III Dynamic Epistemic Logic and Logical Geometry 215

1 The Dynamic Turn in Epistemic Logic

The main topic of this thesis is the so-called dynamic turn in epistemic logic. In particular, I will argue that although this dynamic turn was initially mainly inspired by technical issues in computer science and game theory, it also has great *philosophical* relevance.

This introductory chapter is organized as follows. Section 1.1 provides a brief history of epistemic logic, focusing on the role and importance of the dynamic turn within this history. As an illustration of the dynamic turn, Section 1.2 discusses one of the simplest and most well-known systems of dynamic epistemic logic, viz. public announcement logic. Section 1.3 distinguishes between a weak and a strong interpretation of the dynamic turn. Section 1.4 presents the main results obtained in this thesis. Finally, Section 1.5 provides a chapter-per-chapter overview of the thesis, and indicates how the various chapters relate to each other and to the main line of argumentation.

1.1 A Brief History of Epistemic Logic

The origins. The starting point of 'modern' epistemic logic is commonly taken to be Hintikka's seminal *Knowledge and Belief* (1962), in which knowledge is treated as a modal operator.[1] We thus work with formulas of the form $K\varphi$, which intuitively means that the (unnamed) agent knows that φ. If there are several

[1] Many logical issues pertaining to epistemic and modal operators were already discussed in great detail, but of course informally, in medieval philosophy (Boh 1993, Knuuttila 1993, Boh 2000, Martens 2011, Uckelman 2011a). For a more comprehensive historical overview of epistemic logic, see Gochet and Gribomont (2006).

agents, we add agent indices to the K-operators; for example, $K_a\varphi \wedge \neg K_b\varphi$ means that agent a knows that φ, but agent b does not.

These operators are interpreted on Kripke models, which are triples of the form $\mathbb{M} = \langle W, R, V\rangle$, where W is a non-empty set of states or possible worlds, R is a binary relation over W, and V is a valuation function that specifies for each atomic proposition the states in which it is true. The relation R is the agent's *accessibility* relation; it specifies which states are compatible with her knowledge in any given state. For example, $(w, v) \in R$ means that state v is compatible with the agent's knowledge in state w; i.e. based on all the knowledge that she has in state w, the agent is unable to exclude that v might be the actual state. (If there are several agents, then for each agent i there is a distinct relation R_i in the Kripke model.) The most important semantic clause looks as follows:

$$\mathbb{M}, w \models K\varphi \quad \text{iff} \quad \text{for all } v \in W\text{: if } (w, v) \in R \text{ then } \mathbb{M}, v \models \varphi. \qquad (1.1)$$

Informally, this clause says that the agent knows that φ in a state w iff φ is true in all states that are accessible from w. Note that this clause should not be seen as a philosophically substantial *definition* of knowledge, since that would plainly result in circularity (the agent's knowledge in a state w is 'defined' in terms of the accessibility relation R, which is itself explicated in terms of (compatibility with) the agent's knowledge in w).

As a toy example, consider a model $\mathbb{M} = \langle W, R, V\rangle$ with just three possible worlds: $W = \{w, v, u\}$. In w it holds that Barack Obama is currently visiting Wisconsin, in v it holds that he is currently visiting Vermont, and finally, in u it holds that he is currently visiting Utah. We thus introduce three atomic propositions $\mathsf{visitWisconsin}$, $\mathsf{visitVermont}$ and $\mathsf{visitUtah}$, and specify the following valuation function V:

$$V(\mathsf{visitWisconsin}) = \{w\}, \;\; V(\mathsf{visitVermont}) = \{v\}, \;\; V(\mathsf{visitUtah}) = \{u\}.$$

Furthermore, the model specifies that $(w, w) \in R$ and $(w, v) \in R$, but $(w, u) \notin R$. This means that based on all the knowledge that she has in state w, the agent is able to exclude that u might be the actual state, but unable to exclude that w or v might be the actual state. Hence, she knows that Obama is not visiting Utah, but she does not know that he is visiting Wisconsin (although this actually happens to be true in w). By applying (1.1), we can check that these statements are actually true at the state w in the model $\mathbb{M}$; for example:

- $\mathbb{M}, w \models K\neg\mathsf{visitUtah}$,

- $\mathbb{M}, w \models \neg K\mathsf{visitWisconsin}$,
- $\mathbb{M}, w \models K(\mathsf{visitWisconsin} \vee \mathsf{visitVermont})$.

The fact that (1.1) cannot be taken as a definition of knowledge does not mean that it is philosophically useless. Its main advantage is that it allows us to establish correspondences (in a mathematically precise sense) between philosophical principles about knowledge on the one hand, and mathematical properties of the accessibility relation on the other. Typical examples of such philosophical principles include (i) *factivity* (if the agent knows that φ, then φ is true), (ii) *positive introspection* (if the agent knows that φ, then she knows that she knows this) and (iii) *negative introspection* (if the agent does not know that φ, then she knows that she does not know this). These philosophical principles can be shown to correspond to the mathematical properties of *reflexivity*, *transitivity* and *Euclideanness*, respectively. The table below provides a summary:[2]

	philosophical principle	mathematical property
(i)	$K\varphi \to \varphi$	$\forall w \in W\colon wRw$
(ii)	$K\varphi \to KK\varphi$	$\forall w, v, u \in W\colon \big(wRv \text{ and } vRu\big) \Rightarrow wRu$
(iii)	$\neg K\varphi \to K\neg K\varphi$	$\forall w, v, u \in W\colon \big(wRv \text{ and } wRu\big) \Rightarrow vRu$

Using correspondences such as these, one can argue for or against a philosophical principle about knowledge by arguing for or against the mathematical property of the accessibility relation that it corresponds to. For example, one can argue against positive introspection by pointing out that epistemic accessibility fails to be transitive. Let r, o and y be possible worlds in which the agent is seeing a piece of paper that is respectively red, orange and yellow. It might well be that because of her limited perceptual capacities, the agent is unable to distinguish between the red and orange pieces of paper, i.e. even if the piece of paper in front of her is actually red, the agent cannot exclude that it might be orange: $(r, o) \in R$. Similarly, she is unable to distinguish between the orange and yellow pieces of paper, i.e. even if the piece of paper in front of her is actually orange, the agent cannot exclude that it is might be yellow: $(o, y) \in R$. However, from $(r, o) \in R$ and $(o, y) \in R$ we cannot automatically conclude that also $(r, y) \in R$ (failure of transitivity of the accessibility relation R). After all, the visual contrast between the red and yellow pieces of paper might be so clear that the agent

[2] I will often write wRv instead of $(w, v) \in R$.

is in fact able to distinguish between them, i.e. if the piece of paper in front of her is actually red, the agent can exclude the possibility that it is yellow: $(r, y) \notin R$. It is easy to see that in this model, when the agent is presented with a red piece of paper, she will know that the paper is not yellow, but she will not know that she knows this (failure of positive introspection).

In a similar vein, one can also introduce a belief operator, and study various philosophical principles about the interaction between knowledge and belief. Typical examples include *doxastic-epistemic introspection* (if the agent believes that φ, then she knows that she believes this; formally: $B\varphi \to KB\varphi$) and *doxastic confidence* (if the agent believes that φ, then she believes that she knows this; formally: $B\varphi \to BK\varphi$). It can be proved that certain of these interaction principles (in combination with principles about the individual behavior of knowledge and belief) lead to a collapse of knowledge and belief, in the sense that $K\varphi$ and $B\varphi$ become equivalent (van der Hoek 1993, Halpern et al. 2009). Many philosophers maintain that knowledge and belief are distinct mental states, and hence, they cannot simultaneously accept all of the principles that lead to the collapse.

In summary, then, this line of work in epistemic logic is concerned with issues that are sometimes quite technical, but that are also philosophically relevant. This does not mean that the technical results themselves constitute philosophical positions, but rather that they can be usefully applied to conceptually elucidate various philosophical debates. Two of the grand syntheses of this line of work are Lenzen (1978) and Lenzen (1980).[3]

Applications in computer science and game theory. As time progressed, epistemic logic started being used more and more by computer scientists, game theorists, etc. These scholars make use of epistemic logic to formally analyze the intricate epistemic aspects of situations that frequently arise in their fields of study. Typical examples from multi-agent systems (artificial intelligence) are the so-called 'muddy children' puzzle, the 'sum and product' puzzle, and the 'Byzantine generals' puzzle (Halpern and Moses 1990, van Ditmarsch et al. 2007, 2008). Typical examples from protocol verification (cryptography) are the so-called 'dining cryptographers' puzzle and the 'Russian cards' puzzle (van Ditmarsch 2003, van der Meyden 2011, Pucella forthcoming). In game theory,

[3] Especially the latter is relevant for our current purposes, since it also studies the connection between epistemic notions and probabilistic notions, such as 'being 100% certain that φ'. Most of the systems developed in this thesis also have both an epistemic and a probabilistic component.

epistemic logic is typically used to analyze the epistemic conditions of solution concepts for various types of games, such as the backward induction solution for games of perfect information (Aumann 1995, Halpern 2001, Kaneko 2002, de Bruin 2008a, Baltag et al. 2009).

To illustrate these kinds of applications, I now describe the Byzantine generals and Russian cards scenarios[4] (without going into their formal analysis, of course):

Example 1.1 (Byzantine generals). Two army divisions are camped on two hilltops, overlooking the enemy city in the valley between them. It is commonly known by the divisions' generals that if both divisions attack the enemy simultaneously, they will certainly win the battle, while if only one division attacks, it will certainly lose the battle. The generals wish to coordinate a simultaneous attack somewhere the next day. They will only attack if it is common knowledge between them that both of them will attack. The generals can only communicate by means of a messenger; however, each time this messenger carries a message from one general to the other, he has to pass through the valley, and runs the risk of getting caught by the enemy—thus leaving the message undelivered. The first general sends the messenger to the second general, with the proposal to attack at 8AM. Because of the possibility of the messenger getting caught, the first general does not know whether his message actually reached the second general. To eliminate this uncertainty, the second general sends the messenger back to the first to confirm that he has indeed received the message. But of course, when going back from the second to the first general, the messenger again runs the risk of being caught by the enemy, etc. Assume that the messenger actually never gets caught (i.e. each message that is sent by one general actually reaches the other one). What will happen?[5]

Example 1.2 (Russian cards). Consider three agents: Ann, Bob, and Cath. There

[4] These examples might look far removed from real-world applications of computer science. However, they mainly function as small-scale, abstract and/or idealized scenarios to capture certain intuitions. In this sense, their role is similar to that of fake barns, cleverly disguised mules and the like in contemporary epistemology. In Demey (forthcoming c) I have suggested that the medieval philosopher William of Ockham already used such scenarios in his epistemological theorizing.

[5] For the interested reader: it can be shown that despite the messenger actually never getting caught, the generals do *not* attack. Each time the messenger successfully reaches the other general, an additional layer of knowledge is produced, but the generals never reach the common knowledge that is required to actually launch the attack.

are seven cards on the table, with the numbers 1 to 7 on them. The cards are turned such that the numbers are not visible. Ann and Bob each draw three cards, making sure that no one else sees which cards they have drawn. Cath gets the remaining card. For the sake of concreteness, let's say that Ann has cards 1, 2 and 3; Bob has cards 4, 5 and 6; and Cath has card 7. If all has gone well, each agent now knows which cards she has, but does not know which cards the two other players have.[6] Can Ann and Bob inform each other about which cards they have, *without* revealing any information to Cath? All communication has to be public, i.e. everything that Ann and Bob say to each other can be heard by Cath.[7]

It should be quite clear from these examples that this line of work in epistemic logic is concerned with topics that are of great importance in computer science, game theory, etc., but that have hardly any direct philosophical relevance. Some of the grand syntheses of the use of epistemic logic in computer science and game theory are Fagin et al. (1995), Meyer and van der Hoek (1995), de Bruin (2010) and Perea (2012).

The dynamic turn. Many of the applications of epistemic logic in game theory and computer science mentioned above involve not only reasoning about agents' knowledge at a single point in time, but also about how this knowledge changes over time. Hence, to adequately formalize these examples, we need systems of epistemic logic that are able to represent this epistemic dynamics. Unfortunately, however, the first systems of epistemic logic were all entirely *static*. For example, Hintikka (1962, pp. 7–8) explicity ruled out occasions

> on which people are engaged in gathering new factual information. Uttered on such an occasion, the sentences 'I don't know whether p' and [later] 'I know that p' need not be inconsistent.

The total absence of dynamics in the early systems of epistemic logic was already severely criticized by Scott (1970, p. 161):

[6] Except, of course, for the fact that if an agent has a certain card, then she trivially knows that the two other players do not have that card.

[7] For the interested reader: a protocol achieving the stated goals looks as follows (many others are possible). First, Ann says: 'The three cards I hold are either cards 1,2,3, or 1,4,5, or 1,6,7, or 2,4,6, or 3,5,7'; next, Bob says: 'Cath holds card 7'. Checking that this protocol indeed 'works' is exactly the task that protocol analysts use epistemic logic for.

> Here is what I consider one of the biggest mistakes of all in modal logic: concentration on a system with just *one* modal operator. The only way to have any philosophically significant results in [...] epistemic logic is to combine those operators with [e.g.] tense operators (otherwise how can you formulate principles of change?).

To be able to adequately analyze the scenarios they were interested in, computer scientists and game theorists thus had to develop new, *dynamic* systems of epistemic logic. These dynamic epistemic logics have proved to be highly successful tools for analyzing intricate epistemic scenarios. This line of work has therefore rapidly been expanding in the past few years, and is commonly referred to as the *dynamic turn in epistemic logic*.

It is important to realize that this dynamic turn in epistemic logic is part of a broader dynamic turn in logics of rational agency, and even in logic in general. For example, according to van Benthem (2003, p. 503),

> over the past decades computer science has also begun to influence the research agenda of logic. [...] modern logic is undergoing a Dynamic Turn, putting activities of inference, evaluation, belief revision or argumentation at centre stage.

A good example of this phenomenon can be seen in the field of preference logic, which models agents' preferences. In one of the earliest works on preference logic, von Wright (1963, p. 23) wrote:[8]

> The preferences which we shall study are a subject's [...] preferences on one occasion only. Thus we exclude [...] the possibility of *changes* in preferences. [von Wright's emphasis]

Von Wright's exclusion of preference changes stands in sharp contrast to contemporary preference logics, in which preference dynamics plays an important role (van Benthem and Liu 2007, Girard 2008, Liu 2011).

1.2 Public Announcement Logic

To illustrate the dynamic turn in epistemic logic, I will now informally discuss public announcement logic, which is by far the simplest system of dynamic epistemic logic, and which was historically the first such system to be studied in

[8] Note the striking similarity between von Wright's quotation and that of Hintikka given earlier.

detail (Plaza 1989, Gerbrandy and Groeneveld 1997). The discussion has intentionally been kept as informal as possible; technical details will be discussed extensively in the remainder of this thesis.

Recall that classical epistemic logic is *static*: it aims to describe the knowledge of one (or several) agent(s) at one point in time. In real life, however, an agent's knowledge can change over time. Consider the following three examples:

Example 1.3. Ann does not know that Paris is the capital of France. She is chatting with her best friend, Bob. During their conversation, Bob tells Ann that Paris is the capital of France. After this dialogue, Ann does know that Paris is the capital of France.

Example 1.4. Cath does not know that Paris is the capital of France. She is chatting with Bob (who knows that Paris is the capital of France). During their conversation, Bob tells Cath that Brussels is the capital of France. After this dialogue, Cath thinks that she knows that Brussels is the capital of France, but actually she does not know this.

Example 1.5. Fred and Tom are competing treasure hunters, looking for a particular treasure. Fred finds a map, thereby learning the treasure's exact location. A few miles away, Tom has a conversation with an old magician, who tells him the treasure's exact location. The day before they want to go dig up the treasure, they accidentally meet each other. There is a strange tension between them...

Example 1.3 is a straightforward case of a *learning* process (i.e. a transition from not-knowing to knowing) through communication. In Example 1.4, Bob is *deceiving* Cath: he knows a certain proposition to be false, yet still he communicates it to Cath. This deception leads Cath to think that she has acquired new knowledge (i.e. that she has gone through a learning process), but actually she hasn't. Finally, Example 1.5 illustrates the difference between *private* and *public* communication: Fred and Tom have both, through independent, private events, learned about the treasure's location. Hence, they both know the treasure's location, but they do not know of each other that they know it. When they meet each other, they do their best not to share their newly acquired knowledge (although both of them actually already possess it).

These examples show the variety and subtlety of dynamic epistemic phenomena. All of these phenomena (including deception, private communication, etc.) can be studied in the general system of dynamic epistemic logic, viz. product update logic (Baltag et al. 1998, Baltag and Moss 2004, Baltag and Smets 2008, van Benthem 2011). In this thesis, however, we will focus on (the epistemic and probabilistic aspects of) one particular type of epistemic dynamics, viz. *public announcements*.[9] Here is a typical example:

Example 1.6. Ann, Bob and Cath are all sitting together in the living room, watching television. The news starts, and the newsreader announces that there will be a train strike tomorrow. Ann, Bob and Cath now know that there will be a train strike tomorrow, and they start discussing the consequences this strike will have for them.

In this scenario, the newsreader's message is a public announcement. This particular type of dynamic phenomenon has at least three distinct features. First of all, the announcement is made by an *outside source* (which will not be explicitly represented in the logic). In the scenario, this outside source is the newsreader on television, not Ann, Bob or Cath. (Contrast this with Examples 1.3 and 1.4.) Second, the announcement is *truthful*: it is indeed true that there will be a train strike tomorrow. Only truths can be publicly announced. (Contrast this with Example 1.4.) Third, the announcement is *public*: afterwards, Ann, Bob and Cath not only know that there will be a train strike tomorrow, but they also know of each other that they know it. (Contrast this with Example 1.5.) This is clear from the fact that without any further ado, they start discussing the consequences of the strike. Ann, for example, does not need to tell Bob and Cath that there will be a strike; she knows that Bob and Cath were also watching television, and have thus heard the newsreader's announcement.

Public announcement logic captures this dynamics by means of a model-transforming operation. The initial situation is modeled by means of a Kripke model $\mathbb{M}$: this represents the state of the world (ontic information) and the agents' knowledge (epistemic information) before any announcement has been made. The public announcement of a formula φ transforms the model $\mathbb{M}$ into a new model $\mathbb{M}|\varphi$. This model transformation can be defined in such a way that if $\mathbb{M}$ accurately represents all the ontic and epistemic information of the initial situation (*before* the public announcement of φ), then $\mathbb{M}|\varphi$ accurately represents all the ontic and epistemic information of the terminal situation (*after* the public announcement of φ).

[9] There are three exceptions to this claim: Subsection 3.4 discusses a probabilistic extension of the general product update mechanism, Chapter 4 studies public announcements, but also so-called 'radical upgrades', and Subsection 8.3.3 focuses on the dynamic modal operators from propositional dynamic logic.

To be able to talk about this model transformation, a dynamic operator $[!\cdot]\cdot$ is introduced into the formal language. Intuitively, the formula $[!\varphi]\psi$ means that after a public announcement of φ (assuming it can be publicly announced at all), it will be the case that ψ. The parenthetical remark is necessary, since if φ happens to be false, then it cannot be publicly announced at all (recall the truthfulness of public announcements). The dual of $[!\varphi]\psi$ is $\langle !\varphi\rangle\psi \equiv \neg[!\varphi]\neg\psi$, which means that φ can actually be publicly announced (i.e. φ *is* true), and after this public announcement of φ, it will be the case that ψ. The semantics of these dynamic operators looks as follows:

$$\begin{array}{lll} \mathbb{M}, w \models [!\varphi]\psi & \text{iff} & \text{if } \mathbb{M}, w \models \varphi \text{ then } \mathbb{M}|\varphi, w \models \psi, \\ \mathbb{M}, w \models \langle !\varphi\rangle\psi & \text{iff} & \mathbb{M}, w \models \varphi \text{ and } \mathbb{M}|\varphi, w \models \psi. \end{array}$$

Hence, to evalutate $[!\varphi]\psi$ and $\langle !\varphi\rangle\psi$ at (a state w in) a model $\mathbb{M}$, we have to check the truth value of ψ at (the same state w in) the transformed model $\mathbb{M}|\varphi$. The conditional structure of the first clause expresses the truthfulness assumption of public announcements.

The exact definition of the updated model $\mathbb{M}|\varphi$ in terms of its 'ingredients' $\mathbb{M}$ and φ will be discussed later. For now, it suffices to note that this model transformation, together with the $[!\varphi]$- and $\langle !\varphi\rangle$-operators to describe it, allows us to formalize scenarios such as Example 1.6 in a highly compact and natural way. For example, if $\mathbb{M}$ represents the situation before the newsreader's announcement of the train strike, w is the actual state, and p expresses that there is a train strike tomorrow, then we have:

$$\mathbb{M}, w \models \neg K_a p \wedge \neg K_b p \wedge [!p]\big(K_a p \wedge K_b p \wedge K_a K_b p \wedge K_b K_a p\big).$$

This says that initially (i.e. before the newsreader's announcement), Ann and Bob do not know that there will be a train strike tomorrow, but after the newsreader's announcement, Ann and Bob *do* know that there will be a train strike tomorrow, and furthermore, they also know of each other that they know this.

1.3 Weak and Strong Interpretations of the Dynamic Turn in Epistemic Logic

In this section, I will distinguish between a weak and a strong interpretation of the dynamic turn, and argue that on the weak interpretation, the dynamic turn is hardly philosophically relevant, but that on the strong interpretation, it does turn out to be highly philosophically relevant. I will also discuss some case studies that illustrate the strong interpretation of the dynamic turn.

The *weak interpretation* of the dynamic turn in epistemic logic stays rather close to the historical facts. According to this interpretation, the primary—if not the *only*—use of dynamic epistemic logics is to analyze scenarios such as those described in Subsection 1.1, which originally inspired the development of these logics. These examples explicitly contain various types of epistemic dynamics; for example, in the Russian cards scenario (recall Example 1.2), it is quite clear that Ann and Bob's communication plays an important role, and should be modeled as a sequence of public announcements. As was explained in Subsection 1.1, the large majority of these examples come from computer science and game theory.[10]

On this interpretation, dynamic epistemic logic is severely restricted in scope: its use lies in analyzing scenarios that mainly (although not exclusively; recall Footnote 10) come from computer science and game theory. It should not be surprising that the results of these analyses—however useful they may be from the perspectives of computer science and game theory—will not be of great *philosophical* significance.

This skepticism about the philosophical importance of dynamic epistemic logics (and other logics that were historically primarily inspired by applications in computer science) can be found in many philosophers' attitudes toward these logics. For example, it seems to be at least implicitly present in the following remarks by Korte et al. (2009, p. 544):

> epistemic logic again became very popular because of the interest of computer scientists in the 1980s. A *philosophically more interesting development* than formal calculi may be in the speculation concerning objective and subject-bound quantification. [my emphasis]

In contrast, the *strong interpretation* of the dynamic turn in epistemic logic

[10] This is not to deny that dynamic epistemic logics have also been applied to explicitly dynamic issues that come from philosophy, rather than computer science or game theory. A good example is the formalization of the medieval theory of *obligationes* in dynamic epistemic logic (Uckelman 2011b,c, 2013). However, such philosophical applications are much more rare than applications to explicitly dynamic scenarios from computer science and game theory. Furthermore, historically speaking, these philosophical applications are much more recent, and certainly do not belong to the examples that originally motivated the development of dynamic epistemic logic.

maintains that dynamic epistemic logics can be used to analyze not only explicitly dynamic scenarios, but also scenarios, notions, theorems, etc. that *prima facie* look entirely *static*. After all, these analyses might reveal that underneath their static appearance, these cases contain a lot of dynamics.[11] Using dynamic epistemic logics, we can make this hidden dynamics fully explicit, and thus obtain more fine-grained conceptual analyses. For example, according to van Benthem (1996, p. 17),

> the motivation for standard logics often contains procedural elements present in textbook presentations — and one can make this implicit dynamics explicit.

This kind of conceptual elucidation is one of the primary uses of logic in philosophy. For example, in Subsection 1.1, I argued for the philosophical significance of the early work in epistemic logic by pointing out how this work has been used to clarify (the relations between) various philosophical notions and positions.[12] In *Must Do Better*, his self-proclaimed sermon on the current state of philosophy, Williamson (2007, pp. 288–291) makes similar comments when discussing the importance of logic—or mathematics in general—for philosophy:[13]

> How can we do better? We can make a useful start by getting the simple things right. Much even of analytic philosophy moves too fast in its haste to reach the sexy bits. Details are not given the care they deserve: crucial claims are vaguely stated, significantly different formulations are treated as though they were equivalent, examples are under-described, arguments are gestured at rather than

[11] Van Benthem (2011) explains this in terms of the distinction between 'process' and 'product': some notions (for example, logical proofs) seem to be *products* (finished derivations), but we should not ignore the *processes* that produced them (actually carrying out the derivations).

[12] For a more concrete example of the use of logic as a means of conceptual elucidation, consider the cases of epistemic arithmetic and modal-epistemic arithmetic. Epistemic arithmetic formalizes the notion of 'provability' by means of a single modal operator K (Shapiro 1985); modal-epistemic arithmetic disentangles this complex notion into the conceptually more primitive notions of 'proof' and 'possibility', which are represented by means of operators P and $\Diamond$, respectively. The statement 'it is provable that φ' is thus no longer formalized as $K\varphi$, but rather as $\Diamond P\varphi$ (Horsten 1993, 1994, Heylen 2013). One of the advantages of this conceptually more fine-grained analysis is that it allows us to investigate the $\Diamond$- and P-operators separately, and then study how the behavior of the 'composite' $\Diamond P$-operator arises out of their interactions.

[13] For a more systematic reflection on formal methods in philosophy, see Dutilh Novaes (2012).

> properly made, their form is left unexplained, and so on. [...] Philosophy can never be reduced to mathematics. But we can often produce mathematical models of fragments of philosophy and, when we can, we should. No doubt the models usually involve wild idealizations. It is still progress if we can agree what consequences an idea has in one very simple case.

On the strong interpretation of the dynamic turn, dynamic epistemic logics can thus be used as tools for conceptual elucidation, by revealing the essentially dynamic aspects of seemingly static notions, theorems, etc. To illustrate this, I will now briefly discuss how dynamic epistemic logics have recently been used to shed new light on two well-known and seemingly static issues: Fitch's paradox of knowability, and the problem of logical omniscience.

Fitch's knowability paradox states that, given some plausible assumptions about knowledge and possibility, the plausible claim that all true statements are *knowable* entails the highly implausible claim that all true statements are *in fact known* (Fitch 1963). This argument has received a great deal of attention from philosophers and logicians (Brogaard and Salerno 2009, Salerno 2009). At first sight, it does not seem to involve any dynamics, since it is concerned with claims about all true statements being knowable/known at a single time; this was emphasized by Fitch himself, who wrote that "the element of time will be ignored in dealing with these various concepts" (Fitch 1963, p. 135). Recently, however, there have been proposals to formalize the modal aspect of knowability not as a metaphysical modality, but rather as a dynamic modality: 'knowable' then no longer means 'known in some possible world', but rather 'known after some announcement' (van Benthem 2004, 2009, Balbiani et al. 2008). On this reading, the claim that all true statements are knowable turns out be false, despite its intuitive plausibility. A more positive result is that there exist exact (but highly non-trivial) characterizations of large classes of formulas that *are* knowable; interestingly, the so-called Moore sentence $p \wedge \neg Kp$ does *not* belong to these classes (Holliday and Icard III 2010).[14]

[14] Moore sentences play a central role in Fitch's argument, which has roughly the following structure: towards a contradiction, suppose that all true statements are knowable, but not all true statements are in fact known. Hence, there exists a statement p that is true but not in fact known, i.e. the Moore sentence $p \wedge \neg Kp$ is true. Furthermore, since *all* true statements are knowable, it follows, in particular, that this Moore sentence is knowable. Using some plausible principles about knowledge, this quickly leads to a contradiction.

The problem of logical omniscience is a problem for all systems of epistemic logic that take knowledge to be a normal modal operator. In such systems, agents are predicted to be logically omniscient: they know all tautologies, and their knowledge is closed under logical consequence. This is a serious problem, and it has even led authors such as Hocutt (1972) to argue that the entire project of epistemic logic is doomed to fail. It should thus not be surprising that there exists a wide variety of proposals to solve this problem (Sim 1997, Artemov and Kuznets 2006, Halpern and Pucella 2011). Many of these proposals come dangerously close to achieving the exact opposite of logical omniscience: the agents end up being logical idiots, who cannot perform even the simplest piece of logical reasoning. In reaction to this, several authors have proposed principles of *logical competence*: if an agent knows the premises of a long and tedious derivation, then she does not know its conclusion instantaneously, but she can come to know it by performing the necessary reasoning steps (Duc 1997, Ågotnes and Alechina 2007, Velázquez-Quesada 2009, van Benthem and Velázquez-Quesada 2010). By explicitly representing these types of inference dynamics, we are thus able to avoid the extremes of logical omniscience as well as logical idiocy.

Examples such as these illustrate that on its strong interpretation, the dynamic turn in epistemic logic is certainly philosophically relevant. It should be emphasized, however, that this philosophical relevance is of a strictly *methodological* nature (dynamic epistemic logic as a tool for conceptual elucidation), and is thus independent of any philosophical relevance that the particular topic under discussion might have. In other words, when dynamic epistemic logic is used to reveal dynamic aspects of some seemingly static notion X, the philosophical relevance of this analysis is to be situated in the clearer conceptual understanding of X that it affords us, rather than in the philosophical implications (if any) of X itself. For example, even if X is a technical notion from game theory that does not seem to have any philosophical significance, then showing that X actually has hitherto unknown dynamic aspects will lead to a clearer conceptual understanding of X, which constitutes genuine philosophical progress, if only in the philosophy *of game theory* (Grüne-Yanoff and Lehtinen 2012).

1.4 Main Results of the Thesis

The main claim of this thesis is that despite its origins in computer science and game theory, the dynamic turn in epistemic logic also has great philosophical significance. The overarching argument for this claim was presented in the previous section. In the remainder of the thesis, I will further develop this argument by providing various new applications of dynamic epistemic logic, and discussing their philosophical significance.

First of all, I will discuss three new examples of the strong interpretation of the dynamic turn, by showing how dynamic epistemic logics can be used to shed new light on seemingly static topics: Aumann's celebrated *agreeing to disagree* theorem (from game theory), the *Lockean thesis* about belief and degrees of belief (from epistemology), and the cognitive and epistemic aspects of *surprise* (from cognitive science). These three topics all have in common that they not only involve the agents' epistemic information, but also their probabilistic information. The dynamic behavior of this probabilistic component is non-trivially related to Bayesian conditionalization (as studied in Bayesian epistemology).

I will also discuss one application of dynamic epistemic logic that is not directly an illustration of the strong interpretation of the dynamic turn. This application lies in *logical geometry*, which can be described as the systematic study of the well-known Aristotelian square of oppositions and its various extensions, variants, etc. I will show how dynamic epistemic logic gives rise to non-trivial Aristotelian squares and larger diagrams, and discuss their importance for the philosophical foundations of logical geometry. These results constitute not only a further argument for the philosophical relevance of dynamic epistemic logic, but they are also prime examples of the recent stream of results on logical geometry (independently from its relation to dynamic epistemic logic). These results can be seen as providing the beginnings of an investigation called for by Kauffman (2001, p. 94):

> [there exists a] remarkable connection of polyhedral geometry with basic logic [. . .] One does not expect to find direct connections of the structure of logical speech with the symmetries of Euclidean Geometry. [. . .] The relationship of logic and geometry demands a deep investigation. This investigation is in its infancy.

To provide a representative overview of these recent findings in logical geometry, I will show how several apparently unrelated notions and theorems can be unified by viewing them from the common perspective of *information*.

1.5 Overview of the Thesis

The thesis is organized as follows.

Part I introduces the main topics of this thesis, viz. (epistemic) logic, probability, and dynamics. Since the three case studies on the dynamic turn all involve probabilities, it is important to get a clear view of the relationship between logic and probability theory. Chapter 2 provides a general overview of the various proposals to combine logic and probability, and shows that they can be organized in a systematical and logically meaningful way.

Next, Chapter 3 zooms in on one particular family of approaches, viz. probabilistic epistemic logics. These logics provide a standard epistemic (possible worlds) analysis of the agents' hard information, and supplement it with a fine-grained probabilistic analysis of their soft information. I introduce various dynamic extensions of these logics, and discuss the subtle relationship between public announcements and Bayesian conditionalization.

One of the case studies on the dynamic turn involves not only probabilistic Kripke models, but also epistemic plausibility models. In the literature, the latter are often defined in related, but subtly different ways. Therefore, Chapter 4 provides a detailed introduction to epistemic plausibility models and their model theory. I then use the model-theoretical results to argue that one way of defining these models is superior to all others, since it achieves a better equilibrium between philosophical applicability and mathematical elegance.

Part II presents the three case studies on (the strong interpretation of) the dynamic turn in epistemic logic. In Chapter 5, I discuss Aumann's celebrated agreeing to disagree theorem, and argue that Aumann's original formulation fails to fully capture the dynamics behind the agreement theorem (both in its formulation and in its semantic setup). I show how a more natural formulation of the theorem can be obtained in a system of probabilistic dynamic epistemic logic. Furthermore, I discuss how explicitly representing the dynamics behind the agreement theorem leads to a significant conceptual elucidation concerning the role of common knowledge in this theorem.

Chapter 6 discusses the Lockean thesis, which states that belief can be defined as 'sufficiently high degree of belief'. A well-known problem of this thesis is that it yields a notion of belief that is not closed under conjunction. After pointing out that this is a static problem, I examine how the Lockean thesis fares from a dynamic perspective. I compare the notion of belief as defined by the

Lockean thesis (interpreted on probabilistic Kripke models) with a ‘strictly qualitative’ notion of belief (interpreted on epistemic plausibility models), and show that both notions exhibit exactly the same dynamic behavior under public announcements. Finally, I argue that this technical observation supports the Lockean thesis, for methodological as well as philosophical reasons.

Chapter 7 discusses the epistemic and cognitive aspects of surprise. After providing a brief overview of existing work on surprise, I argue that the main formal accounts of surprise in logic and artificial intelligence fail to do justice to the essentially dynamic nature of surprise. I then propose a new formalization of surprise, using a system of probabilistic dynamic epistemic logic. I show that this system is able to capture several key aspects of surprise, such as its role in belief revision and its transitory nature. The former can also be captured by other formalizations, but the latter can only be adequately represented in the current system, since it is a manifestation of the dynamic nature of surprise.

Part III deals with logical geometry, both in its relation to the dynamic turn in epistemic logic and as an independent area of interest. In Chapter 8, I show how non-trivial Aristotelian squares and larger diagrams (such as hexagons, octagons, and rhombic dodecahedrons) can be constructed for dynamic epistemic logic, and for dynamic modalities in general. I also discuss the importance of these new diagrams for the philosophical foundations of logical geometry.

To illustrate the recent stream of results in logical geometry (independently from its relation to dynamic epistemic logic), I argue in Chapter 9 that Aristotelian diagrams can fruitfully be seen as being hybrid between two other types of diagrams, viz. opposition and implication diagrams. I develop an informativity perspective on all these types of diagrams, and use it to show that the Aristotelian square is strictly more informative than almost all other diagrams.

Finally, Chapter 10 summarizes the results obtained in this thesis, and assesses to what extent its main goals have been achieved.

Part I

Logic, Probability, and Dynamics

2 Combining Logic and Probability

2.1 Introduction

The very idea of combining logic and probability might look strange at first sight (Hájek 2001). After all, logic is concerned with absolutely certain truths and inferences, whereas probability theory deals with uncertainties. Furthermore, logic offers a *qualitative* (structural) perspective on inference (the deductive validity of an argument is based on the argument's formal structure), whereas probabilities are *quantitative* (numerical) in nature. However, as will be emphasized throughout this chapter, there are natural senses in which probability theory *presupposes* as well as *extends* classical logic. Furthermore, historically speaking, several distinguished theorists such as De Morgan (1847), Boole (1854), Ramsey (1990), de Finetti (1937), Carnap (1950), Jeffrey (1992) and Howson (2003, 2007, 2009) have emphasized the tight connections between logic and probability, or even considered their work on probability as a part of logic itself.[1]

This chapter has three main goals. The first goal is to roughly delineate the field: what exactly do we take to be the scope of terms such as 'probability logic' or 'probabilistic logic'? The second goal is to provide an overview of the major approaches to combining logic and probability theory. The third and final goal is to show that these various approaches can be organized in a systematic and logically meaningful way.[2]

[1] For more extensive historical overviews, see Hailperin (1988, 1991, 1996).

[2] Of course, each classification of a field as large and diverse as probability logic is bound to be arbitrary to some extent. Still, I believe that the classification introduced in this chapter—which is adapted from Demey et al. (2013)—captures some theoretically important distinctions, and can

The most common strategy to obtain a concrete system of probabilistic logic is to start with a classical system of logic and to 'probabilify' it in one way or another, by adding probabilistic features to it. There are various ways in which this probabilification can be implemented. One can study probabilistic semantics for classical languages (which do not have any explicit probabilistic operators), in which case the consequence relation itself gets a probabilistic flavor: deductive validity becomes 'probability preservation', rather than 'truth preservation'. Alternatively, one can add various kinds of probabilistic operators to the logic's object language. The first distinction to be made here is whether these probabilistic operators are qualitative or quantitative in nature. Qualitative (non-numerical) probabilistic operators include unary operators such as 'it is probable that . . . ' and binary operators such as '. . . is more probable than . . . '. As for quantitative probabilistic operators, I will distinguish between first-order and propositional operators. A typical first-order operator talks about the probability that a given individual (randomly selected from the domain) satisfies a certain predicate; in contrast, a typical propositional operator talks about the probability that a given proposition is true. Such propositional operators can be studied in isolation, but they are often studied together with other propositional operators (in particular, modal operators).

The remainder of this chapter is organized as follows. Section 2.2 delineates the scope of probability logic. Sections 2.3–2.6 closely correspond to the various subdivisions of the classification of probabilistic logics that was described above (also see Figure 2.1). Section 2.3 introduces systems that provide probabilistic semantics for a language that is itself fully classical (i.e. does not contain any probabilistic operators). Section 2.4 discusses some systems with qualitative probabilistic operators. Next, Sections 2.5 and 2.6 deal with first-order and propositional quantitative probabilistic operators, respectively. Finally, it should be emphasized that combinations of the latter with other propositional operators—in particular, modal (epistemic) operators—will *not* be discussed in this chapter. Such systems will become very important in the remainder of this thesis, and are therefore studied separately and in much more detail in Chapter 3. Finally, Section 2.7 wraps things up.

therefore serve as a guide for further research. For example, De Bona et al. (2013) further develop this classification and use it to study the expressivity and computational complexity of various probabilistic logics.

Figure 2.1: The classification of probabilistic logics discussed in this chapter.

1. no probabilistic operators in the logic's object language → Section 2.3
2. probabilistic operators in the logic's object language
    (a) qualitative probabilistic operators → Section 2.4
    (b) quantitative probabilistic operators
        i. first-order operators → Section 2.5
        ii. propositional operators
            - in isolation → Section 2.6
            - together with modal operators → Chapter 3

2.2 The Scope of Probability Logic

By integrating the complementary perspectives of qualitative logic and numerical probability theory, we obtain highly expressive accounts of inference. It should therefore come as no surprise that combinations of logic and probability have been fruitfully applied in all fields that study reasoning mechanisms, such as philosophy, artificial intelligence, linguistics, cognitive science and mathematics.[3] The downside of this cross-disciplinary popularity is that terms such as 'probability logic' are used by different researchers in different, non-equivalent ways. Therefore, before moving on to the actual discussion of the various approaches, I will first attempt to delineate the subject matter of this chapter.

The most important distinction is that between *probability logic* and *inductive logic*. Classically, an argument is said to be *(deductively) valid* if and only if it is impossible that its premises are all true, while its conclusion is false. In other words, deductive validity amounts to *truth preservation*: in a valid argument, the truth of the premises guarantees the truth of the conclusion. In some arguments, however, the truth of the premises does not fully guarantee the truth

[3] For example, according to Kowalski, "[i]ntegrating probability and logic is one of the most active areas of research in Artificial Intelligence today" (2011, p. 154), while Oaksford and Chater contend that "logic and probability have complementary and compatible roles in helping to explain human reasoning [...] both may be needed to fully understand human reasoning, normatively and psychologically" (2010, p. 22).

of the conclusion, but still renders it highly likely. A typical example looks as follows:

> *The first swan I saw was white*
> *The second swan I saw was white*
> ⋮
> *The 1000th swan I saw was white*
>
> ―――――――――――――――――
>
> *All swans are white*

Such arguments are studied in *inductive logic*, which makes extensive use of probabilistic notions, and is therefore considered by some authors to be related to probability logic. There is some discussion about the exact relation between inductive logic and probability logic, which is summarized in the introduction of Kyburg (1994). The dominant position, which is also adopted here, is that probability logic entirely belongs to deductive logic, and hence should not be concerned with inductive reasoning (Adams and Levine 1975). Still, most work on inductive logic falls within the 'probability preservation' approach, and is thus closely connected to the systems that will be discussed in Section 2.3.[4]

I will also steer clear of the philosophical debate over the exact nature of probability. The formal systems discussed here are compatible with all of the common interpretations of probability, but obviously, in concrete applications, certain interpretations of probability will fit more naturally than others. For example, the probabilistic models discussed in Section 2.6 are, by themselves, neutral about the nature of probability, but when they are used to describe the behavior of a physical system, they are typically interpreted in an objective way, whereas modeling multi-agent scenarios is accompanied most naturally by a subjective interpretation of probabilities (as agents' degrees of belief).[5]

Finally, although the success of probability logic is largely due to its various applications, I will not deal with these applications in any detail. For example, I will not assess the use of probability as a formal representation of belief in philosophy (Bayesian epistemology) or artificial intelligence (knowledge representation), and its advantages and disadvantages with respect to alternative

[4] Recent overviews of inductive logic can be found in Fitelson (2006), Romeijn (2011) and Hawthorne (2012).

[5] For more detailed discussions about the philosophical interpretation of probability, see Gillies (2000), Eagle (2010), Hájek (2011) and Childers (2013).

representations, such as generalized probability theory (for quantum theory) and fuzzy logic.[6]

2.3 Probabilistic Semantics for a Classical Language

In this section, I will present a first family of probability logics, which are used to study questions of 'probability preservation' (or dually, 'uncertainty propagation'). These systems do not extend the object language with any probabilistic operators, but rather deal with a 'classical' propositional language that only contains the usual (Boolean) connectives.

Definition 2.1. Let Prop be a countable set of atomic propositions. The language $\mathcal{L}(\mathsf{Prop})$ is defined by means of the following Backus-Naur form (BNF):

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

where $p \in \mathsf{Prop}$. Usually, I will simply write $\mathcal{L}$ instead of $\mathcal{L}(\mathsf{Prop})$. It is well-known that the other Boolean connectives can be defined in terms of $\wedge$ and $\neg$; for example, we define $\varphi \to \psi :\equiv \neg(\varphi \wedge \neg\psi)$ and $\varphi \vee \psi :\equiv \neg(\neg\varphi \wedge \neg\psi)$.

The main idea is that the premises of a valid argument can be uncertain, in which case (deductive) validity imposes no conditions on the (un)certainty of the conclusion. For example, the following argument is deductively valid (it is an instance of the well-known *modus ponens* argument scheme):

if it will rain tomorrow, I will get wet
it will rain tomorrow

I will get wet

However, if the argument's second premise is uncertain, its conclusion will typically also be uncertain. Probabilistic semantics represent such uncertainties as probabilities, and study how they 'flow' from the premises to the conclusion. In other words, we are not concerned with *truth preservation*, but rather with *probability preservation*. The following three subsections discuss systems that deal with increasingly more general versions of this issue.

[6] For more information about these topics, see Gerla (1994), Paris (1994), Halpern (2003), De Raedt et al. (2008), Vennekens et al. (2009), Hajek (2010), Hájek and Hartmann (2010), Hartmann and Sprenger (2010), Wilce (2012), Huber (2013) and Koons (2013).

2.3.1 A Basic Probabilistic Semantics

Let's begin by introducing the notion of a probability function for the propositional language $\mathcal{L}(\mathsf{Prop})$.[7]

Definition 2.2. A *probability function (for $\mathcal{L}$)* is a function $P\colon \mathcal{L} \to \mathbb{R}$ satisfying the following constraints:

- $P(\varphi) \geq 0$ for all $\varphi \in \mathcal{L}$, (non-negativity)
- if $\models \varphi$, then $P(\varphi) = 1$, (tautologies)
- if $\models \neg(\varphi \wedge \psi)$, then $P(\varphi \vee \psi) = P(\varphi) + P(\psi)$. (finite additivity)

Note that in the second and third constraint, the $\models$-symbol denotes (semantic) validity in classical propositional logic. The definition of probability functions thus requires notions from classical logic, and in this sense probability theory can be said to *presuppose* classical logic (Adams 1998, p. 22). It can easily be shown that if P satisfies these constraints, then $P(\varphi) \in [0, 1]$ for all formulas $\varphi \in \mathcal{L}$, and $P(\varphi) = P(\psi)$ for all formulas $\varphi, \psi \in \mathcal{L}$ that are logically equivalent (i.e. such that $\models \varphi \leftrightarrow \psi$). This shows that probability functions are essentially *semantic* entities: if φ and ψ are logically equivalent, they are merely two syntactically different ways of expressing the same proposition, and are thus assigned the same probability.

I now turn to the basic probabilistic semantics, as defined in Leblanc (1983). The argument with (a set of) premises Γ and conclusion φ will henceforth be denoted (Γ, φ).

Definition 2.3. The argument (Γ, φ) is *probabilistically valid*, written $\Gamma \models_p \varphi$, iff

$$\begin{gathered}\text{for all probability functions } P\colon \mathcal{L} \to \mathbb{R}\text{:}\\ \text{if } P(\gamma) = 1 \text{ for all } \gamma \in \Gamma\text{, then also } P(\varphi) = 1.\end{gathered}$$

Probabilistic semantics thus replaces the valuations $v\colon \mathcal{L} \to \{0, 1\}$ of classical propositional logic with probability functions $P\colon \mathcal{L} \to \mathbb{R}$, which take values

[7] In mathematics, probability functions are usually defined as measures on a σ-algebra of subsets of a given set (called the *sample space*); see Section 2.6. In logical contexts, however, it is often deemed more natural to define probability functions 'directly' on the logic's object language (Williamson 2002).

in the real unit interval $[0, 1]$. The classical truth values of *true* (1) and *false* (0) can thus be regarded as the endpoints of the unit interval $[0, 1]$, and likewise, valuations $v\colon \mathcal{L} \to \{0, 1\}$ can be regarded as degenerate probability functions $P\colon \mathcal{L} \to [0, 1]$. In this sense, classical logic is a special case of probability logic, or equivalently, probability logic is an *extension* of classical logic.

It can be shown that classical propositional logic (CPL) is (strongly) sound and complete with respect to probabilistic semantics:

$$\Gamma \models_p \varphi \text{ if and only if } \Gamma \vdash_{\mathsf{CPL}} \varphi.$$

Some authors interpret probabilities as generalized truth values (Reichenbach 1949, Leblanc 1983). According to this view, probability logic is just a particular kind of many-valued logic, and probabilistic validity boils down to 'truth preservation': truth (i.e. probability 1) carries over from the premises to the conclusion. Other logicians, such as Tarski (1936) and Adams (1998, p. 15), have noted that probabilities cannot be seen as generalized truth values, because probability functions are not 'compositional'; for example, $P(\varphi \wedge \psi)$ cannot be expressed as a function of $P(\varphi)$ and $P(\psi)$. More discussion on this topic can be found in Hailperin (1984).

Another possibility is to interpret a sentence's probability as a measure of its (un)certainty. For example, the sentence 'Jones is in Spain at the moment' is either true or false, but it can have any degree of certainty, ranging from 0 (maximal uncertainty) to 1 (maximal certainty).[8] According to this interpretation, the following theorem follows from the strong soundness and completeness of probabilistic semantics:

Theorem 2.1. *Consider a deductively valid argument* (Γ, φ). *If all premises in* Γ *have probability 1, then the conclusion* φ *also has probability 1.*

This theorem can be seen as a first, very partial clarification of the issue of probability preservation (or uncertainty propagation). It says that if there is no uncertainty whatsoever about the premises, then there cannot be any uncertainty about the conclusion either. In the next two subsections, I will consider more interesting cases, in which there is non-zero uncertainty about the premises, and ask how it carries over to the conclusion.

[8] Note that 0 is actually a kind of certainty, viz. certainty about falsity. However, in this chapter, I follow Adams' terminology and interpret 0 as maximal uncertainty (Adams 1998, p. 31).

Finally, it should be noted that although this subsection only discussed probabilistic semantics for classical propositional logic, there are also probabilistic semantics for a variety of other logics, such as intuitionistic propositional logic (van Fraassen 1981, Morgan and Leblanc 1983), modal logics (Morgan 1982a,b, 1983, Cross 1993), classical first-order logic (Leblanc 1979, 1983, van Fraassen 1981), relevant logic (van Fraassen 1983) and nonmonotonic logic (Pearl 1991). All of these systems share a key feature: the logic's semantics is probabilistic in nature, but probabilities are *not* explicitly represented in the object language; hence, they are much closer in nature to the system discussed here than to the systems presented in later sections.

Most probabilistic semantics for non-classical logics are not based on unary probabilities $P(\varphi)$, but rather on conditional probabilities $P(\varphi \mid \psi)$. The conditional probability $P(\varphi \mid \psi)$ is taken as primitive (rather than being defined as $\frac{P(\varphi\wedge\psi)}{P(\psi)}$, as is usually done) to avoid problems when $P(\psi) = 0$. Goosens (1979) provides an overview of various axiomatizations of probability theory in terms of such primitive notions of conditional probability.

2.3.2 Adams' Probability Logic

In the previous subsection I discussed a first principle of probability preservation, which says that if all premises have probability 1, then the conclusion also has probability 1. Of course, more interesting cases arise when the premises are less than absolutely certain. Consider the valid argument with premises $p \vee q$ and $p \rightarrow q$, and conclusion q.[9] One can easily show that

$$P(q) = P(p \vee q) + P(p \rightarrow q) - 1.$$

In other words, if one knows the probabilities of the argument's premises, then one can calculate the exact probability of its conclusion, and thus provide a complete answer to the question of probability preservation for this particular argument. For example, if $P(p \vee q) = 0.6$ and $P(p \rightarrow q) = 0.8$, then $P(q) = 0.4$. In general, however, it will not be possible to calculate the *exact* probability of the conclusion, given the probabilities of the premises.

Example 2.1. Consider the valid argument with premises p and $p \rightarrow q$, and conclusion q. Furthermore, consider the probability functions P_1 and P_2, which assign the following values to the Boolean combinations of p and q:

[9] Recall that the symbol '$\rightarrow$' denotes the truth-conditional material conditional.

φ	$P_1(\varphi)$	$P_2(\varphi)$
$p \wedge q$	0.2	0.2
$p \wedge \neg q$	0.3	0.3
$\neg p \wedge q$	0.1	0.4
$\neg p \wedge \neg q$	0.4	0.1

It is easy to check that $P_1(p) = P_2(p) = 0.5$ and $P_1(p \to q) = P_2(p \to q) = 0.7$, while $P_1(q) = 0.3 \neq 0.6 = P_2(q)$.

This example shows that the probabilities of the premises do *not* uniquely determine the probability of the conclusion. Hence, the best we can hope for in general is a (tight) upper and/or lower *bound* for the conclusion's probability. I will now discuss Adams' (1998) methods to compute such bounds.

Adams' results can be stated more easily in terms of *uncertainty* rather than *certainty* (probability). Given a probability function $P\colon \mathcal{L} \to [0,1]$, the corresponding *uncertainty function* U_P is defined as

$$U_P : \mathcal{L} \to [0,1]\colon \varphi \mapsto U_P(\varphi) := 1 - P(\varphi).$$

If the probability function P is clear from the context, I will often simply write U instead of U_P. In the remainder of this subsection (and in the next one as well), it will be assumed that all arguments have only finitely many premises (which is not a significant restriction, given the compactness property of classical propositional logic). Adams' first main result, which was originally established by Suppes (1965), can now be stated as follows:

Theorem 2.2. *Consider a valid argument* (Γ, φ) *and a probability function* P*. Then the uncertainty of the conclusion* φ *cannot exceed the sum of the uncertainties of the premises* $\gamma \in \Gamma$*. Formally:*

$$U(\varphi) \leq \sum_{\gamma \in \Gamma} U(\gamma).$$

First of all, note that this theorem subsumes Theorem 2.1 as a special case: if $P(\gamma) = 1$ for all $\gamma \in \Gamma$, then $U(\gamma) = 0$ for all $\gamma \in \Gamma$, so $U(\varphi) \leq \sum U(\gamma) = 0$, and thus $P(\varphi) = 1$. Furthermore, note that the upper bound on the uncertainty of the conclusion depends on $|\Gamma|$, i.e. on the number of premises. If a valid argument has a small number of premises, each of which only has a small uncertainty (i.e. a high certainty), then its conclusion will also have a reasonably small uncertainty (i.e. a reasonably high certainty). Conversely, if a valid argument has premises with small uncertainties, then its conclusion can only be highly uncertain if the argument has a large number of premises.[10] To put the matter more concretely, note that if a valid argument has three premises which each have uncertainty 0.1, then adding a premise which also has uncertainty 0.1 will not influence the argument's validity, but it *will* raise the upper bound on the conclusion's uncertainty from 0.3 to 0.4—thus allowing the conclusion to be more uncertain than was originally the case. Finally, the upper bound provided by Theorem 2.2 is *optimal*, in the sense that (under the right conditions) the uncertainty of the conclusion can coincide with its upper bound:

Theorem 2.3. *Consider a valid argument* (Γ, φ)*, and assume that the premise set* Γ *is consistent, and that every premise* $\gamma \in \Gamma$ *is relevant (i.e.* $\Gamma - \{\gamma\} \not\models \varphi$*). Then there exists a probability function* $P\colon \mathcal{L} \to [0,1]$ *such that*

$$U_P(\varphi) = \sum_{\gamma \in \Gamma} U_P(\gamma).$$

The upper bound provided by Theorem 2.2 can also be used to define a probabilistic notion of validity. An argument (Γ, φ) is said to be *Adams-probabilistically valid*, written $\Gamma \models_a \varphi$, if and only if

for all probability functions $P\colon \mathcal{L} \to \mathbb{R}$: $U_P(\varphi) \leq \sum_{\gamma \in \Gamma} U_P(\gamma)$.

Adams-probabilistic validity has an alternative, equivalent characterization in terms of probabilities rather than uncertainties. This characterization says that the argument (Γ, φ) is Adams-probabilistically valid if and only if the conclusion's probability can get arbitrarily close to 1 if the premises' probabilities are sufficiently high. Formally: $\Gamma \models_a \varphi$ if and only if

for all $\varepsilon > 0$ there exists a $\delta > 0$ such that for all probability functions P:
if $P(\gamma) > 1 - \delta$ for all $\gamma \in \Gamma$, then $P(\varphi) > 1 - \varepsilon$.

It can be shown that classical propositional logic is (strongly) sound and complete with respect to Adams' probabilistic semantics:

$\Gamma \models_a \varphi$ if and only if $\Gamma \vdash_{\mathsf{CPL}} \varphi$.

[10] A famous illustration of this converse principle is the *lottery paradox*, which was introduced by Kyburg (1965) and is also discussed in Sorensen (2011). I will briefly touch upon the epistemic significance of the lottery paradox in Chapter 6.

Adams (1998, p. 154) also defines another logic which is sound and complete with respect to his probabilistic semantics. However, this system involves a non-truth-functional connective (the *probability conditional*), and therefore falls outside the scope of this section.[11]

Example 2.2. The argument A with premises p, q, r, s and conclusion $p \wedge (q \vee r)$ is valid. Assume that $P(p) = 0.9, P(q) = P(r) = 0.8$ and $P(s) = 0.6$. Then Theorem 2.2 says that

$$U(p \wedge (q \vee r)) \leq 0.1 + 0.2 + 0.2 + 0.4 = 0.9.$$

This upper bound on the uncertainty of the conclusion is rather disappointing, and it exposes the main weakness of Theorem 2.2. One of the reasons why the upper bound is so high is that its calculation takes into account the premise s, which has a rather high uncertainty (0.4). However, this premise is irrelevant, in the sense that the conclusion already follows from the other three premises. In other words, we can regard $p \wedge (q \vee r)$ not only as the conclusion of the valid argument A, but also as the conclusion of the (equally valid) argument A', which has premises p, q, r. In the latter case, Theorem 2.2 yields an upper bound of $0.1 + 0.2 + 0.2 = 0.5$, which is already much lower.

The weakness of Theorem 2.2 is thus that it takes into account (the uncertainty of) *irrelevant* or inessential premises. To obtain an improved version of this theorem, a more fine-grained notion of 'essentialness' is necessary. In argument A in Example 2.2, premise s is absolutely irrelevant. Similarly, premise p is absolutely relevant, in the sense that without this premise, the conclusion $p \wedge (q \vee r)$ is no longer derivable. Finally, the premise subset $\{q, r\}$ is 'in between': together q and r are relevant (if both premises are left out, the conclusion is no longer derivable), but each of them separately can be left out (while keeping the conclusion derivable).

The notion of essentialness is formalized as follows:

Definition 2.4. Given a valid argument (Γ, φ) and a premise $\gamma \in \Gamma$,

- a premise set $\Gamma' \subseteq \Gamma$ is *essential* iff $\Gamma - \Gamma' \not\models \varphi$,

[11] More information about probabilistic interpretations of conditionals can be found in Eells and Skyrms (1994), Edgington (2006) and Arló Costa (2007).

- the *degree of essentialness* of γ, written $E(\gamma)$, is defined as
$$E(\gamma) := \frac{1}{|S_\gamma|},$$
where $|S_\gamma|$ is the cardinality of the smallest essential premise set that contains γ. If γ does not belong to any minimal essential premise set, then the degree of essentialness of γ is 0.

With these definitions, a refined version of Theorem 2.2 can be established:

Theorem 2.4. *Consider a valid argument* (Γ, φ)*. Then the uncertainty of the conclusion* φ *cannot exceed the weighted sum of the uncertainties of the premises* $\gamma \in \Gamma$*, with the degrees of essentialness as weights. Formally:*

$$U(\varphi) \leq \sum_{\gamma\in\Gamma} E(\gamma)U(\gamma).$$

The proof of Theorem 2.4 is significantly more difficult than that of Theorem 2.2: Theorem 2.2 requires only basic probability theory, whereas Theorem 2.4 is proved using methods from linear programming (Goldman and Tucker 1956, Adams and Levine 1975). Theorem 2.4 subsumes Theorem 2.2 as a special case: if all premises are relevant (i.e. have degree of essentialness 1), then Theorem 2.4 yields the same upper bound as Theorem 2.2. Furthermore, Theorem 2.4 does not take into account irrelevant premises (i.e. premises with degree of essentialness 0) to compute the upper bound; hence, if a premise is irrelevant for the validity of the argument, then its uncertainty will not carry over to the conclusion. Finally, note that since $E(\gamma) \in [0, 1]$ for all $\gamma \in \Gamma$, it holds that

$$\sum_{\gamma\in\Gamma} E(\gamma)U(\gamma) \leq \sum_{\gamma\in\Gamma} U(\gamma),$$

i.e. Theorem 2.4 yields in general a tighter upper bound than Theorem 2.2. To illustrate this, consider again the argument from Example 2.2, which has premises p, q, r, s and conclusion $p \wedge (q \vee r)$, and recall that $P(p) = 0.9, P(q) = P(r) = 0.8$ and $P(s) = 0.6$. One can calculate the degrees of essentialness of the premises: $E(p) = 1, E(q) = E(r) = 0.5$ and $E(s) = 0$. Hence, Theorem 2.4 yields that

$$U(p \wedge (q \vee r)) \leq (1 \times 0.1) + (0.5 \times 0.2) + (0.5 \times 0.2) + (0 \times 0.4) = 0.3,$$

which is a tighter upper bound for the uncertainty of $p \wedge (q \vee r)$ than any of the bounds obtained above via Theorem 2.2 (viz. 0.9 and 0.5).

2.3.3 Further Generalizations

Given the uncertainties (and degrees of essentialness) of the premises of a valid argument, Adams' theorems allow us to compute an *upper* bound for the uncertainty of the conclusion. Of course, these results can also be expressed in terms of probabilities rather than uncertainties; they then yield a *lower* bound for the probability of the conclusion. For example, when expressed in terms of probabilities rather than uncertainties, Theorem 2.4 looks as follows:

$$P(\varphi) \geq 1 - \sum_{\gamma \in \Gamma} E(\gamma)(1 - P(\gamma)).$$

Adams' results are restricted in at least two ways. Their first limitation is that they only provide a *lower* bound for the probability of the conclusion (given the probabilities of the premises). In a sense, this is the most important bound: it represents the conclusion's probability in the 'worst-case scenario', which might be useful information in practical applications. However, in some applications it might also be informative to have an *upper* bound for the conclusion's probability. For example, if one knows that this probability has an upper bound of 0.1, then one might rationally make another decision than one would have made if the probability's upper bound had been 0.9. I will now describe such a case, using the terminology of decision theory (Peterson 2009):

Example 2.3. An agent has to decide between walking to work vs. taking the bus. Taking the bus requires buying a ticket, whereas walking is free. However, if it starts raining while she is walking, she will be wet when arriving; if she takes the bus, she will not get wet. The utilities of the two actions are thus dependent on whether or not it starts raining; for example, they might sensibly look as follows:

	rain	*no rain*
walk	1	5
bus	2	2

Let p be the probability that it will rain (the probability that it will not rain is thus $1 - p$). The actions' expected utilities are as follows:

$$\begin{array}{lclcl} EU(\mathsf{walk}) & = & 1 \times p \; + \; 5 \times (1 - p) & = & -4p + 5, \\ EU(\mathsf{bus}) & = & 2 \times p \; + \; 2 \times (1 - p) & = & 2. \end{array}$$

Now suppose that the agent has only partial information about the probability of rain: she does not know the actual value of p, but only an upper bound for it. Consider the following two cases:

1. the upper bound is 0.9, i.e. $p \leq 0.9$: the expected utility of walking to work then has the following upper bound:

$$EU(\mathsf{walk}) \;=\; -4p+5 \;\leq\; -4\times 0.9 \;+\; 5 \;=\; 1.4,$$

2. the upper bound is 0.1, i.e. $p \leq 0.1$: the expected utility of walking to work then has the following upper bound:

$$EU(\mathsf{walk}) \;=\; -4p+5 \;\leq\; -4\times 0.1 \;+\; 5 \;=\; 4.6.$$

The expected utility of taking the bus is 2, independently of the value of p. The agent chooses the action that maximizes her expected utility. In the first case, she will certainly choose to take the bus, since $EU(\mathsf{walk}) \leq 1.4 < 2 = EU(\mathsf{bus})$. In the second case, however, things are not so clear. After all, all she knows for sure is that $EU(\mathsf{walk}) \leq 4.6$, which can be divided into the subcases $EU(\mathsf{walk}) \in [0,2]$ and $EU(\mathsf{walk}) \in (2,4.6]$. Note that in the first subcase, it holds that

$$EU(\mathsf{walk}) \leq 2 = EU(\mathsf{bus}),$$

while in the second subcase it holds that

$$EU(\mathsf{bus}) = 2 < EU(\mathsf{walk}).$$

Hence, which action maximizes expected utility (and should thus be chosen by the agent) depends on which subcase actually obtains. In the absence of any other information, it is rational to assume that the first subcase obtains with probability $\frac{2}{4.6} \approx 0.43$, and the second one with probability $\frac{2.6}{4.6} \approx 0.57$. Hence, the second subcase is more likely to occur, so the agent should choose to walk to work.

In sum, then, in case the upper bound on p is 0.9, the agent should definitely choose to take the bus, but in case this upper bound is 0.1, the agent can rationally choose to walk to work. Hence, this example clearly shows that the upper bounds on an agent's probabilities might influence her decisions.

The second limitation on Adams' results is that they presuppose that the premises' exact probabilities are known. In practical applications, however, there might only be partial information about the probability of a premise γ: its exact value is not known, but it *is* known to have a lower bound a and an upper bound b (Walley 1991). In such applications, it would be useful to have a method to calculate (optimal) lower and upper bounds for the probability of the conclusion in terms of the lower and upper bounds of the probabilities of the premises.

Hailperin (1965, 1984, 1986, 1996) and Nilsson (1986) use methods from linear programming to show that these two restrictions can be overcome. Their most important result is the following:

Theorem 2.5. *Consider an argument* (Γ, φ)*, with* $|\Gamma| = n$*. There exist functions* $L_{\Gamma,\varphi} : \mathbb{R}^{2n} \to \mathbb{R}$ *and* $U_{\Gamma,\varphi} : \mathbb{R}^{2n} \to \mathbb{R}$ *such that for any probability function* P*, the following holds: if* $a_i \leq P(\gamma_i) \leq b_i$ *for* $1 \leq i \leq n$*, then:*

1. $L_{\Gamma,\varphi}(a_1, \dots, a_n, b_1, \dots, b_n) \leq P(\varphi) \leq U_{\Gamma,\varphi}(a_1, \dots, a_n, b_1, \dots, b_n)$.
2. *The bounds in item 1 are optimal, in the sense that there exist probability functions* P_L *and* P_U *such that* $a_i \leq P_L(\gamma_i), P_U(\gamma_i) \leq b_i$ *for* $1 \leq i \leq n$*, and* $L_{\Gamma,\varphi}(a_1, \dots, a_n, b_1, \dots, b_n) = P_L(\varphi)$ *and* $P_U(\varphi) = U_{\Gamma,\varphi}(a_1, \dots, a_n, b_1, \dots, b_n)$.
3. *The functions* $L_{\Gamma,\varphi}$ *and* $U_{\Gamma,\varphi}$ *are effectively determinable from the Boolean structure of the sentences in* $\Gamma \cup \{\varphi\}$.

Proof. I will now give a proof sketch of this theorem.[12] The main idea is to transform the problem of finding optimal lower and upper bounds for the probability of φ to a linear programming problem (the same strategy is used by Adams to prove Theorem 2.4). Since we are only concerned with finitely many sentences $\gamma_1, \dots, \gamma_n, \varphi$, we can restrict ourselves to a sublanguage $\mathcal{L}(p_1, \dots, p_m) \subseteq \mathcal{L}(\mathsf{Prop})$, which contains only formulas based on the propositional atoms $p_1, \dots, p_m$. Consider the sentences σ_j, which are defined as follows:

$$\sigma_j := \pm p_1 \wedge \cdots \wedge \pm p_m.$$

[12] Many ideas found in this proof sketch will return in subsequent chapters. For example, the idea of transforming some conditions on probabilities into a linear programming problem is also central to the completeness proofs of the probabilistic epistemic logic described in Chapter 3 and its various extensions developed in Part II. Furthermore, the notion of a state description will be used extensively in Chapter 9.

(We write $+p_i$ for p, and $-p_i$ for $\neg p_i$.) Adams (1998) calls these sentences *state descriptions*, after Carnap (1947). It is clear that there are 2^m such sentences. They have the following logical properties:

$$\begin{array}{ll} \vdash_{\mathsf{CPL}} \bigvee_{j=1}^{2^m} \sigma_j, & \\ \vdash_{\mathsf{CPL}} \neg(\sigma_j \wedge \sigma_k) & \text{for } 1 \leq j \neq k \leq 2^m, \\ \vdash_{\mathsf{CPL}} \alpha \leftrightarrow \bigvee_{\sigma \vdash_{\mathsf{CPL}} \alpha} \sigma & \text{for all } \alpha \in \mathcal{L}(p_1, \dots, p_m). \end{array}$$

One can then easily show the following for any probability function $P\colon \mathcal{L}(p_1, \dots, p_m) \to [0, 1]$ and formula $\alpha \in \mathcal{L}(p_1, \dots, p_m)$:

$$\begin{array}{rclcl} \sum_{j=1}^{j=2^m} P(\sigma_j) & = & P\left(\bigvee_{j=1}^{2^m} \sigma_j\right) & = & 1, \\ P(\alpha) & = & P\left(\bigvee_{\sigma \vdash_{\mathsf{CPL}} \alpha} \sigma\right) & = & \sum_{\sigma \vdash_{\mathsf{CPL}} \alpha} P(\sigma). \end{array}$$

Let's use variables x_j to represent $P(\sigma_j)$ (for $1 \leq j \leq 2^m$). The problem of finding a least upper bound (resp. a greatest upper bound) for $P(\varphi)$ given that $a_i \leq P(\gamma_i) \leq b_i$ (for $1 \leq i \leq n$) can now be reformulated as the problem of finding the maximal (resp. minimal) value of the expression

$$\sum_{\sigma_j \vdash_{\mathsf{CPL}} \varphi} x_j \tag{2.1}$$

subject to the following constraints:

$$\begin{cases} a_i \leq \sum_{\sigma_j \vdash_{\mathsf{CPL}} \gamma_i} x_j & \text{for } 1 \leq i \leq n, \\ b_i \geq \sum_{\sigma_j \vdash_{\mathsf{CPL}} \gamma_i} x_j & \text{for } 1 \leq i \leq n, \\ \sum_{j=1}^{2^m} x_j = 1, & \\ x_j \geq 0 & \text{for } 1 \leq j \leq 2^m. \end{cases} \tag{2.2}$$

Note that expression (2.1) and the conditions in (2.2) are linear in the variables x_j. The remainder of the proof consists in applying methods from linear programming to solve these linear optimization problems. □

This result can also be used to define yet another probabilistic notion of validity, which I will call *Hailperin-probabilistic validity* or simply *h-validity*. This notion is not defined with respect to formulas, but rather with respect to *pairs* consisting of a formula and a subinterval of $[0,1]$. If X_i is the interval associated with premise $\gamma_i \in \Gamma$ and Y is the interval associated with the conclusion φ, then the argument (Γ, φ) is said to be *h*-valid, written $\Gamma \models_h \varphi$, if and only if

$$\text{for all probability functions } P\text{: if } P(\gamma_i) \in X_i \text{ for } 1 \leq i \leq n, \text{ then } P(\varphi) \in Y.$$

In Haenni et al. (2011) this is written as

$$\gamma_1^{X_1}, \ldots, \gamma_n^{X_n} \mathrel{|\!\!\approx} \varphi^Y$$

and called the *standard probabilistic semantics*.

Nilsson's work on probabilistic logic (1986, 1993) has sparked a lot of research on probabilistic reasoning in artificial intelligence (Hansen and Jaumard 2000, Haenni et al. 2011). However, it should be noted that although Theorem 2.5 states that the functions $L_{\Gamma,\varphi}$ and $U_{\Gamma,\varphi}$ are *effectively* determinable from the sentences in $\Gamma \cup \{\varphi\}$, the *computational complexity* of this problem is quite high (Georgakopoulos et al. 1988, Kavvadias and Papadimitriou 1990), and thus finding these functions quickly becomes computationally unfeasible in real-world applications. Contemporary approaches based on probabilistic argumentation systems and probabilistic networks are better capable of handling these computational challenges. Furthermore, probabilistic argumentation systems are closely related to Dempster-Shafer theory (Dempster 1968, Shafer 1976, Haenni and Lehmann 2003). However, an extended discussion of these approaches is beyond the scope of this chapter; a recent survey can be found in Haenni et al. (2011).

2.4 Qualitative Probabilistic Operators

In the remaining sections of this chapter, I will discuss systems that add probabilistic operators to the object language (instead of providing a probabilistic semantics for an object language which itself stays fully classic). In this section, the focus is on qualitative (i.e. non-numerical) operators.

There are several applications in which qualitative theories of probability might be useful, or even necessary. In some situations there are no frequencies available to use as estimates for the probabilities, or it might be practically impossible to obtain those frequencies. Furthermore, people—even experts in their fields—are often willing to *compare* the probabilities of two statements ('φ is more probable than ψ'), without being able to assign explicit probabilities to each of the statements *individually*. Many concrete examples are given by Szolovits and Pauker (1978) and Halpern and Rabin (1987). When modeling and analyzing such situations, qualitative probabilistic operators can be very useful.

2.4.1 Systems with Unary and Binary Operators

One of the earliest qualitative probabilistic logics is found in Hamblin (1959). The classical language $\mathcal{L}(\mathsf{Prop})$ is extended with a unary operator $\Box$, which is read as 'probably'. Hence a formula such as $\Box\varphi$ is to be read as 'probably φ'. This notion of 'probable' can be formalized as *sufficiently high (numerical) probability* (i.e. $P(\varphi) \geq t$, for some threshold value $0.5 < t \leq 1$). An alternative formalization of the 'probably'-operator is in terms of *plausibility*, which is a non-quantitative generalization of probability. In probabilistic terms, we can compare the probabilities of two statements φ and ψ at several levels of preciseness. For example, if $P(\varphi) = 0.8$ and $P(\psi) = 0.4$, we can make the following statements (with increasing levels of preciseness):

1. φ and ψ have different probabilities,
2. φ is more probable than ψ,
3. φ is 0.4 more probable than ψ,
4. φ is 2 times as probable as ψ.

Using the well-known terminology of types of measurement scales (Stevens 1946), these statements can be called 'nominal', 'ordinal', 'interval' and 'ratio' statements, respectively. The first two are qualitative, whereas the last two are quantitative in nature. In plausibilistic terms, only the first two statements can be made:

1. φ and ψ have different plausibilities,
2. φ is more plausible than ψ.

By abstracting away from the quantitative structure of probability theory, plausibility theory thus allows for coarser, more general models of uncertainty. One particular version of the plausibilistic approach to uncertainty will be addressed in much greater detail in Chapter 4.

Burgess (1969) further develops these qualitative probabilistic systems, focusing on the 'high numerical probability'-interpretation. Both Hamblin and Burgess introduce additional operators into their systems (expressing, for example, metaphysical necessity and/or knowledge), and study the interaction between the 'probably'-operator and these other modal operators.

However, the 'probably'-operator already displays some interesting features on its own (independent from any other operators). For example, if it is interpreted as 'sufficiently high probability', then it fails to satisfy the principle $(\Box\varphi \wedge \Box\psi) \rightarrow \Box(\varphi \wedge \psi)$. This means that it is not a *normal* modal operator, and thus cannot be given a Kripke (relational) semantics (Chellas 1980). Herzig and Longin (2003) and Arló Costa (2005) provide weaker systems of *neighborhood semantics* for such 'probably'-operators.

Another route is taken by Segerberg (1971) and Gärdenfors (1975a,b), who introduce a *binary* operator $\geq$. The formula $\varphi \geq \psi$ is to be read as 'φ is at least as probable as ψ'; formally: $P(\varphi) \geq P(\psi)$. The key idea is that one can completely characterize the behavior of $\geq$ without having to use the 'underlying' probabilities of the individual formulas. It should be noted that with comparative probability (a binary operator), one can also express some absolute probabilistic properties (unary operators). For example, (i) $\varphi \geq \top$ expresses that φ has probability 1, (ii) $\varphi \geq \neg\varphi$ expresses that the probability of φ is at least 0.5, and (iii) $\neg(\neg\varphi \geq \varphi)$ expresses that the probability of φ is strictly greater than 0.5.

2.4.2 Linguistic Issues

Unary and binary qualitative probabilistic operators have also been studied from a more linguistically oriented perspective. Kratzer (1991) argues that operators such as 'probably' and 'likely' (considered as natural language expressions, rather than quasi-technical terms) are semantically speaking indeed modal operators, on a par with epistemic and deontic modals such as 'possibly', 'might', 'must', etc.

Yalcin (2010) assesses various proposals for the semantics of probabilistic operators, by investigating which inference patterns they (in)validate, such as the *chancy modus ponens* pattern for the unary 'probably'-operator:

if p then q
probably p

---

probably q,

but also the *conditional to comparative* pattern for the binary 'is at least as probable as'-operator:

if p then q

---

q is at least as probable as p,

and the *positive form transfer* pattern that links the unary and binary operators:

q is at least as probable as p
probably p

---

probably q.

Yalcin also addresses the embedding potential of probability operators. For example, the formula $\Box\Box p$ is trivially well-formed because of the recursive definition of the logics' object languages, but are sentences such as

'probably, John is likely to be sleeping'

meaningful in natural languages such as English or Dutch? In its most natural reading, such a sentence is equivalent to one containing just a single probability operator, which is in line with the broader literature on modal concord (Zeijlstra 2007, Huitink 2012).

Yalcin's own analysis takes a binary, comparative operator $\geq$ as primitive, which is then used to define a unary 'probably'-operator, as was explained above ($\Box\varphi :\equiv \varphi \geq \neg\varphi$). This fits naturally with the broader literature on gradable adjectives (Kennedy 2007), in which the comparative form (e.g. '...is taller than ...') is typically taken to be theoretically basic, and then used to define the absolute form (e.g. '...is tall').

2.5 First-Order Probabilistic Operators

We now turn to systems that add quantitative probabilistic operators to the object language. In this section, we will focus on first-order operators. Consider the following example from Bacchus (1990):

‘More than 75% of all birds fly.’

This sentence has a straightforward probabilistic interpretation: when one randomly selects a bird, then the probability that the selected bird flies is more than 0.75. First-order probabilistic operators are needed to express these sort of statements.

2.5.1 A Basic System of First-Order Probabilistic Logic

In this subsection, we will study a basic system of first-order probabilistic logic. Its object language is as simple as possible, to allow full focus on the probabilistic quantifiers. The language is very much like the language of classical first-order logic, but rather than the familiar universal and existential quantifier, it contains a probabilistic quantifier.

We start by fixing a set Var of individual variables, a set Fun of function symbols, and a set Pred of predicate symbols. All function symbols f and predicate symbols R have arities $ar(f), ar(R) \in \mathbb{N}$.[13] The language contains two kinds of syntactical objects, viz. *terms* and *formulas*. The set of terms $\mathcal{T}(\mathsf{Var}, \mathsf{Fun})$ is defined by means of the following BNF:

$$t ::= x \mid f(t_1, \dots, t_{ar(f)})$$

where $x \in \mathsf{Var}$ and $f \in \mathsf{Fun}$. The language $\mathcal{L}^{\mathsf{BFOPL}}(\mathsf{Var}, \mathsf{Fun}, \mathsf{Pred})$ of basic first-order probabilistic logic is then defined by means of the following BNF:

$$\varphi ::= R(t_1, \dots, t_{ar(R)}) \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid Px(\varphi) \geq q$$

where $R \in \mathsf{Pred}$, $t_1, \dots, t_{ar(R)} \in \mathcal{T}(\mathsf{Var}, \mathsf{Fun})$ and $q \in [0,1] \cap \mathbb{Q}$. (The number q is restricted to be rational, in order to ensure that the language is countable.)

[13] Nullary function symbols and predicate symbols are also called *individual constants* and *propositions*, respectively.

Formulas of the form $Px(\varphi) \geq q$ should be read as: 'when an object is randomly selected, the probability that it satisfies φ is at least q'. Every free occurrence of x in φ is bound by the operator. We make use of the following abbreviations:

$$\begin{array}{lll} Px(\varphi) \leq q & \text{for} & Px(\neg\varphi) \geq 1-q, \\ Px(\varphi) > q & \text{for} & \neg(Px(\varphi) \leq q), \\ Px(\varphi) < q & \text{for} & \neg(Px(\varphi) \geq q), \\ Px(\varphi) = q & \text{for} & Px(\varphi) \geq q \wedge Px(\varphi) \leq q. \end{array}$$

This language is interpreted with respect to very simple first-order models and assignment functions, which are defined as follows:

Definition 2.5. A *probabilistic first-order model* is a triple $\mathbb{M} := \langle D, I, P\rangle$. Here, D is a finite, non-empty set called the *domain*. Furthermore, I is an *interpretation function*, which assigns to each $R \in \mathsf{Pred}$ a set $I(R) \subseteq D^{ar(R)}$ and to each $f \in \mathsf{Fun}$ a function $I(f)\colon D^{ar(f)} \to D$. Finally, P is a function that assigns a value $P(d) \in [0,1]$ to each element $d \in D$; it is required that $\sum_{d\in D} P(d) = 1$.

Definition 2.6. Given a probabilistic first-order model $\langle D, I, P\rangle$, an *assignment function* is a function $g\colon \mathsf{Var} \to D$. For any assignment function g, $x \in \mathsf{Var}$ and $d \in D$, the assignment function $f[x \mapsto d]$ is defined as follows:

$$f[x \mapsto d]\colon \mathsf{Var} \to D\colon y \mapsto \begin{cases} f(y) & \text{if } y \neq x, \\ d & \text{if } y = x. \end{cases}$$

Given a model $\mathbb{M} = \langle D, I, P\rangle$ and an assignment function g, the interpretation $[\![\, t \,]\!]^{\mathbb{M},g}$ of all terms $t \in \mathcal{T}(\mathsf{Var}, \mathsf{Fun})$ is defined as follows:

$$\begin{array}{lll} [\![\, x \,]\!]^{\mathbb{M},g} & := & g(x), \\ [\![\, f(t_1, \ldots, t_{ar(f)}) \,]\!]^{\mathbb{M},g} & := & I(f)\left([\![\, t_1 \,]\!]^{\mathbb{M},g}, \ldots, [\![\, t_{ar(f)} \,]\!]^{\mathbb{M},g}\right). \end{array}$$

The logic's semantics can now be defined as follows:

$$\begin{array}{lll} \mathbb{M}, g \models R(t_1, \ldots, t_{ar(R)}) & \text{iff} & \left([\![\, t_1 \,]\!]^{\mathbb{M},g}, \ldots, [\![\, t_{ar(R)} \,]\!]^{\mathbb{M},g}\right) \in I(R), \\ \mathbb{M}, g \models \neg\varphi & \text{iff} & \mathbb{M}, g \not\models \varphi, \\ \mathbb{M}, g \models \varphi \wedge \psi & \text{iff} & \mathbb{M}, g \models \varphi \text{ and } \mathbb{M}, g \models \psi, \\ \mathbb{M}, g \models Px(\varphi) \geq q & \text{iff} & \sum_{d\,:\,\mathbb{M}, g[x\mapsto d]\models\varphi} P(d) \geq q. \end{array}$$

The only non-standard clause is the final one, for the probabilistic quantifier. To understand it better, consider the following example:

Example 2.4. A vase contains 10 marbles: 6 are black and 4 are white. Intuitively, the probability that a randomly selected marble is black is 0.6, and the probability that a randomly selected marble is white is 0.4. To formalize this, define a probabilistic first-order model $\mathbb{M} := \langle D, I, P\rangle$ by putting $D := \{m_1, \dots, m_{10}\}$, $I(\mathit{black}) = \{m_1, \dots, m_5\}$, $I(\mathit{white}) = \{m_6, \dots, m_{10}\}$,[14] and $P(m_i) = 0.1$ for each $1 \leq i \leq 10$ (this captures the assumption that each marble is equally likely to be selected). For any assignment function g, variable $x \in \mathsf{Var}$ and $d \in D$, it holds that

$$\begin{aligned} \mathbb{M}, g[x \mapsto d] \models \mathit{black}(x) \quad & \text{iff} \quad [\![\, x \,]\!]^{\mathbb{M}, g[x \mapsto d]} \in I(\mathit{black}) \\ & \text{iff} \quad g[x \mapsto d](x) \in I(\mathit{black}) \\ & \text{iff} \quad d \in \{m_1, \dots, m_6\}. \end{aligned}$$

Hence

$$\sum_{d \,:\, \mathbb{M}, g[x \mapsto d] \models \mathit{black}(x)} P(d) = P(m_1) + \cdots + P(m_6) = 0.6.$$

It now follows by the semantic clause for the probability quantifier that

$$\mathbb{M}, g \models Px(\mathit{black}(x)) = 0.6.$$

Completely analogously, one can show that

$$\mathbb{M}, g \models Px(\mathit{white}(x)) = 0.4.$$

2.5.2 Three Extensions

The logic presented in the previous section is too simple to capture many forms of reasoning about probabilities. I will now discuss three extensions.

Quantifying over more than one variable. First of all, one would like to be able to reason about cases where more than one object is selected from the domain. Consider the following variant on Example 2.4: one picks a marble from the vase, puts it back, and then picks another marble from the vase. The probability

[14] We assume that *black* and *white* are unary predicate symbols in the object language.

that the first marble is black and the second one is white is $0.6 \times 0.4 = 0.24$, but this cannot be expressed in the simple language introduced in Subsection 2.5.1.

One can therefore introduce a probabilistic quantifier that deals with multiple variables simultaneously, thus obtaining formulas of the form $Px_1, \ldots, x_n(\varphi) \geq q$. To interpret such formulas in a probabilistic first-order model $\langle D, I, P \rangle$, the probability function P should assign probabilities not only to elements $d \in D$, but also to n-tuples $(d_1, \ldots, d_n) \in D^n$ (for all $n \in \mathbb{N}$). The simplest way to extend P is by assuming that all selections are *independent* and *with replacement*; one can then simply extend P to n-tuples by putting $P(d_1, \ldots, d_n) := \prod_{i=1}^{i=n} P(d_i)$. This approach is taken by Bacchus (1990) and Halpern (1990). Extending the semantics to this new operator is straightforward:[15]

$$\mathbb{M}, g \models Px_1 \ldots x_n(\varphi) \geq q \quad \text{iff} \quad \sum_{\substack{(d_1,\ldots,d_n)\,: \\ \mathbb{M}, g[x_1 \mapsto d_1, \ldots, x_n \mapsto d_n] \models \varphi}} P(d_1, \ldots, d_n) \geq q$$

For example, for the model $\mathbb{M}$ defined in Example 2.4 (and an arbitrary assignment function g), it holds that

$$\mathbb{M}, g \models Px, y(\mathit{black}(x) \wedge \mathit{white}(y)) = 0.24.$$

There also exist more general approaches to extending the measure from single elements of the domain to tuples of elements, which do *not* assume that the selections are independent and with replacement. Such alternatives are explored by Hoover (1978) and Keisler (1985).

Conditional probabilities. Recall the sentence from the beginning of this section: ‘more than 75% of all birds fly’. This cannot be adequately captured in a model where the domain contains objects that are not birds. These non-bird objects should not matter to what one wishes to express, but the probability quantifiers discussed so far quantify over the entire domain. In order to restrict quantification, one must add conditional probability operators, thus obtaining formulas of the form $Px(\varphi \mid \psi) \geq q$. These formulas have the following semantics:

[15] The assignment function $g[x_1 \mapsto d_1, \ldots, x_n \mapsto d_n]$ is defined as expected, viz.

$$g[x_1 \mapsto d_1, \ldots, x_n \mapsto d_n](y) := \begin{cases} d_i & \text{if } y = x_i \text{ for some } 1 \leq i \leq n, \\ y & \text{otherwise.} \end{cases}$$

$$\mathbb{M}, g \models Px(\varphi \mid \psi) \geq q \quad \text{iff} \quad \text{if } \textstyle\sum_{d \,:\, \mathbb{M}, g[x \mapsto d] \models \psi} P(d) > 0, \\ \text{then } \frac{\sum_{d \,:\, \mathbb{M}, g[x \mapsto d] \models \varphi \wedge \psi} P(d)}{\sum_{d \,:\, \mathbb{M}, g[x \mapsto d] \models \psi} P(d)} \geq q.$$

With these operators, the formula $Px(\mathit{fly}(x) \mid \mathit{bird}(x)) > 0.75$ expresses that more than 75% of all birds fly.

Probabilities as terms. The basic probabilistic first-order logic has formulas of the form $Px(\varphi) \geq q$, which say that the probability of randomly selecting an object that satisfies φ is at least q. In general, however, one might want to say other things about the probability of randomly selecting an object that satisfies φ; for example, one might want to compare this probability with the probability of randomly selecting an object that satisfies some other formula ψ.

In such cases, it is more convenient to treat probabilities as terms in their own right. For each formula φ, the expression $Px(\varphi)$ is thus added to the set of terms $\mathcal{T}(\mathsf{Var}, \mathsf{Fun})$.[16] For any model $\mathbb{M}$ and assignment function g, the interpretation of $Px(\varphi)$ is defined as follows:

$$[\![\, Px(\varphi) \,]\!]^{\mathbb{M}, g} := \sum_{d \,:\, \mathbb{M}, g[x \mapsto d] \models \varphi} P(d).$$

One can then extend the language with arithmetical operations such as addition and multiplication, and with operators such as equality and various inequalities to compare probability terms.

Such extensions require that the language is *sorted*, i.e. that it contains two separate classes of terms: one for probabilities, numbers and the results of arithmetical operations on such terms, and one for the 'ordinary' domain of discourse which the probabilistic operators quantify over. I will not present such a language and semantics in detail here; details can be found in Bacchus (1990).

In the context of Example 2.4, one might wish to say that a randomly selected marble is 1.5 times more likely to be black than to be white. In other words, the probability that a randomly selected marble is black is 1.5 times higher than the probability that a randomly selected marble is white. This can easily be expressed in a sorted language (which contains probabilities as terms, 1.5 as a constant symbol, and × as a function symbol):

$$\mathbb{M}, g \models Px(\mathit{black}(x)) = 1.5 \times Px(\mathit{white}(x)).$$

[16] Hence, the terms and formulas are defined by a mutual recursion.

2.5.3 Metatheoretical Results

It is hard to provide proof systems for first-order probabilistic logics, because the validity problem for these logics is generally not decidable, and not even semi-decidable (Abadi and Halpern 1994). Compare the situation with classical first-order logic, which is semi-decidable but not decidable (Boolos et al. 2007):

- if an inference in classical first-order logic is valid, then it is guaranteed that one can find out in finite time,
- however, if an inference in classical first-order logic is not valid, then it is *not* guaranteed that one can find out in finite time.

Since first-order probabilistic logic is not even semi-decidable, neither validity nor invalidity of inferences is finitely discoverable, i.e.:

- if an inference in first-order probabilistic logic is valid, then it is *not* guaranteed that one can find out in finite time,
- if an inference in first-order probabilistic logic is not valid, then it is *not* guaranteed that one can find out in finite time.

Despite these limitations, there exist many interesting metatheoretical results for various first-order probabilistic logics. For example, Hoover (1978) and Keisler (1985) study completeness results. Bacchus (1990) and Halpern (1990) also provide complete axiomatizations, and study combinations of first-order probabilistic logics and modal probabilistic logics.

2.6 Propositional Probabilistic Operators

In the previous section, we focused on systems that add first-order probabilistic operators to the object language. In this section, we will study systems that deal with another type of quantitative probabilistic operators, viz. propositional operators.

Since propositional probabilistic operators transform a given formula φ into another formula (e.g. $P(\varphi) \geq 0.6$), the recursive structure of the object language allows for the 'nesting' of such probabilistic formulas (e.g. formulas of the form $(P(\varphi) \geq 0.6) \geq 0.4$). Furthermore, it is highly natural to study the interaction between propositional probabilistic operators and other propositional operators, such as epistemic operators. However, as was already noted in Section 2.1, these extensions will be studied in much more detail in Chapter 3. Therefore, in this section, we will focus on a rather basic system of probabilistic propositional logic, which does not have an epistemic component and does not allow reasoning about higher-order (nested) probabilities. This will allow us to make some technical observations in their most general form (which will also apply to the more complex systems that are introduced in Chapter 3).

Subsection 2.6.1 introduces probability spaces and probabilistic models. Subsection 2.6.2 defines the language that is interpreted on these models, and Subsection 2.6.3 discusses its expressivity. Subsection 2.6.4 provides a complete axiomatization.

2.6.1 Probabilistic Models

We begin by introducing probabilistic models, which are based on a set of states S. There are two ways of adding probabilistic information to such a set S. The first is to define probabilities directly on the states of S; the second is to define probabilities on (a subcollection of) the subsets of S. These two approaches are formalized in the notions of *discrete probability structure* and *probability space*, respectively.

Definition 2.7. A *discrete probability structure* is a tuple $\langle S, p\rangle$, where S is a non-empty, finite set, whose elements will usually be called ‘states’ or ‘possible worlds’, and $p\colon S \to [0,1]$ is a function such that $\sum_{s\in S} p(s) = 1$.

Definition 2.8. A *probability space* is a tuple $\langle S, \mathcal{A}, \mu\rangle$, where S is an arbitrary (potentially uncountable) set called the *sample space*, $\mathcal{A} \subseteq \wp(S)$ is a *σ-algebra* over S,[17] and $\mu\colon \mathcal{A} \to [0,1]$ is a *probability measure*, i.e. a countably additive function[18] such that $\mu(S) = 1$. Finally, the elements of $\mathcal{A}$ are called the *measurable sets* of the space.

The function p on individual states in a discrete probability structure is naturally extended to a function p^+ on sets of states, by putting

$$p^+\colon \wp(S) \to [0,1]\colon X \mapsto p^+(X) := \sum_{x\in X} p(x).$$

[17] I.e. $\mathcal{A}$ has the following properties: (i) $S \in \mathcal{A}$, (ii) if $X, Y \in \mathcal{A}$, then $X \cap Y \in \mathcal{A}$, and (iii) if $\mathcal{X} \subseteq \mathcal{A}$ and $\mathcal{X}$ is countable, then $\bigcap \mathcal{X} \in \mathcal{A}$.

[18] A real-valued set function is *countably additive* iff for any countable collection of sets A_i that are pairwise disjoint ($A_i \neq A_j$ for each $i \neq j$), it holds that $f(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} f(A_i)$.

Note that it follows immediately that $p^+(S) = 1$. Given this construction, it is easy to check that every discrete probability structure gives rise to a probability space, by taking the σ-algebra of the space to consist of *all* subsets: if $\langle S, p\rangle$ is a discrete probability structure, then $\langle S, \wp(S), p^+\rangle$ is a probability space. Hence, probability spaces can be seen as a generalization of discrete probability structures—or vice versa, discrete probability structures can be seen as a special case of probability spaces, viz. those whose σ-algebra is the powerset of their sample space.

Discrete probability structures, defined by point-functions, have the advantage of simplicity, as well as having numerous countable settings for examples and applications. The probability spaces, with their set-functions defined on σ-algebras, have the advantage of generality. The purpose of the σ-algebra is to restrict the domain of the probability set-function from the entire powerset of the sample space to a smaller σ-algebra (cf. *supra*). This is unavoidable, for example, when we wish to define a uniform probability distribution over an infinite set: one cannot assign equal probability to all singletons while maintaining countable additivity (since the sum of all probabilities must be at most 1).

In the remainder of this subsection, I will make use of both discrete probability structures and probability spaces. However, in the remaining chapters of this thesis, I will work exclusively with discrete probability functions, since the additional machinery of σ-algebras is unnecessary for the technical and philosophical results to be presented.

Let Prop be a countable set of atomic propositions (just as in Subsection 2.3). Probabilistic models and discrete probabilistic models are defined as follows:

Definition 2.9. A *discrete probabilistic model* is a tuple $\mathbb{M} = \langle S, p, V\rangle$, where $\langle S, p\rangle$ is a discrete probability structure, and $V\colon \mathsf{Prop} \to \wp(S)$ is a valuation. The class of all discrete probabilistic models will be denoted $\mathcal{C}_d^{PPL}$.

Definition 2.10. A *probabilistic model* is a tuple $\mathbb{M} = \langle S, \mathcal{A}, \mu, V\rangle$, where $\langle S, \mathcal{A}, \mu\rangle$ is a probability space, and $V\colon \mathsf{Prop} \to \mathcal{A}$ is a valuation. The class of all probabilistic models will be denoted $\mathcal{C}^{PPL}$.

The details of discrete probability structures and probability spaces have already been discussed above. The valuation V determines which atomic propositions are true at which states; intuitively, $s \in V(p)$ means that p is true at the state s. Note that in the definition of probabilistic model, the valuation is required to map atomic propositions to the σ-algebra $\mathcal{A}$, instead of to the full powerset $\wp(S)$

(as is usually done in modal settings). This restriction ensures that every Boolean formula has a probability,[19] which will be used in the definition of the formal semantics.

2.6.2 Language and Semantics

The basic probability language is defined in layers. For a given set of atomic propositions Prop, let $\mathcal{L}_B(\mathsf{Prop})$ be the set of Boolean formulas, given by the following BNF:

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

where $p \in \mathsf{Prop}$. Next, let $\mathcal{T}(\mathsf{Prop})$ be a set of terms, given by the following BNF:

$$t \ ::= \ aP(\varphi) \mid t + t$$

where $a \in \mathbb{Q}$ is a rational number, and $\varphi \in \mathcal{L}_B(\mathsf{Prop})$. Finally, let $\mathcal{L}_P(\mathsf{Prop})$ be the set of probability formulas, given by the grammar:

$$f \ ::= \ t \geq b \mid \neg f \mid f \wedge f$$

where $b \in \mathbb{Q}$ and $t \in \mathcal{T}(\mathsf{Prop})$. (The numbers a, b are restricted to be rational, in order to ensure that the language is countable.)

Formulas of the form $t \geq b$ are called *probability formulas*. We allow for linear combinations in probability formulas, because this additional expressivity is useful when looking for a complete axiomatization (Fagin et al. 1990), and because it allows us to make comparative judgments such as 'φ is at least twice as probable as ψ': this is expressed by the formula $P(\varphi) \geq 2P(\psi)$. The added expressivity of linear combinations will be addressed in more detail in Subsection 2.6.3.

The formula $P(\varphi) \geq 2P(\psi)$ is actually an abbreviation for $P(\varphi) - 2P(\psi) \geq 0$. In general, we introduce the following abbreviations:

[19] The σ-algebra of measurable sets thus plays a role similar to that of the modal algebra of admissible sets in a general Kripke model (Blackburn et al. 2001, Definition 1.32): both are meant to restrict the range of the valuation function, to prevent certain 'problematic' sets from becoming truth sets of formulas.

$$\begin{array}{lll}
\sum_{\ell=1}^{n} a_\ell P(\varphi_\ell) \geq b & \text{for} & a_1 P(\varphi_1) + \cdots + a_n P(\varphi_n) \geq b, \\
a_1 P_i(\varphi_1) \geq a_2 P_i(\varphi_2) & \text{for} & a_1 P(\varphi_1) + (-a_2) P(\varphi_2) \geq 0, \\
\sum_{\ell=1}^{n} a_\ell P(\varphi_\ell) \leq b & \text{for} & \sum_{\ell=1}^{n} (-a_\ell) P(\varphi_\ell) \geq -b, \\
\sum_{\ell=1}^{n} a_\ell P(\varphi_\ell) < b & \text{for} & \neg\big(\sum_{\ell=1}^{n} a_\ell P(\varphi_\ell) \geq b\big), \\
\sum_{\ell=1}^{n} a_\ell P(\varphi_\ell) > b & \text{for} & \neg\big(\sum_{\ell=1}^{n} a_\ell P(\varphi_\ell) \leq b\big), \\
\sum_{\ell=1}^{n} a_\ell P(\varphi_\ell) = b & \text{for} & \sum_{\ell=1}^{n} a_\ell P(\varphi_\ell) \geq b \wedge \sum_{\ell=1}^{n} a_\ell P(\varphi_\ell) \leq b.
\end{array}$$

The formal semantics is defined in layers, just like the language itself. Given a probabilistic model $\mathbb{M}$ (discrete or otherwise) with domain S and valuation V, we first define a function $\llbracket \cdot \rrbracket^{\mathbb{M}} \colon \mathcal{L}_B \to \wp(S)$ by putting

$$\begin{array}{lll}
\llbracket p \rrbracket^{\mathbb{M}} & = & V(p), \\
\llbracket \neg\varphi \rrbracket^{\mathbb{M}} & = & S - \llbracket \varphi \rrbracket^{\mathbb{M}}, \\
\llbracket \varphi \wedge \psi \rrbracket^{\mathbb{M}} & = & \llbracket \varphi \rrbracket^{\mathbb{M}} \cap \llbracket \psi \rrbracket^{\mathbb{M}}.
\end{array}$$

It is easy to check that if $\mathbb{M}$ is a non-discrete probabilistic model, the sets $\llbracket \varphi \rrbracket^{\mathbb{M}}$ are measurable for all $\varphi \in \mathcal{L}_B$.

Given a probabilistic model $\mathbb{M} = \langle S, \mathcal{A}, \mu, V \rangle$, the semantics of $\mathcal{L}_P$ looks as follows:

$$\begin{array}{lll}
\mathbb{M} \models \sum_{\ell=1}^{n} a_\ell P(\varphi_\ell) \geq b & \text{iff} & \sum_{\ell=1}^{n} a_\ell \mu(\llbracket \varphi_\ell \rrbracket^{\mathbb{M}}) \geq b, \\
\mathbb{M} \models \neg f & \text{iff} & \mathbb{M} \not\models f, \\
\mathbb{M} \models f_1 \wedge f_2 & \text{iff} & \mathbb{M} \models f_1 \text{ and } \mathbb{M} \models f_2.
\end{array}$$

If $\mathbb{M}$ is a *discrete* probabilistic model, i.e. $\mathbb{M} = \langle S, p, V \rangle$, then the first semantic clause makes use of the additive lifting p^+ of the probability function p:

$$\mathbb{M} \models \sum_{\ell=1}^{n} a_\ell P(\varphi_\ell) \geq b \quad \text{iff} \quad \sum_{\ell=1}^{n} a_\ell p^+(\llbracket \varphi_\ell \rrbracket^{\mathbb{M}}) \geq b.$$

It should be emphasized that $\models$ is only defined for $\mathcal{L}_P$-formulas. Hence, for a propositional atom p and a probabilistic model $\mathbb{M}$, it makes sense to ask whether $\mathbb{M} \models P(p) \geq 0.6$, but not whether $\mathbb{M} \models p$. More importantly, note that probability formulas are of the form $P(\varphi) \geq k$, where $\varphi \in \mathcal{L}_B$ is a Boolean combination of propositional atoms. In other words, higher-order probabilities cannot be expressed in $\mathcal{L}_P$: formulas such as $P(P(q) \geq 0.7) \geq 0.6$ are not well-formed.

I will finish this subsection by showing that despite these limitations, this framework is quite powerful, and can be used to naturally formalize rather intricate scenarios.[20]

Example 2.5. Three indistinguishable balls are simultaneously dropped down a tube. Inside the tube, each ball gets stuck (independently of the other balls) with some small probability ε. In other words, each ball rolls out at the other end of the tube with probability $1 - \varepsilon$, and does not roll out with probability ε. Because the three balls are indistinguishable, we cannot know *which* ball(s) got stuck; we can only count the *number* of balls that roll out of the tube. (Of course, if we count 3 (resp. 0) balls rolling out, then we do know that no (resp. all) balls have gotten stuck.) What is the probability that exactly 2 balls will roll out?

We define the sample space as $S := \{(s_1, s_2, s_3) \mid s_i \in \{0, 1\}\}$, where

$$s_i := \begin{cases} 1 & \text{if ball } i \text{ rolls out,} \\ 0 & \text{if ball } i \text{ does not roll out.} \end{cases}$$

We consider the propositional atoms $\mathsf{rollout}_n$, which are to be read as 'exactly n balls roll out'. Obviously, we put $V(\mathsf{rollout}_n) := \{(s_1, s_2, s_3) \in S \mid s_1 + s_2 + s_3 = n\}$. Let $\mathcal{A}$ be the σ-algebra generated by $\{V(\mathsf{rollout}_n) \mid 0 \leq n \leq 3\}$. This reflects the fact that we can only observe the *number* of balls rolling out of the tube; for example, the singleton set $\{(1, 1, 0)\}$ (which contains the information that b_3 got stuck, and b_1 and b_2 roll out) is not in $\mathcal{A}$. Finally, we define a probability measure μ by putting $\mu(V(\mathsf{rollout}_n)) := \binom{3}{n}(1 - \varepsilon)^n \varepsilon^{3-n}$ (i.e. a binomial distribution).[21] The probabilistic model $\mathbb{M} := \langle S, \mathcal{A}, \mu, V \rangle$ fully captures the scenario. For example, if $\varepsilon = 0.1$, one can check that $\mathbb{M} \models P(\mathsf{rollout}_2) = 0.243$.

2.6.3 The Expressivity of Linear Combinations

The set of terms $\mathcal{T}(\mathsf{Prop})$ introduced in the previous subsection contains not only terms of the form $P(\varphi)$ (with $\varphi \in \mathcal{L}_B(\mathsf{Prop})$), but also linear combinatons: $a_1 P(\varphi_1) + \cdots + a_n P(\varphi_n)$. Besides having technical motivations, this leads to the language $\mathcal{L}_P(\mathsf{Prop})$ being highly expressive. For example, (i) it can express comparative probability judgments (of the form $P(\varphi) \geq P(\psi)$), and (ii) the $\geq$-comparison suffices to define all others (cf. the abbreviations stated above for $>$, $\leq$, etc.).

[20] Example 2.5 is loosely based on Example 1.2.5 from Geiss and Geiss (2009).

[21] The binomial coefficients are defined as $\binom{m}{n} = \frac{m!}{n!(m-n)!}$.

However, this expressivity gain should not be exaggerated. For example, if we restrict ourselves to 'single' probabilities—i.e. terms of the form $P(\varphi)$, and thus probability formulas of the form $P(\varphi) \geq b$—, all the comparisons are already definable. Let's first consider the case of $P(\varphi) \leq b$. With the abbreviations mentioned above in mind, this can be defined as $-P(\varphi) \geq -b$. However, this involves scalar multiplication (with -1) of $P(\varphi)$, and thus already takes us outside the narrow realm of 'single' probabilities. There exists, however, an alternative definition that stays inside this realm, viz. $P(\neg\varphi) \geq 1 - b$.[22] Defining the other comparisons is now straightforward:

$$\begin{array}{lll} P(\varphi) \leq b & \text{for} & P(\neg\varphi) \geq 1 - b, \\ P(\varphi) < b & \text{for} & \neg\,(P(\varphi) \geq b), \\ P(\varphi) > b & \text{for} & \neg\,(P(\varphi) \leq b), \\ P(\varphi) = b & \text{for} & P(\varphi) \geq b \wedge P(\varphi) \leq b. \end{array}$$

Next, consider the (finite)[23] additivity property of probability measures:

$$\mu(X \cup Y) = \mu(X) + \mu(Y) \text{ for all disjoint } X, Y \in \mathcal{A}.$$

Using linear combinations of probabilities, this can be expressed almost 'literally' in the object language:

$$P(\varphi \vee \psi) = P(\varphi) + P(\psi) \text{ whenever } \neg(\varphi \wedge \psi) \text{ is a tautology.} \tag{2.3}$$

The axiomatization that is introduced in the next subsection uses another, equivalent expression, which also involves linear combinations:

$$P(\varphi \wedge \psi) + P(\varphi \wedge \neg\psi) = P(\varphi).$$

This suggests an alternative way of expressing additivity, which does *not* make use of linear combinations:

$$\big(P(\varphi \wedge \psi) = a \wedge P(\varphi \wedge \neg\psi) = b\big) \to P(\varphi) = a + b. \tag{2.4}$$

[22] Note that if $\varphi \in \mathcal{L}_B$, then $\neg\varphi \in \mathcal{L}_B$ as well, and thus $P(\neg\varphi) \geq 1 - b$ is a perfectly well-formed $\mathcal{L}_P$-formula.

[23] Probability measures actually satisfy the stronger *countable* additivity requirement. Since this requirement cannot easily be captured in a finitary logic (but see Goldblatt (2010) for an example where countable additivity is captured by means of an infinitary rule in a finitary logic), it is customary to focus on formulas for *finite* additivity.

Heifetz and Mongin (2001) provide an axiomatization of probabilistic propositional logic along these lines. Note that since (2.3) makes use of a linear combination of probability terms, it does not explicitly contain the numbers a and b, and is therefore able to express additivity in a *single* formula (for given φ and ψ, of course). The alternative formulation (2.4), which does *not* make use of linear combinations, should be seen as a *scheme*: it corresponds to the (countable) set of formulas

$$\Big\{\Big(\big(P(\varphi \wedge \psi) = a \wedge P(\varphi \wedge \neg\psi) = b\big) \to P(\varphi) = a + b\Big) \in \mathcal{L}_P(\mathsf{Prop}) \ \Big| \ a, b \in [0,1] \cap \mathbb{Q}\Big\}.$$

Finally, note that linear combinations do not make the language more powerful at *distinguishing* between models. For any language $\mathcal{L}$ and models $\mathbb{M}_1, \mathbb{M}_2$ (on which $\mathcal{L}$ is interpretable), we define:

$$\mathbb{M}_1 \text{ and } \mathbb{M}_2 \text{ are } \textit{$\mathcal{L}$-equivalent} \quad \text{iff} \quad \forall \varphi \in \mathcal{L}\colon \mathbb{M}_1 \models \varphi \Leftrightarrow \mathbb{M}_2 \models \varphi.$$

Informally, two models are $\mathcal{L}$-equivalent if $\mathcal{L}$ cannot distinguish between them.

Let $\mathcal{L}_P^*(\mathsf{Prop})$ be the language that is obtained from $\mathcal{L}_P(\mathsf{Prop})$ by only allowing probability formulas of the form $P(\varphi) \geq a$ (in other words, linear combinations of probability terms are not allowed). Then one can show the following:

Lemma 2.1. *Consider arbitrary probabilistic models* $\mathbb{M}_1 = \langle S_1, \mathcal{A}_1, \mu_1, V_1\rangle$ *and* $\mathbb{M}_2 = \langle S_2, \mathcal{A}_2, \mu_2, V_2\rangle$. *These models are* $\mathcal{L}_P(\mathsf{Prop})$*-equivalent iff they are* $\mathcal{L}_P^*(\mathsf{Prop})$*-equivalent.*

Proof. If $\mathbb{M}_1$ and $\mathbb{M}_2$ are $\mathcal{L}_P$-equivalent, then they are trivially $\mathcal{L}_P^*$-equivalent as well, since $\mathcal{L}_P^* \subseteq \mathcal{L}_P$. We now prove the other direction.

Consider an arbitrary propositional formula $\varphi \in \mathcal{L}_B$. For a reductio, suppose that $\mu_1([\![\varphi]\!]^{\mathbb{M}_1}) \neq \mu_2([\![\varphi]\!]^{\mathbb{M}_2})$. Without loss of generality, assume that $\mu_1([\![\varphi]\!]^{\mathbb{M}_1}) > \mu_2([\![\varphi]\!]^{\mathbb{M}_2})$ (the other case is completely analogous). Since $\mathbb{Q}$ is dense in $\mathbb{R}$, there exists a $k \in \mathbb{Q}$ such that $\mu_1([\![\varphi]\!]^{\mathbb{M}_1}) > k > \mu_2([\![\varphi]\!]^{\mathbb{M}_2})$. It now follows that $\mathbb{M}_1 \models P(\varphi) \geq k$, while $\mathbb{M}_2 \not\models P(\varphi) \geq k$, which contradicts the assumption that these models are $\mathcal{L}_P^*$-equivalent. We therefore conclude that $\mu_1([\![\varphi]\!]^{\mathbb{M}_1}) = \mu_2([\![\varphi]\!]^{\mathbb{M}_2})$. Since $\varphi \in \mathcal{L}_B$ was chosen arbitrarily, this holds for *all* propositional formulas. Hence, for any probability formula $a_1P(\varphi_1) + \cdots + a_nP(\varphi_n) \geq b$, we have:

Figure 2.2: Componentwise axiomatization of probabilistic propositional logic.

1. propositional component
    - all propositional tautologies and the modus ponens rule
2. probabilistic component
    - $P(\varphi) \geq 0$
    - $P(\top) = 1$
    - $P(\varphi \wedge \psi) + P(\varphi \wedge \neg\psi) = P(\varphi)$
    - if $\vdash \varphi \leftrightarrow \psi$ then $\vdash P(\varphi) = P(\psi)$
3. linear inequalities component
    - $\sum_{\ell=1}^{n} a_\ell P(\varphi_\ell) \geq b \leftrightarrow \sum_{\ell=1}^{n} a_\ell P(\varphi_\ell) + 0P(\varphi_{n+1}) \geq b$
    - $\sum_{\ell=1}^{n} a_\ell P(\varphi_\ell) \geq b \leftrightarrow \sum_{\ell=1}^{n} a_{p(\ell)} P(\varphi_{p(\ell)}) \geq b$ (for any permutation p of $1, \ldots, n$)
    - $\sum_{\ell=1}^{n} a_\ell P(\varphi_\ell) \geq b \wedge \sum_{\ell=1}^{n} a'_\ell P(\varphi_\ell) \geq b' \rightarrow \sum_{\ell=1}^{n} (a_\ell + a'_\ell) P(\varphi_\ell) \geq b + b'$
    - $\sum_{\ell=1}^{n} a_\ell P(\varphi_\ell) \geq b \leftrightarrow \sum_{\ell=1}^{n} d a_\ell P(\varphi_\ell) \geq db$ (for any $d > 0$)
    - $\sum_{\ell=1}^{n} a_\ell P(\varphi_\ell) \geq b \vee \sum_{\ell=1}^{n} a_\ell P(\varphi_\ell) \leq b$
    - $\sum_{\ell=1}^{n} a_\ell P(\varphi_\ell) \geq b \rightarrow \sum_{\ell=1}^{n} a_\ell P(\varphi_\ell) > b'$ (for any $b' < b$)

$$
\begin{aligned}
\mathbb{M}_1 \models \textstyle\sum_{\ell=1}^{n} a_\ell P(\varphi_\ell) \geq b \quad & \text{iff} \quad \textstyle\sum_{\ell=1}^{n} a_\ell \mu_1(\llbracket \varphi_\ell \rrbracket^{\mathbb{M}_1}) \geq b \\
& \text{iff} \quad \textstyle\sum_{\ell=1}^{n} a_\ell \mu_2(\llbracket \varphi_\ell \rrbracket^{\mathbb{M}_2}) \geq b \\
& \text{iff} \quad \mathbb{M}_2 \models \textstyle\sum_{\ell=1}^{n} a_\ell P(\varphi_\ell) \geq b.
\end{aligned}
$$

Since $\mathbb{M}_1$ and $\mathbb{M}_2$ agree on all probability formulas, they also agree on all Boolean combinations of such formulas, and are thus $\mathcal{L}_P$-equivalent. □

2.6.4 Proof System

A proof system for probabilistic propositional logic is given in Figure 2.2. Beyond the propositional component, there is a probabilistic component, which is a straightforward translation into $\mathcal{L}_P$ of the well-known Kolmogorov axioms of probability, together with a rule stating that provably equivalent formulas have identical probabilities. This component thus ensures that the formal symbol $P(\,\cdot\,)$ behaves like a real probability function. Finally, the linear inequalities component is mainly a technical tool to ensure that the logic is strong enough to capture the behavior of linear inequalities of probabilities.

Fagin et al. (1990) show that this logic is sound and complete:

Theorem 2.6. *Probabilistic propositional logic, as axiomatized in Figure 2.2, is sound and weakly complete with respect to the class $\mathcal{C}^{PPL}$ of probabilistic models, and also with respect to the class $\mathcal{C}_d^{PPL}$ of discrete probabilistic models.*

The notion of completeness used in this theorem is *weak* completeness ($\vdash \varphi$ iff $\models \varphi$), rather than *strong* completeness ($\Gamma \vdash \varphi$ iff $\Gamma \models \varphi$). These two notions do not coincide in probabilistic propositional logic, because this logic is not *compact*. For example, every finite subset of the set

$$\{P(p) > 0\} \cup \{P(p) \leq k \,|\, k > 0\}$$

is satisfiable, but the entire set is not. (Similar remarks apply to the other logics that will be discussed in later chapters.)

The proof of Theorem 2.6 involves establishing the existence of a satisfying model for a consistent probability formula f. To do this, it is shown that f is provably equivalent to a conjunction of probability formulas or negations of probability formulas. Hence, f is satisfiable iff the system of linear inequalities corresponding to this conjunction of (negated) probability formulas, together with the equalities and inequalities given by the Kolmogorov axioms, has a solution.[24] The satisfying model has a finite number of states, its σ-algebra is the powerset of its domain, and probabilities are assigned to singletons according to the solution of the linear system. This model is based on a probability space, and is thus a *non-discrete* probabilistic model. However, since all subsets of its domain are measurable, it can also be viewed as a *discrete* probabilistic model. It follows that the axiomatization in Figure 2.2 is complete with respect to both kinds of probabilistic models ($\mathcal{C}^{PPL}$ and $\mathcal{C}_d^{PPL}$).

[24] Recall the proof sketch of Theorem 2.5.

2.7 Conclusion

This chapter has provided an overview of a wide variety of approaches to combining logic and probability theory. After delineating the field of probabilistic logic, I discussed the major approaches in the field, and showed that they can be organized in a systematical and logically meaningful way (recall Figure 2.1).

There is one major class of systems that was not discussed in this chapter, viz. systems whose object language contains both propositional probabilistic operators and other propositional operators (in particular, epistemic operators). These probabilistic epistemic logics will be studied in the next chapter.

3 Dynamic Epistemic Logic with Probabilities

3.1 Introduction

Epistemic logic and probability theory both provide formal accounts of information. Epistemic logic takes a *qualitative* perspective on information, and works with a modal operator K. Formulas such as $K\varphi$ can be interpreted as 'the agent knows that φ', 'the agent believes that φ', or, more generally speaking, 'φ follows from the agent's current information'. Probability theory, on the other hand, takes a *quantitative* perspective on information, and works with numerical probability functions P. Formulas such $P(\varphi) = k$ can be interpreted as 'the probability of φ is k'. In the present context, probabilities will usually be interpreted subjectively, and can thus be taken to represent the agent's degrees of belief or credences.

With respect to one and the same formula φ, epistemic logic is able to distinguish between three epistemic attitudes: knowing its truth ($K\varphi$), knowing its falsity ($K\neg\varphi$), and being ignorant about its truth value ($\neg K\varphi \wedge \neg K\neg\varphi$).[1] Probability theory, however, distinguishes infinitely many epistemic attitudes with respect to φ, viz. assigning it probability k ($P(\varphi) = k$), for every $k \in [0, 1]$. In this sense probability theory can be said to provide a much more *fine-grained* perspective on information.

While epistemic logic thus is a coarser account of information, it certainly has a wider scope. From its very origins in Hintikka (1962), epistemic logic has not only been concerned with knowledge about 'the world', but also with knowl-

[1] In Part III, we will see that these three formulas, together with their negations, form a (strong) Sesmat-Blanché hexagon.

edge about knowledge, i.e. with *higher-order information*. Typical discussions focus on principles such as positive introspection ($K\varphi \to KK\varphi$). In contrast, probability theory rarely talks about principles involving higher-order probabilities, such as $P(\varphi) = 1 \to P(P(\varphi) = 1) = 1$.[2] This issue becomes even more pressing in multi-agent scenarios. Natural examples might involve an agent a not having any information about a proposition φ, while being certain that another agent, b, does have this information. In epistemic logic this is naturally formalized as

$$\neg K_a\varphi \wedge \neg K_a\neg\varphi \wedge K_a(K_b\varphi \vee K_b\neg\varphi).$$

A formalization in probability theory might look as follows:

$$P_a(\varphi) = 0.5 \wedge P_a(P_b(\varphi) = 1 \vee P_b(\varphi) = 0) = 1.$$

However, because this statement makes use of 'nested' probabilities, it is rarely used in standard treatments of probability theory.

An additional theme is that of dynamics, i.e. *information change*. The agents' information is not eternally the same; rather, it should be changed in the light of new incoming information. Probability theory typically uses Bayesian updating to represent information change (but other, more complicated update mechanisms are available as well). Dynamic epistemic logic interprets new information as changing the epistemic model, and uses the new, updated model to represent the agents' updated information states. Once again, the main difference is that dynamic epistemic logic takes (changes in) higher-order information into account, whereas probability theory does not.

For all these reasons, the project of *probabilistic epistemic logic* seems very appealing. Such systems inherit the fine-grained perspective on information from probability theory, and the representation of higher-order information from epistemic logic. Their *dynamic* versions provide a unified perspective on changes in first- and higher-order information. In other words, they can be thought of as

[2] A notable exception is 'Miller's principle', which states that $P_1(\varphi \,|\, P_2(\varphi) = b) = b$. The probability functions P_1 and P_2 can have various interpretations, such as the probabilities of two agents, subjective probability (credence) and objective probability (chance), or the probabilities of one agent at different moments in time—in the last two cases, the principle is also called the 'principal principle' or the 'principle of reflection', respectively. These principles have been widely discussed in Bayesian epistemology and philosophy of science (Miller 1966, Lewis 1980, van Fraassen 1984, Halpern 1991, Meacham 2010). Regardless of one's agreement or disagreement with these principles, arguing for or against them requires a language in which they can at least be expressed, i.e. in which higher-order probabilities are allowed.

incorporating the complementary perspectives of (dynamic) epistemic logic and probability theory, thus yielding richer and more detailed accounts of information and information flow.

The remainder of this chapter is organized as follows. Section 3.2 introduces the static framework of probabilistic epistemic logic, and discusses its intuitive interpretation and technical features. Section 3.3 focuses on a rather straightforward type of dynamics, namely public announcements. It describes a probabilistic version of the well-known system of public announcement logic, and compares public announcement and Bayesian conditionalization. In Section 3.4 a more general update mechanism is introduced. This is a probabilistic version of the 'product update' mechanism in dynamic epistemic logic. Finally, Section 3.5 indicates some applications and potential avenues of further research for the systems discussed in this chapter.

3.2 Probabilistic Epistemic Logic

In this section, I introduce the static framework of probabilistic epistemic logic, which will be 'dynamified' in Sections 3.3 and 3.4. Subsection 3.2.1 discusses the models on which the logic is interpreted. Subsection 3.2.2 defines the formal language and its semantics. Finally, Subsection 3.2.3 provides a complete axiomatization.

3.2.1 Probabilistic Kripke Models

Consider a finite set I of agents, and a countable set Prop of atomic propositions. Throughout this chapter, these sets will be kept fixed, so they will often be left implicit.

Definition 3.1. A *probabilistic Kripke frame* is a tuple $\mathbb{F} = \langle W, R_i, \mu_i \rangle_{i \in I}$, where W is a non-empty finite set of states, $R_i \subseteq W \times W$ is agent i's epistemic accessibility relation, and $\mu_i \colon W \to (W \rightharpoonup [0,1])$ assigns to each state $w \in W$ a partial function $\mu_i(w) \colon W \rightharpoonup [0,1]$, such that

$$\sum_{v \in \mathrm{dom}(\mu_i(w))} \mu_i(w)(v) = 1.$$

Definition 3.2. A *probabilistic Kripke model* is a tuple $\mathbb{M} = \langle \mathbb{F}, V \rangle$, where $\mathbb{F}$ is a probabilistic Kripke frame (with set of states W), and $V \colon \mathsf{Prop} \to \wp(W)$ is a valuation.

Note that in principle, no conditions are imposed on the agents' epistemic accessibility relations. However, as is usually done in the literature on (probabilistic) dynamic epistemic logic, we will henceforth assume these relations to be *equivalence relations* (so that the corresponding knowledge operators satisfy the principles of the modal logic $\mathsf{S5}$).

The function $\mu_i(w)$ represents agent i's probabilities (i.e. degrees of belief) at state w. For example, $\mu_i(w)(v) = k$ means that at state w, agent i assigns probability k to state v being the actual state. From a mathematical perspective, this is not the most general approach: one can also define a probability space[3] $\mathbb{P}_i(w) = \langle S_i(w), \mathcal{A}_i(w), \mu_i(w) \rangle$ for each agent i and state w, and let $\mu_i(w)$ assign probabilities to sets in the σ-algebra $\mathcal{A}_i(w)$, rather than to individual states in the sample space $S_i(w)$. In this way, one can easily drop the requirement that frames and models have finitely many states. This approach is taken in Fagin and Halpern (1994) for static probabilistic epistemic logic, and extended to dynamic settings in Sack (2009); see Demey and Sack (forthcoming) for an extensive overview. However, all the characteristic features of probabilistic (dynamic) epistemic logic already arise in the simpler approach. Therefore, in the remainder of this thesis, I will stick to the simpler approach, and take $\mu_i(w)$ to assign probabilities to individual states. These functions are additively extended from individual states to sets of states, by putting for each set $X \subseteq \mathrm{dom}(\mu_i(w))$:[4]

$$\mu_i(w)(X) := \sum_{x \in X} \mu_i(w)(x).$$

A consequence of the simple approach is that all sets $X \subseteq \mathrm{dom}(\mu_i(w))$ have a definite probability $\mu_i(w)(X)$, whereas in the more general approach, sets X not belonging to the σ-algebra $\mathcal{A}_i(w)$ of $\mathbb{P}_i(w)$ are not assigned any definite probability at all. A similar distinction can be made at the level of individual states. Because $\mu_i(w)$ is a partial function, states $v \in W - \mathrm{dom}(\mu_i(w))$ are not assigned any definite probability at all. An even simpler approach involves

[3] Recall Definition 2.8 on p. 63.

[4] In Section 2.6, I wrote p for the point-function and p^+ for its additive lifting to a set-function. Henceforth, I will no longer notationally distinguish between these two. This should cause no confusion, since they obviously agree on single states: $p(w) = p^+(\{w\})$.

putting $\mu_i(w)(v) = 0$, rather than leaving it undefined. In this way, the function $\mu_i(w)$ can be assumed to be total after all (i.e. $\mathrm{dom}(\mu_i(w)) = W$). From a mathematical perspective, these two approaches are equivalent. From an informal perspective, however, there is a clear difference: $\mu_i(w)(v) = 0$ means that agent i is certain (at state w) that v is not the actual state, whereas $\mu_i(w)(v)$ being undefined means that agent i has no opinion whatsoever (at state w) about v being the actual state. Again, because all the characteristic features of probabilistic (dynamic) epistemic logic already arise without this intuitive distinction, I will opt for the even simpler approach, and henceforth assume that all probability functions are total.

To summarize: the approach adopted in this chapter (and in the remainder of this thesis) is the simplest one possible, in the sense that definite probabilities are assigned to 'everything': (i) to all *sets* (there is no σ-algebra to rule out some sets from having a definite probability), and (ii) to all *states* (the probability functions $\mu_i(w)$ are total on their domain W, so no states are ruled out from having a definite probability).

3.2.2 Language and Semantics

The language $\mathcal{L}^s(\mathsf{Prop})$ of (static) probabilistic epistemic logic is defined by means of the following BNF:

$$\varphi \ ::= \ p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_i\varphi \mid a_1P_i(\varphi) + \cdots + a_nP_i(\varphi) \geq b$$

where $p \in \mathsf{Prop}, i \in I, 1 \leq n < \omega$, and $a_1, \ldots, a_n, b \in \mathbb{Q}$. As usual, we only allow rational numbers as values for $a_1, \ldots, a_n, b$ in order to keep the language countable. The formula $K_i\varphi$ expresses that agent i knows that φ, or, more generally, that φ follows from agent i's information. Its dual is defined as $\hat{K}_i\varphi := \neg K_i\neg\varphi$, and means that φ is consistent with agent i's information.

Formulas of the form $a_1P_i(\varphi_1)+\cdots+a_nP_i(\varphi_n) \geq b$ are called *i-probability formulas*. An i-probability formula is said to be *atomic* iff it is of the form $a_1P_i(p_1) + \cdots + a_nP_i(p_n) \geq b$ with $p_1, \ldots, p_n \in \mathsf{Prop}$, i.e. iff the arguments of its probability operators are propositional atoms (rather than arbitrary formulas). Note that mixed agent indices are not allowed; for example, $P_a(p) + P_b(q) \geq b$ is *not* a well-formed formula of $\mathcal{L}^s$. Intuitively, $P_i(\varphi) \geq b$ means that agent i assigns probability at least b to φ. We allow for linear combinations in i-probability formulas, because this additional expressivity is useful when looking for a complete axiomatization (Fagin and Halpern 1994), and because it allows us to express comparative judgments such as 'agent i considers φ to be at least twice as probable as ψ': $P_i(\varphi) \geq 2P_i(\psi)$. This last formula is actually an abbreviation for $P_i(\varphi) - 2P_i(\psi) \geq 0$. In general, we introduce the following abbreviations:[5]

$$\begin{array}{lll}
\sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) \geq b & \text{for} & a_1 P_i(\varphi_1) + \cdots + a_n P_i(\varphi_n) \geq b, \\
a_1 P_i(\varphi_1) \geq a_2 P_i(\varphi_2) & \text{for} & a_1 P_i(\varphi_1) + (-a_2) P_i(\varphi_2) \geq 0, \\
\sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) \leq b & \text{for} & \sum_{\ell=1}^{n} (-a_\ell) P_i(\varphi_\ell) \geq -b, \\
\sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) < b & \text{for} & \neg\big(\sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) \geq b\big), \\
\sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) > b & \text{for} & \neg\big(\sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) \leq b\big), \\
\sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) = b & \text{for} & \sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) \geq b \wedge \sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) \leq b.
\end{array}$$

Note that because of its recursive definition, the language $\mathcal{L}^s$ can express the agents' higher-order information of any sort: higher-order knowledge (for example $K_a K_b \varphi$), but also higher-order probabilities (for example $P_a(P_b(\varphi) \geq 0.5) = 1$), and higher-order information that mixes knowledge and probabilities (for example, $K_a(P_b(\varphi) \geq 0.5)$ and $P_a(K_b \varphi) = 1$).

The formal semantics for $\mathcal{L}^s$ is defined as follows. Consider an arbitrary probabilistic Kripke model $\mathbb{M}$ and a state w in $\mathbb{M}$. We will often abbreviate $[\![\, \varphi \,]\!]^{\mathbb{M}} := \{v \in W \mid \mathbb{M}, v \models \varphi\}$. Then:

$$\begin{array}{lll}
\mathbb{M}, w \models p & \text{iff} & w \in V(p), \\
\mathbb{M}, w \models \neg\varphi & \text{iff} & \mathbb{M}, w \not\models \varphi, \\
\mathbb{M}, w \models \varphi \wedge \psi & \text{iff} & \mathbb{M}, w \models \varphi \text{ and } \mathbb{M}, w \models \psi, \\
\mathbb{M}, w \models K_i \varphi & \text{iff} & \text{for all } v\text{: if } (w, v) \in R_i \text{ then } \mathbb{M}, v \models \varphi, \\
\mathbb{M}, w \models \sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) \geq b & \text{iff} & \sum_{\ell=1}^{n} a_\ell \mu_i(w)([\![\, \varphi_\ell \,]\!]^{\mathbb{M}}) \geq b.
\end{array}$$

The following notions are also defined in the standard way:

$$\begin{array}{lll}
\mathbb{M} \models \varphi & \text{iff} & \mathbb{M}, w \models \varphi \text{ for all } w \in W, \\
\mathbb{F} \models \varphi & \text{iff} & \langle \mathbb{F}, V \rangle \models \varphi \text{ for all valuations } V \text{ on the frame } \mathbb{F}, \\
\models \varphi & \text{iff} & \mathbb{F} \models \varphi \text{ for all frames } \mathbb{F}.
\end{array}$$

[5] These are essentially the same abbreviations as for propositional probabilistic logic (Subsection 2.6.2), with the exception that the probability operators now carry agent indices.

I will now discuss some typical principles about the interaction between knowledge and probability, and show how they correspond to various properties of probabilistic relational frames.[6]

Definition 3.3. Let $\mathbb{F} = \langle W, R_i, \mu_i \rangle_{i \in I}$ be a probabilistic Kripke frame.

1. $\mathbb{F}$ is *uniform* iff for all states w, v: if $(w,v) \in R_i$ then $\mu_i(w) = \mu_i(v)$,
2. $\mathbb{F}$ is *consistent* iff for all states w, v: if $(w,v) \notin R_i$ then $\mu_i(w)(v) = 0$,
3. $\mathbb{F}$ is *prudent* iff for all states w, v: if $(w,v) \in R_i$, then $\mu_i(w)(v) > 0$,
4. $\mathbb{F}$ is *live* iff for all states w: $\mu_i(w)(w) > 0$.

Lemma 3.1. *Let $\mathbb{F}$ be a probabilistic Kripke frame. Then the following hold:*

1. *$\mathbb{F}$ is uniform iff for all atomic i-probability formulas φ:*
$$\mathbb{F} \models \big(\varphi \to K_i\varphi\big) \wedge \big(\neg\varphi \to K_i\neg\varphi\big),$$
2. *$\mathbb{F}$ is consistent iff $\mathbb{F} \models K_ip \to P_i(p) = 1$,*
3. *$\mathbb{F}$ is prudent iff $\mathbb{F} \models \hat{K}p \to P_i(p) > 0$,*
4. *$\mathbb{F}$ is live iff $\mathbb{F} \models p \to P_i(p) > 0$.*

Uniformity asserts that the agents' probabilities are entirely determined by their epistemic information: if an agent cannot epistemically distinguish between two states, then she should have the same probability functions at those states. This property corresponds to an epistemic-probabilistic introspection principle, stating that agents know their own probabilistic setup (i.e. probability formulas and their negations).

Consistency asserts that the agents assign probability 0 to all states that they do not consider possible. This seems rational: if an agent knows that a certain state is not actual, then it would be a 'waste' to assign any non-zero probability to it. This property corresponds to the principle that knowledge implies certainty (i.e. probability 1). From a more conceptual perspective, consistency yields the following lemma:[7]

[6] See Halpern (2003) for a further discussion of these and other properties, and their correspondence to knowledge/probability interaction principles. Furthermore, from a technical perspective, Lemma 3.1 illustrates how the notion of *frame correspondence* from modal logic (van Benthem 1983, 2001a) can be extended into the probabilistic realm.

[7] I make use of the standard abbreviation $R_i[w] := \{v \in W \mid (w,v) \in R_i\}$.

Lemma 3.2. *Let $\mathbb{F} = \langle W, R_i, \mu_i, V\rangle_{i\in I}$ be an arbitrary probabilistic Kripke frame. If $\mathbb{F}$ satisfies consistency, then for all $w \in W$ and $X \subseteq W$ it holds that*

$$\mu_i(w)(X) = \mu_i(w)(X \cap R_i[w]).$$

Although this lemma is almost trivial to prove, its conceptual importance cannot be underestimated. According to Blackburn et al. (2001, p. ix), one of the key properties of modal (and thus epistemic) logic is its *locality*: to find out whether a formula holds at a given state w, one only needs to check whether it holds at states that are accessible from w. In other words, only states inside $R_i[w]$ are relevant for determining the truth values of propositions at w. In general, probabilistic epistemic logic is *not* local in this sense. For example, to check whether $\mathbb{M}, w \models P_i(q) \geq 0.4$, we need to check whether $\mu_i(w)([\![q]\!]^{\mathbb{M}}) = \sum_{v\,:\,\mathbb{M},v\models q} \mu_i(w)(v) \geq 0.4$, which requires checking *all* states of $\mathbb{M}$ (for each state v, check whether $\mathbb{M}, v \models q$, and if so, add $\mu_i(w)(v)$ to the sum). If (the frame underlying) $\mathbb{M}$ satisfies consistency, however, then Lemma 3.2 states that we only have to check the states inside $R_i[w]$ when calculating this sum. In general: if we only work with frames that satisfy consistency, then probabilistic epistemic logic *is* local, just as modal logic.

I now turn to the next frame property: prudence. This asserts that the agents assign non-zero probability to all states that are epistemically indistinguishable from the actual state. After all, it would be quite 'bold' for an agent to assign probability 0 to a state that, to the best of her knowledge, might turn out to be the actual state.[8] This property corresponds to the principle that epistemic possibility implies probabilistic possibility (non-zero probability), or, read contrapositively, that an agent can only assign probability 0 to propositions that she knows to be false:

$$P_i(p) = 0 \to K_i \neg p.$$

Yet another, equivalent formulation is that an agent can only be certain of propositions that she knows to be true:

$$P_i(p) = 1 \to K_i p.$$

[8] However, there also exist counterexamples to this prudence principle. Kooi (2003, p. 384) gives the example of tossing a fair coin an infinite number of times. A state in which the coin lands tails every time is epistemically possible (we can perfectly imagine that this would happen), yet probabilistically impossible (it seems perfectly reasonable to assign probability 0 to it).

This last formula is exactly the converse of the formula corresponding to consistency, thus revealing the close connection between prudence and consistency.[9]

Liveness asserts that agents assign non-zero probability to the actual state. If one assumes that each state is indistinguishable from itself (i.e. that the epistemic indistinguishability relation R_i is reflexive), then liveness is a direct consequence of prudence. Furthermore, liveness corresponds to the principle that agents should assign non-zero probability to all true propositions ($p \to P_i(p) > 0$). Note, trivially perhaps, that if one assumes that knowledge is factive (which is exactly the principle corresponding to the reflexivity of R_i), then this principle follows immediately from the principle corresponding to prudence.

The properties of consistency and liveness seem particularly plausible. Furthermore, these properties will return often in Part II. It is therefore useful to introduce a separate name for probabilistic Kripke frames satisfying these two properties:

Definition 3.4. A probabilistic Kripke frame is said to be *well-behaved* iff it is uniform and live. A probabilistic Kripke model $\langle \mathbb{F}, V\rangle$ is said to be well-behaved iff its underlying frame $\mathbb{F}$ is.

3.2.3 Proof System

Probabilistic epistemic logic can be axiomatized in a highly modular fashion. An overview is given in Figure 3.1. The propositional, probabilistic and linear inequalities components are exactly as in propositional probabilistic logic (Figure 2.6.4 on p. 70), and should thus not need any further comments.[10] The epistemic component ensures that K_i is an $\mathsf{S5}$ modal operator.

[9] It should be noted that if a probabilistic Kripke frame $\langle W, R_i, \mu_i\rangle_{i\in I}$ satisfies consistency as well as prudence, then for all states $w, v \in W$, it holds that $(w, v) \in R_i$ iff $\mu_i(w)(v) > 0$. This means that the relation R_i is definable in terms of the probability function μ_i, and can thus be dropped from the models altogether. The resulting structures are of the form $\langle W, \mu_i, V\rangle_{i\in I}$, and are essentially a type of *probabilistic transition systems*, which have been extensively studied in theoretical computer science (Larsen and Skou 1991, de Vink and Rutten 1999, Jonsson et al. 2001).

[10] The only differences are that the probability operators now carry agent indices, and, more importantly, that formulas 'inside' $P_i(\,\cdot\,)$ are no longer restricted to be Boolean combinations of propositional atoms. The fact that removing this restriction does not cause any trouble for the axiomatization seems to suggest that it was not needed in the first place.

Figure 3.1: Componentwise axiomatization of probabilistic epistemic logic.

1. propositional component
   - all propositional tautologies and the modus ponens rule
2. epistemic component
   - the S5 axioms and rules for the K_i-operators
3. probabilistic component
   - $P_i(\varphi) \geq 0$
   - $P_i(\top) = 1$
   - $P_i(\varphi \wedge \psi) + P_i(\varphi \wedge \neg\psi) = P_i(\varphi)$
   - if $\vdash \varphi \leftrightarrow \psi$ then $\vdash P_i(\varphi) = P_i(\psi)$
4. linear inequalities component
   - $\sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) \geq b \leftrightarrow \sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) + 0P_i(\varphi_{n+1}) \geq b$
   - $\sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) \geq b \leftrightarrow \sum_{\ell=1}^{n} a_{p(\ell)} P_i(\varphi_{p(\ell)}) \geq b$
     (for any permutation p of $1, \ldots, n$)
   - $\sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) \geq b \wedge \sum_{\ell=1}^{n} a'_\ell P_i(\varphi_\ell) \geq b' \rightarrow$
     $\sum_{\ell=1}^{n} (a_\ell + a'_\ell) P_i(\varphi_\ell) \geq b + b'$
   - $\sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) \geq b \leftrightarrow \sum_{\ell=1}^{n} d a_\ell P_i(\varphi_\ell) \geq db$ (for any $d > 0$)
   - $\sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) \geq b \vee \sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) \leq b$
   - $\sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) \geq b \rightarrow \sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) > b'$ (for any $b' < b$)

The logic's soundness and completeness can be proved using standard techniques (Fagin and Halpern 1994):

Theorem 3.1. *Probabilistic epistemic logic, as axiomatized in Figure 3.1, is sound and complete with respect to the class of probabilistic Kripke models, and also with respect to the class of well-behaved probabilistic Kripke models.*

The proof of this theorem involves constructing a canonical model, i.e. a probabilistic Kripke model in which every consistent formula is satisfiable. By construction, this model is well-behaved, and thus immediately yields completeness with respect to the subclass of well-behaved models as well.

3.3 Probabilistic Public Announcement Logic

In this section, I will discuss a first 'dynamification' of probabilistic epistemic logic, by introducing public announcements into the logic. Subsection 3.3.1 discusses updated probabilistic Kripke models, and introduces a public announcement operator into the formal language to talk about these models. Subsection 3.3.2 provides a complete axiomatization, and Subsection 3.3.3 focuses on the role of higher-order information in public announcement dynamics.

3.3.1 Semantics

Public announcements form one of the simplest types of epistemic dynamics. They concern the truthful and public announcement of some piece of information φ by an external source. That the announcement is *truthful* means that the announced information φ has to be true; that it is *public* means that all agents $i \in I$ learn about it simultaneously and commonly. Finally, the announcement's source is called 'external' because it is not one of the agents $i \in I$ (and will thus not be explicitly represented in the formal language).

Public announcement logic (Plaza 1989, Gerbrandy and Groeneveld 1997, van Ditmarsch et al. 2007) represents these announcements as updates that change Kripke models, and introduces a dynamic public announcement operator into the formal language to describe these updated models. This strategy can straightforwardly be extended into the probabilistic realm.

Syntactically, we add a dynamic operator $[!\cdot]\cdot$ to the static language $\mathcal{L}^s(\mathsf{Prop})$, thus obtaining the new language $\mathcal{L}^!(\mathsf{Prop})$. The formula $[!\varphi]\psi$ means that after any truthful public announcement of φ, it will be the case that ψ. Its dual is defined as $\langle !\varphi\rangle\psi := \neg[!\varphi]\neg\psi$, and means that φ can truthfully and publicly be announced, and afterwards ψ will be the case. These formulas thus allow us to express 'now' (i.e. *before* any dynamics has taken place) what will be the case 'later' (*after* the dynamics has taken place). These formulas are interpreted on a probabilistic Kripke model $\mathbb{M}$ and state w as follows:

$$\begin{array}{lll} \mathbb{M}, w \models [!\varphi]\psi & \text{iff} & \text{if } \mathbb{M}, w \models \varphi \text{ then } \mathbb{M}|\varphi, w \models \psi, \\ \mathbb{M}, w \models \langle !\varphi\rangle\psi & \text{iff} & \mathbb{M}, w \models \varphi \text{ and } \mathbb{M}|\varphi, w \models \psi. \end{array}$$

Note that these clauses mention not only the model $\mathbb{M}$, but also the updated model $\mathbb{M}|\varphi$. The model $\mathbb{M}$ represents the agents' information *before* the public announcement of φ; the model $\mathbb{M}|\varphi$ represents their information *after* the public announcement of φ. Hence, the public announcement of φ *itself* is represented by the update mechanism $\mathbb{M} \mapsto \mathbb{M}|\varphi$, which is formally defined as follows:

Definition 3.5. Consider a probabilistic Kripke model $\mathbb{M} = \langle W, R_i, \mu_i, V\rangle_{i\in I}$ and a formula $\varphi \in \mathcal{L}^!$. Then the *updated probabilistic Kripke model* $\mathbb{M}|\varphi := \langle W^\varphi, R_i^\varphi, \mu_i^\varphi, V^\varphi\rangle_{i\in I}$ is defined as follows:

- $W^\varphi := W$,
- $R_i^\varphi := R_i \cap (W \times [\![\varphi]\!]^{\mathbb{M}})$ for every agent $i \in I$,
- $\mu_i^\varphi \colon W^\varphi \to (W^\varphi \to [0,1])$ is defined (for every agent $i \in I$) by

$$\mu_i^\varphi(w)(v) := \begin{cases} \frac{\mu_i(w)(\{v\}\cap[\![\varphi]\!]^{\mathbb{M}})}{\mu_i(w)([\![\varphi]\!]^{\mathbb{M}})} & \text{if } \mu_i(w)([\![\varphi]\!]^{\mathbb{M}}) > 0 \\ \mu_i(w)(v) & \text{if } \mu_i(w)([\![\varphi]\!]^{\mathbb{M}}) = 0, \end{cases}$$

- $V^\varphi := V$.

The main effect of the public announcement of φ in a model $\mathbb{M}$ is that all links to $\neg\varphi$-states are deleted; hence these states are no longer accessible for any of the agents. Also note that the public announcement of φ does not change the 'ground facts': the valuation V^φ is the same as the old valuation V.[11] This procedure is standard; I will therefore focus on the probabilistic components μ_i^φ.

[11] Public announcements can thus cause changes (e.g. going from not-knowing that φ to knowing that φ), but these changes are always on the epistemic/probabilistic level, never on the ontic level. Informally: public announcements (can) change the agents' (epistemic and probabilistic) *information about* the world, but they do not change the world *itself*.

First of all, it should be noted that the case distinction in the definition of $\mu_i^{\varphi}(v)(u)$ is made for strictly technical reasons, viz. to ensure that there are no 'dangerous' divisions by 0. In all examples and applications, we will be using the 'interesting' case $\mu_i(v)([\![\varphi]\!]^{\mathbb{M}}) > 0$. Still, for general theoretical reasons, *something* has to be said about the case $\mu_i(v)([\![\varphi]\!]^{\mathbb{M}}) = 0$. Leaving $\mu_i^{\varphi}(v)(u)$ undefined would lead to truth value gaps in the logic, and thus greatly increase the difficulty of finding a complete axiomatization. The approach taken here is to define $\mu_i^{\varphi}(v)(u)$ simply as $\mu_i(v)(u)$ in case $\mu_i(v)([\![\varphi]\!]^{\mathbb{M}}) = 0$—so the public announcement of φ has *no effect* whatsoever on $\mu_i(v)$. The intuitive idea behind this definition is that an agent i simply *ignores* new information if she previously assigned probability 0 to it. Technically speaking, this definition will yield a relatively simple axiomatization.

One can easily check that if $\mathbb{M}$ is a probabilistic Kripke model, then $\mathbb{M}|\varphi$ is a probabilistic Kripke model as well. We focus on $\mu^{\varphi}(v)$ (for some arbitrary state $v \in W^{\varphi}$). If $\mu_i(v)([\![\varphi]\!]^{\mathbb{M}}) = 0$, then $\mu_i^{\varphi}(v)$ is $\mu_i(v)$, which is a probability function on $W = W^{\varphi}$. If $\mu_i(v)([\![\varphi]\!]^{\mathbb{M}}) > 0$, then for any $u \in W^{\varphi}$,

$$\mu_i^{\varphi}(v)(u) = \frac{\mu_i(v)(\{u\} \cap [\![\varphi]\!]^{\mathbb{M}})}{\mu_i(v)([\![\varphi]\!]^{\mathbb{M}})},$$

which is positive because $\mu_i(v)(\{u\} \cap [\![\varphi]\!]^{\mathbb{M}})$ is positive, and at most 1, because $\mu_i(v)(\{u\} \cap [\![\varphi]\!]^{\mathbb{M}}) \leq \mu_i(v)([\![\varphi]\!]^{\mathbb{M}})$—and hence $\mu_i^{\varphi}(v)(u) \in [0,1]$. Furthermore,

$$\sum_{u \in W^{\varphi}} \mu_i^{\varphi}(v)(u) = \sum_{u \in W} \frac{\mu_i(v)(\{u\} \cap [\![\varphi]\!]^{\mathbb{M}})}{\mu_i(v)([\![\varphi]\!]^{\mathbb{M}})} = \sum_{\substack{u \in W: \\ \mathbb{M},u \models \varphi}} \frac{\mu_i(v)(u)}{\mu_i(v)([\![\varphi]\!]^{\mathbb{M}})} = 1.$$

It should be noted that the definition of $\mu_i^{\varphi}(v)$—in the interesting case when $\mu_i(v)([\![\varphi]\!]^{\mathbb{M}}) > 0$—can also be expressed in terms of conditional probabilities:

$$\mu_i^{\varphi}(v)(u) = \frac{\mu_i(v)(\{u\} \cap [\![\varphi]\!]^{\mathbb{M}})}{\mu_i(v)([\![\varphi]\!]^{\mathbb{M}})} = \mu_i(v)(u \mid [\![\varphi]\!]^{\mathbb{M}}).$$

In general, for any $X \subseteq W^{\varphi}$ we have:

$$\mu_i^{\varphi}(v)(X) = \frac{\mu_i(v)(X \cap [\![\varphi]\!]^{\mathbb{M}})}{\mu_i(v)([\![\varphi]\!]^{\mathbb{M}})} = \mu_i(v)(X \mid [\![\varphi]\!]^{\mathbb{M}}).$$

In other words, after the public announcement of a formula φ, the agents calculate their new, updated probabilities by means of *Bayesian conditionalization* on the information provided by the announced formula φ. This connection between public announcements and Bayesian conditionalization will be explored more thoroughly in Subsection 3.3.3.

If the probabilistic Kripke model $\mathbb{M}$ is well-behaved (recall Definition 3.4), the updated model $\mathbb{M}|\varphi$ can be defined in a much simpler way:

Definition 3.6. Consider a probabilistic Kripke model $\mathbb{M} = \langle W, R_i, \mu_i, V\rangle_{i\in I}$ and a formula $\varphi \in \mathcal{L}^!$, and suppose that $\mathbb{M}$ is well-behaved. Then the *updated probabilistic Kripke model* $\mathbb{M}|\varphi := \langle W^\varphi, R_i^\varphi, \mu_i^\varphi, V^\varphi\rangle_{i\in I}$ can be defined as follows:

- $W^\varphi := [\![\varphi]\!]^{\mathbb{M}}$,
- $R_i^\varphi := R_i \cap ([\![\varphi]\!]^{\mathbb{M}} \times [\![\varphi]\!]^{\mathbb{M}})$ for every agent $i \in I$,
- $\mu_i^\varphi(w)(v) := \frac{\mu_i(w)(v)}{\mu_i(w)([\![\varphi]\!]^{\mathbb{M}})}$ for every agent $i \in I$ and states $w, v \in W^\varphi$,
- $V^\varphi(p) := V(p) \cap [\![\varphi]\!]^{\mathbb{M}}$ for every $p \in \mathsf{Prop}$.

Informally, Definition 3.5 formalizes a public announcement of φ by cutting all links to $\neg\varphi$-states, while Definition 3.6 removes these states themselves. Both techniques are well-known in the literature on public announcement logic; in non-probabilistic settings they are equivalent to each other.

Note that μ_i^φ is well-defined (there is no danger of dividing by 0); after all, $\mu_i^\varphi(w)$ is only defined for states $w \in W^\varphi = [\![\varphi]\!]^{\mathbb{M}}$, and since $\mathbb{M}$ is well-behaved and thus satisfies liveness, it follows that $\mu_i(w)([\![\varphi]\!]^{\mathbb{M}}) \geq \mu_i(w)(w) > 0$. Just as with Definition 3.5, one can easily show that the updated model defined in Definition 3.6 is a probabilistic Kripke model. Furthermore, this update mechanism preserves well-behavedness:

Lemma 3.3. *If $\mathbb{M}$ is well-behaved, then $\mathbb{M}|\varphi$ is also well-behaved.*

Proof. We first check consistency. Consider arbitrary states $w, v \in W^\varphi$ and suppose that $(w, v) \notin R_i^\varphi$. Since $R_i^\varphi = R_i \cap (W^\varphi \times W^\varphi)$, it follows that $(w, v) \notin R_i$, and hence (by the consistency of $\mathbb{M}$) $\mu_i(w)(v) = 0$. We thus find that

$$\mu_i^\varphi(w)(v) = \frac{\mu_i(w)(v)}{\mu_i(w)([\![\varphi]\!]^{\mathbb{M}})} = 0.$$

Next, we check liveness. Consider an arbitrary state $w \in W^\varphi$. Since $W^\varphi \subseteq W$, it follows by the liveness of $\mathbb{M}$ that $\mu_i(w)(w) > 0$, and hence

$$\mu_i^\varphi(w)(w) = \frac{\mu_i(w)(w)}{\mu_i(w)([\![\varphi]\!]^{\mathbb{M}})} > 0. \qquad \square$$

To illustrate the naturalness and explanatory power of probabilistic public announcement logic, I will finish this subsection by discussing two examples. Example 3.1 is quite simple, but Example 3.2 is a more complex scenario, viz. the well-known *Monty Hall puzzle*.[12]

Example 3.1. Consider the following scenario. An agent does not know whether p is the case, i.e. she cannot distinguish between p-states and $\neg p$-states. (In fact, p happens to be true.) Furthermore, the agent has no specific reason to think that one state is more probable than any other; therefore it is reasonable for her to assign equal probabilities to all states. This example can be formalized by the following model: $\mathbb{M} = \langle W, R, \mu, V\rangle, W = \{w, v\}, R = W \times W, \mu(w)(w) = \mu(w)(v) = \mu(v)(w) = \mu(v)(v) = 0.5$, and $V(p) = \{w\}$. (We work with only one agent in this example, so agent indices can be dropped.) This model is a faithful representation of the scenario described above. For example, we have:

$$\mathbb{M}, w \models \neg Kp \wedge \neg K\neg p \wedge P(p) = 0.5 \wedge P(\neg p) = 0.5.$$

Now suppose that p is publicly announced (which is indeed possible, since p was assumed to be actually true). By applying Definition 3.5 we obtain the updated model $\mathbb{M}|p$, with $W^p = W, R = \{(w, w)\}$, and

$$\mu^p(w)([\![p]\!]^{\mathbb{M}|p}) = \mu^p(w)(w) = \frac{\mu(w)(\{w\} \cap [\![p]\!]^{\mathbb{M}})}{\mu(w)([\![p]\!]^{\mathbb{M}})} = \frac{\mu(w)(w)}{\mu(w)(w)} = 1.$$

Using this updated model $\mathbb{M}|p$, we find that

$$\mathbb{M}, w \models [!p]\big(Kp \wedge P(p) = 1 \wedge P(\neg p) = 0\big).$$

So after the public announcement of p, the agent has come to know that p is in fact the case. She has also adjusted her probabilities: she now assigns probability

[12] The Monty Hall puzzle has caused a large intellectual stir, the details of which cannot be recounted here. More information about this fascinating history can be found in the introduction of Kooi (1999) and Rosenhouse (2009). My presentation of the puzzle follows Kooi (2003, p. 402ff.).

1 to p being true, and probability 0 to p being false. These are the results that one would intuitively expect, so Definition 3.5 seems to yield an adequate representation of the epistemic and probabilistic effects of public announcements.

Example 3.2. The *Monty Hall puzzle* is about the following scenario:

> Suppose you're on a game show, and you're given the choice of three doors. Behind one door is a car, behind the others, goats. You pick a door, say number 1, and the host [Monty Hall], who knows what's behind the doors opens another door, say number 3, which has a goat. He says to you, "Do you want to pick door number 2?" Is it to your advantage to switch your choice of doors?
>
> —Craig F. Whitaker, as cited in Kooi (2003, p. 402).

It turns out that it is indeed advantageous to switch, because of the following argument:

> Suppose you initially pick the door with the car, then you should not switch. This happens in one third of the cases. Suppose on the other hand you initially pick a door that contains a goat, which happens in two third of the cases. Monty Hall cannot open the door with the car and he cannot open the door you picked. He has to open the other door with a goat. So, if you pick a door with a goat, Monty Hall has only one option. After he opens that door, the remaining unopened door you did not pick must contain the car. Therefore, if you initially pick a door with a goat, switching will guarantee that you win the car. You pick such a door in two third of the cases. Hence by switching you lose in one third of the cases and you win in two third of the cases. Kooi (2003, p. 403).

This informal argument—and especially its conclusion—seems highly counterintuitive to many people, and is therefore very controversial (recall Footnote 12). It is therefore useful to formalize the argument in a formal logical system. Since the argument involves probabilities as well as dynamics (such as 'opening a door'), the formal system needs to be able to express and reason about all of these aspects. Probabilistic public announcement logic is exactly such a system; we will now show how it can be used to formalize the argument.

There are two agents: the participant p and the quiz master q. For $1 \leq i \leq 3$, the propositional atoms $\mathsf{car}_i, \mathsf{choose}_i$ and open_i express that the car is behind door i, that the participant initially chooses door i, and that the quiz master opens door i, respectively. The rules of the game state that there is exactly one car, that the participant initially chooses exactly one door, and that the quiz master opens exactly one door. These can be formalized as follows ($\veebar$ expresses exclusive disjunction):

$$\begin{aligned} \mathsf{onecar} &:\equiv \mathsf{car}_1 \veebar \mathsf{car}_2 \veebar \mathsf{car}_3; \\ \mathsf{onechoice} &:\equiv \mathsf{choose}_1 \veebar \mathsf{choose}_2 \veebar \mathsf{choose}_3; \\ \mathsf{oneopen} &:\equiv \mathsf{open}_1 \veebar \mathsf{open}_2 \veebar \mathsf{open}_3. \end{aligned}$$

Although the informal argument does not mention this explicitly, it relies on the (very natural) assumption that the participant initially assigns the same probability to the car being behind each particular door, viz. $\frac{1}{3}$. Furthermore, after the participant has chosen a door (but *before* the quiz master has opened another door), she has not gained any additional information about the car's location, and should thus still assign probability $\frac{1}{3}$ to it being behind each particular door. These two assumptions are formalized as follows:

$$\begin{aligned} \mathsf{equal} &:\equiv P_p(\mathsf{car}_1) = \frac{1}{3} \wedge P_p(\mathsf{car}_2) = \frac{1}{3} \wedge P_p(\mathsf{car}_3) = \frac{1}{3}; \\ \mathsf{stillequal} &:\equiv [!\mathsf{choose}_1]\mathsf{equal} \wedge [!\mathsf{choose}_2]\mathsf{equal} \wedge [!\mathsf{choose}_3]\mathsf{equal}. \end{aligned}$$

Next, consider the way in which the quiz master decides which door to open. If the participant has chosen door i, then the quiz master should open a door j such that (i) the car is not behind it and (ii) it is another door than the one initially chosen by the participant (i.e. $i \neq j$). If there is a goat behind door i, then there is exactly one such door j; however, if the car is behind door i, then there are *two* such doors, and the quiz master must (arbitrarily) select exactly one of them as door j. This selection rule can be formalized as follows:

$$\mathsf{selection} :\equiv \bigwedge_{\substack{1\leq i\leq 3\\ 1\leq j\leq 3}} [!\mathsf{choose}_i]\Big(\mathsf{open}_j \leftrightarrow \big(\neg\mathsf{car}_j \wedge \neg\mathsf{choose}_j \wedge \bigwedge_{\substack{1\leq k\leq 3\\ k\neq j}} \neg\mathsf{open}_k\big)\Big).$$

This completes the formalization of the 'rules' of the game show; let us define

$$\mathsf{rules} :\equiv \mathsf{onecar} \wedge \mathsf{onechoice} \wedge \mathsf{oneopen} \wedge \mathsf{equal} \wedge \mathsf{stillequal} \wedge \mathsf{selection}.$$

It is natural to assume that the participant knows the rules of the game she is participating in (the rules have been carefully explained to her, etc.). As a consequence, she will be fully certain of these rules, i.e. assign probability 1 to them.

We define the *general setup* of the game show as consisting of all its rules *plus* the fact that the participant is certain about these rules:

$$\mathsf{generalsetup} :\equiv \mathsf{rules} \wedge P_p(\mathsf{rules}) = 1.$$

With the game show's general setup in place, let us now turn to the particular scenario described above: the participant chooses door 1 and, subsequently, the quiz master opens door 3, thus revealing a goat. Is it rational for the participant to switch from door 1 to door 2 in this case? She should choose the door which she considers most likely to contain the car. Hence, it will be rational to switch from door 1 to door 2 iff she considers it more likely that the car is behind door 2 than that it is behind door 1. Hence, the question boils down to the following:

$$\mathsf{switch} :\equiv [!\mathsf{choose}_1][!\mathsf{open}_3]\big(P_p(\mathsf{car}_1) < P_p(\mathsf{car}_2)\big).$$

It is natural to assume that the participant does not get radically surprised during the scenario, i.e. that she did not assign probability 0 to any of the propositions involved in the scenario. In this particular scenario, this means that the participant assigns non-zero probability to her choosing door 1, and to the quiz master opening door 3 after she has chosen door 1. These scenario-specific assumptions are naturally expressed as follows:

$$\mathsf{scenario} :\equiv P_p(\mathsf{choose}_1) > 0 \wedge [!\mathsf{choose}_1]P_p(\mathsf{open}_3) > 0.$$

It is now a fairly easy exercise in logical reasoning to show that, given the general setup of the game show and this particular scenario, after the participant has chosen door 1 and the quiz master has opened door 3, the participant assigns probability $\frac{2}{3}$ to the car being behind door 2, and only probability $\frac{1}{3}$ to it being behind door 1:

$$\models \mathsf{setup} \wedge \mathsf{scenario} \;\rightarrow\; [!\mathsf{choose}_1][!\mathsf{open}_3]\big(P_p(\mathsf{car}_1) = \frac{1}{3} \wedge P_p(\mathsf{car}_2) = \frac{2}{3}\big).$$

But this means that the participant should indeed switch:

$$\models \mathsf{setup} \wedge \mathsf{scenario} \;\rightarrow\; \mathsf{switch}.$$

Probabilistic public announcement logic thus formally vindicates the informal argument presented at the beginning of this example, despite its counterintuitive conclusion. Furthermore, because of its logical rigor, it forces us to make all the assumptions of this argument fully explicit, including trivial ones such as equal, but also more subtle ones such as $\mathsf{stillequal}$.

Figure 3.2: Axiomatization of probabilistic public announcement logic.

1. static base logic
   - probabilistic epistemic logic, as axiomatized in Figure 3.1
2. necessitation for public announcement
   - if $\vdash \psi$ then $\vdash [!\varphi]\psi$
3. reduction axioms for public announcement

$$\begin{array}{rcl}
[!\varphi]p & \leftrightarrow & \varphi \rightarrow p \\
[!\varphi]\neg\psi & \leftrightarrow & \varphi \rightarrow \neg[!\varphi]\psi \\
[!\varphi](\psi_1 \wedge \psi_2) & \leftrightarrow & [!\varphi]\psi_1 \wedge [!\varphi]\psi_2 \\
[!\varphi]K_i\psi & \leftrightarrow & \varphi \rightarrow K_i[!\varphi]\psi \\
[!\varphi]\sum_\ell a_\ell P_i(\psi_\ell) \geq b & \leftrightarrow & \varphi \rightarrow \\
 & & \big(P_i(\varphi) = 0 \wedge \sum_\ell a_\ell P_i(\langle !\varphi\rangle\psi_\ell) \geq b\big) \vee \\
 & & \big(P_i(\varphi) > 0 \wedge \sum_\ell a_\ell P_i(\langle !\varphi\rangle\psi_\ell) \geq bP_i(\varphi)\big)
\end{array}$$

3.3.2 Proof System

Public announcement logic can be axiomatized by adding a set of *reduction axioms* to the static base logic (van Ditmarsch et al. 2007). These axioms allow us to recursively rewrite formulas containing dynamic public announcement operators as formulas without such operators; hence the dynamic language $\mathcal{L}^!$ is equally expressive as the static $\mathcal{L}^s$. Alternatively, reduction axioms can be seen as 'predicting' what will be the case after the public announcement has taken place in terms of what is the case before the public announcement has taken place.

This strategy can be extended into the probabilistic realm. For the static base logic, we do not simply take some system of epistemic logic (usually S5), but rather the system of probabilistic epistemic logic described in Subsection 3.2.3 (Figure 3.1), and add the reduction axioms shown in Figure 3.2. The first four reduction axioms are familiar from classical (non-probabilistic) public announcement logic. Note that the reduction axiom for i-probability formulas makes, just like Definition 3.5, a case distinction based on whether the agent assigns probability 0 to the announced formula φ. The significance of this reduction axiom, and its connection with Bayesian conditionalization, will be further explored in the next subsection.

If we restrict attention to well-behaved models, then the reduction axiom for i-probability formulas can be somewhat simplified. If $\mathbb{M}$ is well-behaved and $\mathbb{M}, w \models \varphi$, then also $\mathbb{M}, w \models P_i(\varphi) > 0$, and hence, the reduction axiom for i-probability formulas reduces to

$$[!\varphi] \sum_\ell a_\ell P_i(\psi_\ell) \geq b \ \leftrightarrow\ \Big(\varphi \to \sum_\ell a_\ell P_i(\langle !\varphi\rangle \psi_\ell) \geq b P_i(\varphi)\Big).$$

Once again, standard techniques suffice to prove the following (Kooi 2003):

Theorem 3.2. *Probabilistic public announcement logic, as axiomatized in Figure 3.2, is sound and complete with respect to the class of probabilistic Kripke models, and also with respect to the class of well-behaved probabilistic Kripke models.*

3.3.3 Higher-Order Information in Public Announcements

In this subsection, I will discuss the role of higher-order information in probabilistic public announcement logic. This will further clarify the connection, but also the distinction, between (dynamic versions of) probabilistic epistemic logic and probability theory proper.

In the previous subsection, a reduction axiom for i-probability formulas was introduced. This axiom allows us to derive the following principle as a special case:

$$(\varphi \wedge P_i(\varphi) > 0) \ \longrightarrow\ \big([!\varphi] P_i(\psi) \geq b \leftrightarrow P_i(\langle !\varphi\rangle\psi) \geq b P_i(\varphi)\big). \tag{3.1}$$

The antecedent states that φ is true (because of the truthfulness of public announcements) and that agent i assigns a strictly positive probability to it (so that we are in the 'interesting' case of the reduction axiom). To see the meaning of the consequent more clearly, note that $\langle !\varphi\rangle\psi$ is equivalent to $\varphi \wedge [!\varphi]\psi$, and introduce the following abbreviation of conditional probability into the formal language:

$$P_i(\beta \,|\, \alpha) \geq b \ \ :\equiv\ \ P_i(\alpha \wedge \beta) \geq b P_i(\alpha). \tag{3.2}$$

Principle (3.1) can now be rewritten as follows:

$$(\varphi \wedge P_i(\varphi) > 0) \longrightarrow \big([!\varphi]P_i(\psi) \geq b \leftrightarrow P_i([!\varphi]\psi \,|\, \varphi) \geq b\big). \tag{3.3}$$

A similar version can be proved for $\leq$ instead of $\geq$; combining these two we get:

$$(\varphi \wedge P_i(\varphi) > 0) \longrightarrow \big([!\varphi]P_i(\psi) = b \leftrightarrow P_i([!\varphi]\psi \,|\, \varphi) = b\big). \tag{3.4}$$

The consequent thus states a connection between the agent's probability of ψ after the public announcement of φ, and her conditional probability of $[!\varphi]\psi$, given the truth of φ. In other words, after a public announcent of φ, the agent updates her probabilities by Bayesian conditionalization on φ. The subtlety of principle (3.4), however, is that the agent does not take the conditional probability (conditional on φ) of ψ *itself*, but rather of the *updated* formula $[!\varphi]\psi$.

The reason for this is that $[!\varphi]P_i(\psi) = b$ talks about the probability that the agent assigns to ψ after the public announcement of φ has *actually* happened. If we want to describe this probability as a conditional probability, we cannot simply make use of the conditional probability $P_i(\psi \,|\, \varphi)$, because this represents the probability that the agent *would* assign to ψ if a public announcement of φ *would* happen—hypothetically, not actually! Borrowing a slogan from van Benthem: "The former takes place once arrived at one's vacation destination, the latter is like reading a travel folder and musing about tropical islands." (van Benthem 2003, p. 417). Hence, if we want to describe the agent's probability of ψ after an actual public announcement of φ in terms of conditional probabilities, we need to represent the effects of the public announcement of φ on ψ explicitly, and thus take the conditional probability (conditional on φ) of $[!\varphi]\psi$, rather than simply ψ.

One might wonder about the relevance of this subtle distinction between actual and hypothetical public announcements. The point is that the public announcement of φ can have effects on the truth value of ψ. For large classes of formulas ψ, this will not occur: their truth value is not affected by the public announcement of φ. Formally, this means that $[!\varphi]\psi$ is equivalent to ψ, so (the consequent of) principle (3.4) becomes:

$$[!\varphi]P_i(\psi) = b \leftrightarrow P_i(\psi \,|\, \varphi) = b$$

—thus wiping away all differences between the agent's probability of ψ after a public announcement of φ, and her conditional probability of ψ, given φ. A typical class of such formulas (whose truth value is unaffected by the public announcement of φ) is formed by the Boolean combinations of proposition letters, i.e. those formulas which express *ontic* or *first-order information*. Since probability theory proper is usually only concerned with first-order information ('no nested probabilities'), the distinction between actual and hypothetical announcements—or in general, between actual and hypothetical learning of new information—thus vanishes completely, and Bayesian conditionalization can be used as a universal update rule to compute new probabilities after (actually) learning a new piece of information.

However, in probabilistic epistemic logic (and its dynamic versions, such as probabilistic PAL), higher-order information *is* taken into account, and hence the distinction between actual and hypothetical public announcements has to be taken seriously. Therefore, the consequent of principle (3.4) should really use the conditional probability $P_i([!\varphi]\psi \,|\, \varphi)$, rather than just $P_i(\psi \,|\, \varphi)$.[13]

Example 3.3. To illustrate this, consider again the model defined in Example 3.1, and put $\varphi := p \wedge P(\neg p) = 0.5$. It is easy to show that

$$\mathbb{M}, w \models P(\varphi \,|\, \varphi) = 1 \ \wedge \ P([!\varphi]\varphi \,|\, \varphi) = 0 \ \wedge \ [!\varphi]P(\varphi) = 0.$$

Hence the probability assigned to φ after the public announcement is the conditional probability $P([!\varphi]\varphi \,|\, \varphi)$, rather than just $P(\varphi \,|\, \varphi)$. Note that this example indeed involves higher-order information, since we are talking about the probability of φ, which itself contains the probability statement $P(\neg p) = 0.5$ as a conjunct. Finally, this example also shows that learning a new piece of information φ (via public announcement) does *not* automatically lead to the agents being certain about (i.e. assigning probability 1 to) that formula. This is to be contrasted with probability theory, where a new piece of information φ is processed via Bayesian conditionalization, and thus always leads to certainty: $P(\varphi \,|\, \varphi) = 1$. The explanation is, once again, that probability theory is only concerned with

[13] Romeijn (2012) provides an analysis that stays closer in spirit to probability theory proper. He argues that the public announcement of φ induces a shift in the interpretation of ψ (in our terminology: from ψ to $[!\varphi]\psi$, i.e. from $[\![\, \psi \,]\!]^{\mathbb{M}}$ to $[\![\, \psi \,]\!]^{\mathbb{M}|\varphi}$), and shows that such meaning shifts can be modeled using Dempster-Shafer belief functions. Crucially, however, this proposal is able to deal with the case where the formula $P(\psi) = b$ expresses *second-order* information (e.g. when ψ itself is of the form $P_i(p) = b$), but not with the case of higher-order information *in general* (e.g. when ψ is of the form $P_j(P_i(p) = b) = a$, or involves even more deeply nested probabilities) (Romeijn 2012, p. 603).

first-order information, whereas the phenomena described above can only occur at the level of higher-order information.[14,15]

3.4 Probabilistic Dynamic Epistemic Logic

In this section, I will move from a probabilistic version of public announcement logic to a probabilistic version of 'full' dynamic epistemic logic. Subsection 3.4.1 introduces a probabilistic version of the *product update* mechanism that is behind dynamic epistemic logic. Subsection 3.4.2 introduces dynamic operators into the formal language to talk about these product updates, and discusses a detailed example. Subsection 3.4.3, finally, shows how to obtain a complete axiomatization in a fully standard (though non-trivial) fashion.

3.4.1 Probabilistic Product Update

Classical (non-probabilistic) dynamic epistemic logic models epistemic dynamics by means of a product update mechanism (Baltag and Moss 2004, Baltag et al. 1998). The agents' *static* information (what is the current state?) is represented in a Kripke model $\mathbb{M}$, and their *dynamic* information (what type of event is currently taking place?) is represented in an update model $\mathbb{E}$. The agents' new information (after the dynamics has taken place) is represented by means of a product construction $\mathbb{M} \otimes \mathbb{E}$. I will now show how a probabilistic version of this construction can be set up.

Before stating the formal definitions, I will show how they naturally arise as probabilistic generalizations of the classical (non-probabilistic) notions. The probabilistic Kripke models introduced in Definition 3.2 represent the agents' static information, in both its epistemic and its probabilistic aspects. This static probabilistic information is called the *prior probabilities of the states* in van

[14] Similarly, the *success postulate* for belief expansion in the (traditional) AGM framework (Alchourrón et al. 1985, Gärdenfors 1988) states that after expanding one's belief set with a new piece of information φ, the updated (expanded) belief set should always contain this new information. Here, too, the explanation is that AGM is only concerned with first-order information. (Note that we talk about the success postulate for belief *expansion*, rather than belief *revision*, because the former seems to be the best analogue of public announcement in the AGM framework.)

[15] The occurrence of higher-order information is a *necessary* condition for this phenomenon, but not a *sufficient* one: there exist formulas φ that involve higher-order information, but still $\models [!\varphi]P_i(\varphi) = 1$ (or epistemically: $\models [!\varphi]K_i\varphi$).

Benthem et al. (2009). We can thus say that when w is the actual state, agent i considers it epistemically possible that v is the actual state ($(w, v) \in R_i$), and, more specifically, that she assigns probability b to v being the actual state ($\mu_i(w)(v) = b$).

Update models are essentially like Kripke models: they represent the agents' information about events, rather than states. Since probabilistic Kripke models represent both epistemic and probabilistic information about *states*, by analogy probabilistic update models should represent both epistemic and probabilistic information about *events*. Hence, they should not only have epistemic accessibility relations R_i over their set of events E, but also probability functions $\mu_i \colon E \to (E \to [0, 1])$. (Formal details will be given in Definition 3.7.) We can then say that when e is the actually occurring event, agent i considers it epistemically possible that f is the actually occurring event ($(e, f) \in R_i$), and, more specifically, that she assigns probability b to f being the actually occurring event ($\mu_i(e)(f) = b$). This dynamic probabilistic information is called the *observation probabilities* in van Benthem et al. (2009).

Finally, how probable it is that an event e will occur, might vary from state to state. We assume that this variation can be captured by means of a set Φ of (pairwise inconsistent) sentences in the object language (so that the probability that an event e will occur can only vary between states that satisfy *different* sentences of Φ). This will be formalized by adding to the probabilistic update models a set of preconditions Φ, and probability functions $\mathsf{pre} \colon \Phi \to (E \to [0, 1])$. The meaning of $\mathsf{pre}(\varphi)(e) = b$ is that if φ holds, then event e occurs with probability b. In van Benthem et al. (2009) these are called *occurrence probabilities*.[16]

We are now ready to formally introduce probabilistic update models:

Definition 3.7. A *probabilistic update model* is a tuple $\mathbb{E} = \langle E, R_i, \Phi, \mathsf{pre}, \mu_i \rangle_{i \in I}$, where E is a non-empty finite set of events, $R_i \subseteq E \times E$ is agent i's epistemic accessibility relation, $\Phi \subseteq \mathcal{L}^{\otimes}$ is a finite set of pairwise inconsistent sentences called *preconditions*, $\mu_i \colon E \to (E \to [0, 1])$ assigns to each event $e \in E$ a probability function $\mu_i(e)$ over E, and $\mathsf{pre} \colon \Phi \to (E \to [0, 1])$ assigns to each precondition $\varphi \in \Phi$ a probability function $\mathsf{pre}(\varphi)$ over E.

All components of a probabilistic update model have already been commented upon. Note that we use the same symbols R_i and μ_i to indicate agent i's

[16] Occurrence probabilities are often assumed to be *objective frequencies*. This is reflected in the formal setup: the function pre is not agent-dependent.

epistemic and probabilistic information in a probabilistic Kripke model $\mathbb{M}$ and in a probabilistic update model $\mathbb{E}$—from the context it will always be clear which of the two is meant. The language $\mathcal{L}^{\otimes}$ that the preconditions are taken from will be formally defined in the next subsection. (As is usual in this area, there is a non-vicious simultaneous recursion going on here.)

We now introduce occurrence probabilities for events at states:

Definition 3.8. Consider a probabilistic Kripke model $\mathbb{M}$, a state w, a probabilistic update model $\mathbb{E}$, and an event e. Then the *occurrence probability of* e *at* w is defined as

$$\mathsf{pre}(w)(e) := \begin{cases} \mathsf{pre}(\varphi)(e) & \text{if } \varphi \in \Phi \text{ and } \mathbb{M}, w \models \varphi \\ 0 & \text{if there is no } \varphi \in \Phi \text{ such that } \mathbb{M}, w \models \varphi. \end{cases}$$

Since the preconditions are pairwise inconsistent, $\mathsf{pre}(w)(e)$ is always well-defined. The meaning of $\mathsf{pre}(w)(e) = b$ is that in state w, event e occurs with probability b. Note that if two states w and v satisfy the same precondition, then always $\mathsf{pre}(w)(e) = \mathsf{pre}(v)(e)$; in other words, the occurrence probabilities of an event e can only vary ‘up to a precondition’ (cf. supra).

The probabilistic product update mechanism can now be defined as follows:

Definition 3.9. Consider a probabilistic Kripke model $\mathbb{M} = \langle W, R_i, \mu_i, V\rangle_{i\in I}$ and a probabilistic update model $\mathbb{E} = \langle E, R_i, \Phi, \mathsf{pre}, \mu_i\rangle_{i\in I}$. Then the *updated model* $\mathbb{M} \otimes \mathbb{E} := \langle W', R'_i, \mu'_i, V'\rangle_{i\in I}$ is defined as follows:

- $W' := \{(w, e) \mid w \in W, e \in E, \mathsf{pre}(w)(e) > 0\}$,
- $R'_i := \{((w, e), (w', e')) \in W' \times W' \mid (w, w') \in R_i \text{ and } (e, e') \in R_i\}$ for every agent $i \in I$,
- $\mu'_i \colon W' \to (W' \to [0, 1])$ is defined (for every agent $i \in I$) by

  $$\mu'_i(w, e)(w', e') := \begin{cases} \alpha^{-1} \cdot \mu_i(w)(w') \cdot \mathsf{pre}(w')(e') \cdot \mu_i(e)(e') & \text{if } \alpha > 0 \\ 0 & \text{if } \alpha = 0, \end{cases}$$

  where $\alpha := \sum_{\substack{w''\in W\\ e''\in E}} \mu_i(w)(w'') \cdot \mathsf{pre}(w'')(e'') \cdot \mu_i(e)(e'')$,
- $V'(p) := \{(w, e) \in W' \mid w \in V(p)\}$ for every $p \in \mathsf{Prop}$.

I will only comment on the probabilistic component of this definition (all other components are fully classical). After the dynamics has taken place, agent i calculates at state (w, e) her new probability for (w', e') by taking the arithmetical product of (i) her *prior probability* for w' at w, (ii) the *occurrence probability* of e' in w', and (iii) her *observation probability* for e' at e, and then normalizing this product (i.e. dividing it by α). The factors in this product are not *weighted* (or equivalently, they all have weight 1); van Benthem et al. (2009) also discuss weighted versions of this update mechanism, and show how one of these weighted versions corresponds to the rule of *Jeffrey conditioning* from probability theory (Jeffrey 1983). Finally, note that $\mathbb{M} \otimes \mathbb{E}$ might fail to be a probabilistic Kripke model: if $\alpha = 0$, then $\mu'_i(w, e)$ assigns 0 to all states in W'. We will not care here about the interpretation of this feature, but only remark that technically speaking it is harmless and, perhaps most importantly, still allows for a reduction axiom for i-probability formulas (cf. Subsection 3.4.3).

3.4.2 Language and Semantics

To talk about these updated models, dynamic operators $[\mathsf{E}, \mathsf{e}]$ are added to the static language $\mathcal{L}^s$, thus obtaining the new language $\mathcal{L}^{\otimes}$. Here, E, e are formal names for the probabilistic update model $\mathbb{E} = \langle E, R_i, \Phi, \mathsf{pre}, \mu_i \rangle_{i \in I}$ and event $e \in E$; recall the remark about the mutual recursion of the dynamic language and the updated models. The formula $[\mathsf{E}, \mathsf{e}]\varphi$ says that after the event e has occurred, it will be the case that φ. It has the following semantics:

$$\mathbb{M}, w \models [\mathsf{E}, \mathsf{e}]\psi \qquad \text{iff} \qquad \text{if } \mathsf{pre}(w)(e) > 0, \text{ then } \mathbb{M} \otimes \mathbb{E}, (w, e) \models \psi.$$

I will now illustrate the expressive power of this framework by showing how it can be used to adequately model a rather intricate scenario. Note that this example is based on the sense of *sight*; similar examples in van Benthem et al. (2009) and Demey and Sack (forthcoming) are based on the senses of *touch* and *hearing*, respectively. This is no coincidence: because of the fallibility of sense perception, examples based on the senses can easily be used to illustrate the notion of observation probability.[17]

Example 3.4. Consider the following scenario. While strolling through a flee market, you see a painting that you think might be a real Picasso. Of course, the

[17] Examples based on the senses of smell and taste are eagerly awaited.

chance that the painting is actually a real Picasso is very slim, say 1 in 100 000. You know from an art encyclopedia that Picasso signed almost all his paintings with a very characteristic signature. If the painting is a real Picasso, the chance that it bears the characteristic signature is 97%, while if the painting is not a real Picasso, the chance that it bears the characteristic signature is 0% (nobody is capable of imitating Picasso's signature). You immediately look at the painting's signature, but determining whether it is Picasso's characteristic signature is very hard, and—not being an expert art historian—, you remain uncertain and think that the chance is 50% that the painting's signature is Picasso's characteristic one.

Your initial information (before having looked at the painting's signature) can be represented as the following probabilistic Kripke model: $\mathbb{M} = \langle W, R, \mu, V\rangle$, where $W = \{w, v\}, R = W \times W, \mu(w)(w) = \mu(v)(w) = 0.00001, \mu(w)(v) = \mu(v)(v) = 0.99999$, and $V(\mathsf{real}) = \{w\}$. (We work with only one agent in this example, so agent indices can be dropped.) Hence, initially you do not rule out the possibility that the painting in front of you is a real Picasso, but you consider it highly unlikely:

$$\mathbb{M}, w \models \hat{K}\mathsf{real} \wedge P(\mathsf{real}) = 0.00001.$$

The event of looking at the signature can be represented with the following update model: $\mathbb{E} = \langle E, R, \Phi, \mathsf{pre}, \mu\rangle$, where $E = \{e, f\}$, $R = E \times E$, $\Phi = \{\mathsf{real}, \neg\mathsf{real}\}$, $\mathsf{pre}(\mathsf{real})(e) = 0.97$, $\mathsf{pre}(\mathsf{real})(f) = 0.03$, $\mathsf{pre}(\neg\mathsf{real})(e) = 0$, $\mathsf{pre}(\neg\mathsf{real})(f) = 1$, and $\mu(e)(e) = \mu(f)(e) = \mu(e)(f) = \mu(f)(f) = 0.5$. The event e represents 'looking at Picasso's characteristic signature'; the event f represents 'looking at a signature that is not Picasso's characteristic one'.

We now construct the updated model $\mathbb{M}\otimes\mathbb{E}$. Since $\mathbb{M}, v \not\models \mathsf{real}$, it holds that $\mathsf{pre}(v)(e) = \mathsf{pre}(\neg\mathsf{real})(e) = 0$, and hence (v, e) does not belong to the updated model. It is easy to see that the other states (w, e), (w, f) and (v, f) do belong to the updated model. Furthermore, one can easily calculate that $\mu'(w, e)(w, e) = 0.0000003$ and $\mu'(w, e)(w, f) = 0.0000097$, so $\mu'(w, e)([\![\,\mathsf{real}\,]\!]^{\mathbb{M}\otimes\mathbb{E}}) = 0.0000003 + 0.0000097 = 0.00001$, and thus

$$\mathbb{M}, w \models [\mathsf{E}, \mathsf{e}]P(\mathsf{real}) = 0.00001.$$

Hence, even though the painting in front of you is a real Picasso (in state w), after looking at the signature (which is indeed Picasso's characteristic signature!—the event that actually happened was event e) you still assign a probability of 1 in 100 000 to it being a real Picasso.

Note that if you had been an expert art historian, with the same prior probabilities, but with the reliable capability of recognizing Picasso's characteristic signature—let's formalize this as $\mu(e)(e) = 0.99$ and $\mu(e)(f) = 0.01$—, then the same update mechanism would have implied that

$$\mathbb{M}, w \models [\mathsf{E}, \mathsf{e}]P(\mathsf{real}) = 0.00096.$$

In other words, if you had been an expert art historian, then looking at the painting's signature would have been highly informative: it would have led to a significant change in your probabilities.

3.4.3 Proof System

A complete axiomatization for probabilistic dynamic epistemic logic can be found using the standard strategy, viz. by adding a set of reduction axioms to static probabilistic epistemic logic. Implementing this strategy, however, is not entirely trivial. The reduction axioms for non-probabilistic formulas are familiar from classical (non-probabilistic) dynamic epistemic logic, but the reduction axiom for i-probability formulas is more complicated.

First of all, this reduction axiom makes a case distinction on whether a certain sum of probabilities is strictly positive or not. I will now show that this corresponds to the case distinction made in the definition of the updated probability functions (Definition 3.9). In the definition of $\mu_i'(w, e)$, a case distinction is made on the value of the denominator of a fraction, i.e. on the value of the following expression:

$$\sum_{\substack{v \in W \\ f \in E}} \mu_i(w)(v) \cdot \mathsf{pre}(v)(f) \cdot \mu_i(e)(f). \tag{3.5}$$

But this expression can be rewritten as

$$\sum_{\substack{v \in W \\ f \in E \\ \varphi \in \Phi \\ \mathbb{M}, v \models \varphi}} \mu_i(w)(v) \cdot \mathsf{pre}(\varphi)(f) \cdot \mu_i(e)(f).$$

Figure 3.3: Axiomatization of probabilistic dynamic epistemic logic.

1. static base logic
   - probabilistic epistemic logic, as axiomatized in Figure 3.1
2. necessitation for $[\mathsf{E},\mathsf{e}]$
   - if $\vdash \psi$ then $\vdash [\mathsf{E},\mathsf{e}]\psi$
3. reduction axioms

$$\begin{array}{rcl}
[\mathsf{E},\mathsf{e}]p & \leftrightarrow & \mathsf{pre}_{\mathsf{E},\mathsf{e}} \to p \\
[\mathsf{E},\mathsf{e}]\neg\psi & \leftrightarrow & \mathsf{pre}_{\mathsf{E},\mathsf{e}} \to \neg[\mathsf{E},\mathsf{e}]\psi \\
[\mathsf{E},\mathsf{e}](\psi_1 \wedge \psi_2) & \leftrightarrow & [\mathsf{E},\mathsf{e}]\psi_1 \wedge [\mathsf{E},\mathsf{e}]\psi_2 \\
[\mathsf{E},\mathsf{e}]K_i\psi & \leftrightarrow & \mathsf{pre}_{\mathsf{E},\mathsf{e}} \to \bigwedge_{(e,f)\in R_i} K_i[\mathsf{E},\mathsf{f}]\psi \\
[\mathsf{E},\mathsf{e}]\sum_\ell a_\ell P_i(\psi_\ell) \geq b & \leftrightarrow & \mathsf{pre}_{\mathsf{E},\mathsf{e}} \to \\
 & & \big(\sum_{\substack{\varphi\in\Phi \\ f\in E}} k_{i,e,\varphi,f}P_i(\varphi) = 0 \wedge 0 \geq b\big) \ \vee \\
 & & \big(\sum_{\substack{\varphi\in\Phi \\ f\in E}} k_{i,e,\varphi,f}P_i(\varphi) > 0 \wedge \chi\big)
\end{array}$$

   using the following definitions:

   - $\mathsf{pre}_{\mathsf{E},\mathsf{e}} := \bigvee_{\substack{\varphi\in\Phi \\ \mathsf{pre}(\varphi)(e)>0}} \varphi$
   - $k_{i,e,\varphi,f} := \mathsf{pre}(\varphi)(f) \cdot \mu_i(e)(f) \in \mathbb{R}$
   - $\chi := \sum_{\substack{\ell \\ \varphi\in\Phi \\ f\in E}} a_\ell k_{i,e,\varphi,f} P_i(\varphi \wedge \langle\mathsf{E},\mathsf{f}\rangle\psi_\ell) \geq \sum_{\substack{\varphi\in\Phi \\ f\in E}} b k_{i,e,\varphi,f} P_i(\varphi)$

Using the definition of $k_{i,e,\varphi,f}$ (cf. Figure 3.3), this can be rewritten as

$$\sum_{\substack{\varphi\in\Phi\\ f\in E}} \mu_i(w)([\![\,\varphi\,]\!]^{\mathbb{M}}) \cdot k_{i,e,\varphi,f}.$$

Since E and Φ are finite, this sum is finite and corresponds to an expression in the formal language $\mathcal{L}^{\otimes}$, which we will abbreviate as σ:

$$\sigma := \sum_{\substack{\varphi\in\Phi\\ f\in E}} k_{i,e,\varphi,f} P_i(\varphi).$$

This expression can be turned into an i-probability formula by 'comparing' it with a rational number b; for example $\sigma \geq b$. Particularly important are the formulas $\sigma = 0$ and $\sigma > 0$: it are exactly these formulas which are used to make the case distinction in the reduction axiom for i-probability formulas.[18]

Next, the reduction axiom for i-probability formulas provides a statement in each case of the case distinction: $0 \geq b$ in the case $\sigma = 0$, and χ (as defined in Figure 3.3) in the case $\sigma > 0$. We will only discuss the meaning of χ in the 'interesting' case $\sigma > 0$. If $\mathbb{M}, w \models \sigma > 0$, then the value of (3.5) is strictly positive (cf. supra), and we can calculate:

$$\begin{array}{rcl}
\mu_i'(w,e)([\![\,\psi\,]\!]^{\mathbb{M}\otimes\mathbb{E}}) & = & \sum_{\mathbb{M}\otimes\mathbb{E},(w',e')\models\psi} \mu_i'(w,e)(w',e') \\[2ex]
& = & \sum_{\substack{w'\in W, e'\in E\\ \mathbb{M},w'\models\langle\mathsf{E},\mathsf{e}'\rangle\psi}} \frac{\mu_i(w)(w')\cdot\mathsf{pre}(w')(e')\cdot\mu_i(e)(e')}{\sum_{\substack{v\in W\\ f\in E}} \mu_i(w)(v)\cdot\mathsf{pre}(v)(f)\cdot\mu_i(e)(f)} \\[2ex]
& = & \dfrac{\sum_{\substack{\varphi\in\Phi\\ f\in E}} \mu_i(w)([\![\,\varphi\wedge\langle\mathsf{E},\mathsf{f}\rangle\psi\,]\!]^{\mathbb{M}})\cdot k_{i,e,\varphi,f}}{\sum_{\substack{\varphi\in\Phi\\ f\in E}} \mu_i(w)([\![\,\varphi\,]\!]^{\mathbb{M}})\cdot k_{i,e,\varphi,f}}.
\end{array}$$

Hence, in this case ($\sigma > 0$) we can express that $\mu_i'(w,e)([\![\,\psi\,]\!]^{\mathbb{M}\otimes\mathbb{E}}) \geq b$ in the formal language, by means of the following i-probability formula:

$$\sum_{\substack{\varphi\in\Phi\\ f\in E}} k_{i,e,\varphi,f} P_i(\varphi \wedge \langle\mathsf{E},\mathsf{f}\rangle\psi) \geq \sum_{\substack{\varphi\in\Phi\\ f\in E}} b k_{i,e,\varphi,f} P_i(\varphi).$$

[18] Note that E and Φ are components of the probabilistic update model $\mathbb{E}$ named by E; furthermore, the values $k_{i,e,\varphi,f}$ are fully determined by the model $\mathbb{E}$ and event e named by E and e, respectively (consider their definition in Figure 3.3). Hence, any i-probability formula involving the expression σ is fully determined by $\mathbb{E}, e$, and can be interpreted at any probabilistic Kripke model $\mathbb{M}$ and state w.

Moving to linear combinations, we can express that $\sum_\ell a_\ell \mu'_i(w,e)([\![\,\psi_\ell\,]\!]^{\mathbb{M}\otimes\mathbb{E}}) \geq b$ in the formal language using an analogous i-probability formula, namely χ (cf. the definition of this formula in Figure 3.3).

We thus obtain the following theorem (van Benthem et al. 2009):

Theorem 3.3. *Probabilistic dynamic epistemic logic, as axiomatized in Figure 3.3, is sound and complete with respect to the class of probabilistic Kripke models.*

3.5 Further Developments and Applications

Probabilistic extensions of dynamic epistemic logic are a recent development, and there are various open questions and potential applications to be explored. In this section, I discuss a selection of such topics for further research; more suggestions can be found in van Benthem et al. (2009) and van Benthem (2011, Chapter 8).

A typical technical issue is whether other representations of *soft information* can learn something from the probabilistic approach to dynamic epistemic logic. Probabilistic Kripke models represent the agents' soft information via the probability functions μ_i, and interpret formulas of the form $P_i(\varphi) \geq b$. Plausibility models, on the other hand, represent the agents' soft information via a (non-numerical) plausibility ordering $\leq_i$; for example, $w \leq_i v$ means that agent i considers state w at least as plausible as state v (Baltag and Smets 2008, van Benthem 2007, 2011).[19]

The product update for probabilistic Kripke models described in Definition 3.9 takes into account prior probabilities ($\mu_i(w)(v)$ for states w and v), observation probabilities ($\mu_i(e)(f)$ for events e and f), and occurrence probabilities ($\mathsf{pre}(w)(e)$ for a state w and event e). One can also define a product update for plausibility models; a widely used rule to define the updated plausibility ordering is the so-called 'priority update' (Baltag and Smets 2008, van Benthem 2011):

$$(w,e) \leq_i (v,f) \quad \text{iff} \quad e <_i f \text{ or } (e \cong_i f \text{ and } w \leq_i v).$$

The updated plausibility ordering thus gives priority to the plausibility ordering on events, and otherwise preserves the original plausibility ordering on states

[19] The model theory of epistemic plausibility models is discussed in detail in Chapter 4.

as much as possible. In analogy with the probabilistic setting, the plausibility orderings on states and events can be called the 'prior plausibility' and 'observation plausibility', respectively. However, the notion of occurrence probability does *not* seem to have a clear analogue in the framework of plausibility models. Van Benthem (2012) defines a notion of 'occurrence plausibility', which can be expressed as $e \leq_w f$: at state w, event e is at least as plausible as f to occur (this ordering is not agent-dependent; recall Footnote 16). New product update rules thus have to merge *three* plausibility orderings: prior plausibility, observation plausibility, and occurrence plausibility. Van Benthem (2012) makes some proposals for such rules, but finding a fully satisfactory definition remains a major open problem in this area.

Probabilistic extensions of (dynamic) epistemic logic have been fruitfully applied in fields such as game theory and cognitive science. In recent years, epistemic logic has been widely applied to explore the epistemic foundations of game theory (van Benthem 2001b, 2007, Bonanno and Dégremont forthcoming). However, given the importance of probability in game theory (for example, in the notion of mixed strategy), it is quite surprising that rather few of these logical analyses have a probabilistic component. A notable exception is de Bruin (2008a,b, 2010), who uses probabilistic epistemic logic to analyze epistemic characterization theorems for several solution concepts for normal form games and extensive games, such as Nash equilibrium, iterated strict dominance, and backward induction. However, de Bruin's system is entirely *static*. An application of *dynamic* probabilistic epistemic logics in game theory will be discussed in Chapter 5, where I use a version of probabilistic public announcement logic to analyze the roles of communication and common knowledge in Aumann's celebrated agreeing to disagree theorem.

Another potential field of application for probabilistic dynamic epistemic logic is cognitive science. The usefulness of (epistemic) logic for cognitive science has been widely recognized (Pietarinen 2003, van Benthem 2008, Verbrugge et al. forthcoming). Of course, as in any other empirical discipline, one quickly finds out that real-life human cognition is rarely a matter of all-or-nothing, but often involves degrees (probabilities). Furthermore, a recent development in cognitive science is toward probabilistic (Bayesian) models of cognition (Oaksford and Chater 2008). If epistemic logic is to remain a valuable tool here, it will thus have to be a thoroughly 'probabilified' version. For example, Lorini and Castelfranchi (2007) use a version of probabilistic epistemic logic to model the cognitive and epistemic aspects of surprise. In Chapter 7, I will argue that Lorini and Castelfranchi's system fails to fully capture the dynamic nature of surprise, and propose an alternative formalization in the framework of probabilistic public announcement logic.

3.6 Conclusion

In this chapter, I have introduced probabilistic epistemic logic and several of its dynamic versions. These logics provide a standard epistemic (possible worlds) analysis of the agents' hard information, and supplement it with a fine-grained probabilistic analysis of their soft information. Higher-order information of any kind (knowledge about probabilities, probabilities about knowledge, etc.) is represented explicitly. The importance of higher-order information in dynamics is clearly illustrated by the subtle relationship between public announcements and Bayesian conditionalization. The probabilistic versions of both public announcement logic and dynamic epistemic logic with product updates can be completely axiomatized in a standard way (via reduction axioms).

4 The Model Theory of Plausibility Models

4.1 Introduction

Traditional epistemic logic can be seen as a particular branch of modal logic. Its semantics is defined in terms of Kripke models, and philosophical principles about knowledge (e.g. factivity: $K\varphi \to \varphi$) are shown to correspond to properties of the epistemic accessibility relation (e.g. reflexivity). By adding another (doxastic) accessibility relation, *belief* can be treated in this framework as well. Belief is not assumed to be factive, but at least consistent ($\neg B\bot$), which corresponds to requiring the doxastic accessibility relation to be serial instead of reflexive. In this extended framework, one can study the interaction between knowledge and belief, expressed in principles such as $K\varphi \to B\varphi$ (van der Hoek 1993, Halpern 1996). Furthermore, since this framework is still 'just' a modal logic, it inherits the mathematically well-developed model theory of modal logic.

This framework can also be used to model the interaction of (factive) *knowledge* with public announcements (Plaza 1989, Gerbrandy and Groeneveld 1997) and other dynamic epistemic phenomena (Baltag et al. 1998, Baltag and Moss 2004). The dynamics of *belief* (and other non-factive attitudes), however, cannot convincingly be modeled in this framework: if an agent receives a true piece of information φ while previously believing that $\neg\varphi$, then this agent is predicted to go insane and start believing *everything* (rather than performing a realistic process of *belief revision*)—thus contradicting the consistency requirement about belief.[1] To remedy this problem, (epistemic) plausibility models have been in-

[1] More details can be found in van Benthem (2007, Section 3.1).

troduced (technical details will be presented later). In these models, one can again study knowledge, belief (and even other cognitive propositional attitudes), and their various interactions. Furthermore, this framework provides a realistic model of various dynamic phenomena, and thus solves the main problem of the previous approach. Because plausibility models give rise to a much richer semantics than Kripke models,[2] however, they do not straightforwardly inherit the model-theoretical results of modal logic. Therefore, while epistemic plausibility structures are well-suited for modeling purposes, an extensive investigation of their model theory has been lacking so far.

An additional issue is that some of the main authors in this area, viz. Baltag and Smets (2008) and van Benthem (2007), define epistemic plausibility models in subtly different ways. This situation might lead to some unnecessary confusions (for example, see Footnotes 14 and 18). Furthermore, given these two definitions of epistemic plausibility models, one is left wondering whether one of them is superior over the other, and for what reasons.

The main aim of this chapter is to initiate a systematic exploration of the model theory of epistemic plausibility models. Because van Benthem's definition is the most general, I will focus on this notion (at least for now). Furthermore, given that bisimulation is *the* central notion in basic modal logic (Goranko and Otto 2006), it makes sense to start this investigation by focusing on bisimulations for (van Benthem-type) epistemic plausibility models. In the end, however, I will argue that the model-theoretical results also shed new light on the issue of the two definitions of epistemic plausibility models: they provide us with a specific argument for the superiority of Baltag and Smets's definition over that of van Benthem.

The remainder of this chapter is organized as follows. In Section 4.2, I introduce both van Benthem- and Baltag/Smets-type epistemic plausibility models, and discuss some important operators which can be interpreted on such models, and their dynamic behavior. In Section 4.3, bisimulations for van Benthem-type epistemic plausibility models are studied. I first show that the most straightforward definition of bisimulation fails, and then define various better notions of bisimulations (parametrized by a language $\mathcal{L}$). I establish a Hennessy-Milner

[2] In Kripke models, the modal operators are interpreted by universally quantifying over the accessible states; in plausibility models, however, certain operators are interpreted by considering the states that are *minimal* according to the plausibility ordering (cf. the semantic clause for conditional belief versus those for knowledge and safe belief in Definition 4.3). This notion of 'minimality' is completely absent from traditional Kripke semantics and its model theory.

type theorem, and prove several (un)definability results—thus shedding some light on the formal relationships between the various operators that can be interpreted on epistemic plausibility models. A central aspect of these bisimulations, however, turns out to be unsatisfactory for several reasons. In Section 4.4, I discuss these reasons and explore two possible solutions: adding a modality to the language, and putting extra constraints on the models. In Section 4.5, finally, I show that the second of these solutions constitutes a methodological argument to adopt Baltag and Smets's definition of epistemic plausibility model, rather than that of van Benthem.

4.2 Epistemic Plausibility Models

I will start by introducing epistemic plausibility models (EPMs), as defined in van Benthem (2007)[3] and Baltag and Smets (2008). Let I be a non-empty, finite set, whose elements will be called *agents*; furthermore, let Prop be a countably infinite set of atomic propositions. These two sets will be kept fixed throughout the chapter, so they can often be left implicit.

Definition 4.1. A *van Benthem-type epistemic plausibility model* is a structure $\mathbb{M} = \langle W, R_i, \leq_{i,w}, V\rangle_{i\in I}^{w\in W}$, where W is a non-empty set of *states*, $R_i \subseteq W \times W$ is the *epistemic accessibility relation* for agent i, $\leq_{i,w} \subseteq W \times W$ is the *plausibility order* for agent i at state w, and $V \colon \mathsf{Prop} \to \wp(W)$ is a *valuation*.

As usual, $(w, v) \in R_i$, or simply wR_iv, means that agent i cannot epistemically distinguish between states w and v. This relation is assumed to be an equivalence relation. The R_i-equivalence class of a state $w \in W$ is abbreviated as $R_i[w] := \{v \in W \mid wR_iv\}$. Furthermore, $w \leq_{i,s} v$ means that at state s, agent i considers w at least as plausible as v. This relation is assumed to be a well-founded preorder. For each $X \subseteq W$, we define the set of $\leq_{i,s}$-minimal elements as $\mathrm{Min}_{\leq_{i,s}}(X) := \{x \in X \mid \forall y \in X \colon y \leq_{i,s} x \Rightarrow x \leq_{i,s} y\}$. That $\leq_{i,s}$ is

[3] Actually, van Benthem (2007) only introduces Kripke models and plausibility models separately, and does not explicitly combine them into 'full-fledged' epistemic plausibility models (hence, in that paper the semantics of conditional belief and safe belief only involved the plausibility order, and not the epistemic accessibility relation—compare with Definition 4.3 here). However, van Benthem (p.c.) has confirmed that Definitions 4.1 and 4.3 are consistent with his intentions in that paper. Finally, it should be mentioned that more recently, van Benthem (2011, Chapter 7) initially defines EPMs in full generality, but almost immediately puts restrictions on them which are very close to the proposals made in Section 4.4 of this chapter.

a well-founded preorder means that it is reflexive and transitive, and that for each non-empty set $X \subseteq W$, $\mathrm{Min}_{\leq_{i,s}}(X)$ is non-empty as well. Note that the relation $\leq_{i,s}$ is not only dependent on agents, but also on states: it is possible for agent i to have different plausibility orderings at different states. (From Section 4.4 onwards, more constraints will be placed on the plausibility orderings.)

Definition 4.2. A *Baltag/Smets-type epistemic plausibility model* is a structure $\mathbb{M} = \langle W, R_i, \leq_i, V\rangle_{i\in I}$, where W is a non-empty set of *states*, $R_i \subseteq W \times W$ is the *epistemic accessibility relation* for agent i, $\leq_i \subseteq W \times W$ is the *plausibility order* for agent i, and $V\colon \mathsf{Prop} \to \wp(W)$ is a *valuation*.

The epistemic accessibility relation R_i is again assumed to be an equivalence relation. Note that the plausibility orders are no longer state-dependent. We assume that if $w \leq_i v$ then wR_iv, and that for every $w \in W$ the restricted ordering $\leq_i \cap\, (R_i[w] \times R_i[w])$ is a well-preorder. For each $X \subseteq R_i[w]$, we define $\mathrm{Min}_{\leq_i}(X) := \{x \in X \mid \forall y \in X\colon x \leq_i y\}$. That $\leq_i \cap\, (R_i[w] \times R_i[w])$ is a well-preorder means that it is reflexive and transitive, and that for each non-empty set $X \subseteq R_i[w]$, $\mathrm{Min}_{\leq_i}(X)$ is non-empty as well.[4]

Several intuitive epistemic and doxastic notions can be interpreted on EPMs. The three most important ones are: (i) $K_i\varphi$ (agent i *knows that* φ), (ii) $B_i^{\alpha}\varphi$ (agent i *believes that* φ, *conditional on* α), and (iii) $\Box_i\varphi$ (agent i *safely believes that* φ). ‘Normal’ (unconditional) belief can be defined in terms of conditional belief, by putting $B_i\varphi := B_i^{\top}\varphi$. ‘Safe belief’ is the name given by Baltag and Smets (2008) to a doxastic attitude ‘between’ belief and knowledge. This non-introspective attitude is sometimes called ‘defeasible knowledge’; Stalnaker (2006) takes this operator to be a more faithful representation of our ‘everyday notion’ of knowledge than the full-fledged $\mathsf{S5}$-type K_i-operator. The formal semantics for these three notions is as follows:

Definition 4.3. Consider a van Benthem-type EPM $\mathbb{M}$ and state w; then

$$\begin{array}{lll}
\mathbb{M}, w \models K_i\varphi & \text{iff} & \forall v \in R_i[w]\colon \mathbb{M}, v \models \varphi,\\
\mathbb{M}, w \models B_i^{\alpha}\varphi & \text{iff} & \forall v \in \mathrm{Min}_{\leq_{i,w}}([\![\, \alpha \,]\!]^{\mathbb{M}} \cap R_i[w])\colon \mathbb{M}, v \models \varphi,\\
\mathbb{M}, w \models \Box_i\varphi & \text{iff} & \forall v \in R_i[w]\colon v \leq_{i,w} w \Rightarrow \mathbb{M}, v \models \varphi.
\end{array}$$

[4] Baltag and Smets note that on their definition of Min, the non-emptiness of $\mathrm{Min}_{\leq_i}(X)$ (for every $X \subseteq R_i[w]$) entails that $\leq_i$ is (locally) *connected*: if wR_iv then $w \leq_i v$ or $v \leq_i w$. However, one can also impose (local) connectedness as a ‘primitive’ condition on $\leq_i$, and show that under this assumption Baltag and Smets’s definition of Min is equivalent to that of van Benthem (modulo the state-indices on the plausibility orders, of course). The role of local connectedness will be explored in depth in Subsection 4.4.2.

If $\mathbb{M}$ is a Baltag/Smets-type EPM, then we have the same semantics, with the obvious proviso that all state-indices need to be removed from the plausibility orderings.[5]

I now turn to the dynamics. I will focus on two specific dynamic phenomena: *public announcement* (hard information) and *radical upgrade* (soft information). Public announcement of a formula φ in an epistemic plausibility model $\mathbb{M}$ simply removes all $\neg\varphi$-states from the model.[6] Radical upgrade with φ, on the other hand, makes all φ-states more plausible than all $\neg\varphi$-states, and leaves everything within these two zones untouched. Formally, this looks as follows:

Definition 4.4. Let $\mathbb{M} = \langle W, R_i, \leq_{i,w}, V\rangle_{i\in I}^{w\in W}$ be an arbitrary van Benthem-type EPM, and φ an arbitrary formula. We then define the following two van Benthem-type EPMs (in the definition of the model $\mathbb{M}!\varphi$, the formula φ is additionally assumed to be true in at least one state $w \in W$):

1. $\mathbb{M}!\varphi := \langle W^{!\varphi}, R_i^{!\varphi}, \leq_{i,w}^{!\varphi}, V^{!\varphi}\rangle_{i\in I}^{w\in W^{!\varphi}}$, where
   - $W^{!\varphi} := [\![\varphi]\!]^{\mathbb{M}} = \{w \in W \mid \mathbb{M}, w \models \varphi\}$,
   - $R_i^{!\varphi} := R_i \cap ([\![\varphi]\!]^{\mathbb{M}} \times [\![\varphi]\!]^{\mathbb{M}})$ for all $i \in I$,
   - $\leq_{i,w}^{!\varphi} := \leq_{i,w} \cap ([\![\varphi]\!]^{\mathbb{M}} \times [\![\varphi]\!]^{\mathbb{M}})$ for all $i \in I$ and $w \in W^{!\varphi}$,
   - $V^{!\varphi}(p) := V(p) \cap [\![\varphi]\!]^{\mathbb{M}}$ for all $p \in \mathsf{Prop}$;
2. $\mathbb{M} \Uparrow \varphi := \langle W^{\Uparrow\varphi}, R_i^{\Uparrow\varphi}, \leq_{i,w}^{\Uparrow\varphi}, V^{\Uparrow\varphi}\rangle_{i\in I}^{w\in W^{\Uparrow\varphi}}$, where
   - $W^{\Uparrow\varphi} := W$,
   - $R_i^{\Uparrow\varphi} := R_i$ for all $i \in I$,
   - $\leq_{i,w}^{\Uparrow\varphi} := \big(\leq_{i,w} \cap ([\![\varphi]\!]^{\mathbb{M}} \times [\![\varphi]\!]^{\mathbb{M}})\big) \cup \big(\leq_{i,w} \cap ([\![\neg\varphi]\!]^{\mathbb{M}} \times [\![\neg\varphi]\!]^{\mathbb{M}})\big)$ $\cup \big([\![\varphi]\!]^{\mathbb{M}} \times [\![\neg\varphi]\!]^{\mathbb{M}}\big)$ for all $i \in I$ and $w \in W^{\Uparrow\varphi}$,
   - $V^{\Uparrow\varphi}(p) := V(p)$ for all $p \in \mathsf{Prop}$.

[5] Furthermore, for Baltag/Smets-type EPMs the semantic clauses mentioned in the definition contain some redundancies. For example, in the clause for safe belief, it is not necessary to require that $v \in R_i[w]$, because this follows already from the condition that $v \leq_i w$.

[6] The definition of a public announcement in an EPM closely resembles that of a public announcement in a well-behaved probabilistic Kripke model (recall Definition 3.6 on p. 86). In this chapter, I will write $\mathbb{M}!\varphi$ instead of $\mathbb{M}|\varphi$, to emphasize the contrast with $\mathbb{M} \Uparrow \varphi$.

If $\mathbb{M}$ is a Baltag/Smets-type EPM, then $\mathbb{M}!\varphi$ and $\mathbb{M} \Uparrow \varphi$ are defined in exactly the same way, with the provisos (i) that all state-indices need to be removed from the plausibility orderings, and (ii) that in the definition of $\leq_i^{\Uparrow\varphi}$ the third part of the union needs to be added 'locally', i.e. $\big(\llbracket\varphi\rrbracket^{\mathbb{M}} \times \llbracket\neg\varphi\rrbracket^{\mathbb{M}}\big)$ needs to be replaced with

$$\bigcup_{v \in W^{\Uparrow\varphi}} \big((\llbracket\varphi\rrbracket^{\mathbb{M}} \cap R_i[v]) \times (\llbracket\neg\varphi\rrbracket^{\mathbb{M}} \cap R_i[v])\big).$$

It is easy to check that if $\mathbb{M}$ is a (van Benthem- or Baltag/Smets-type) EPM, then $\mathbb{M}!\varphi$ and $\mathbb{M} \Uparrow \varphi$ are (van Benthem- or Baltag/Smets-type) EPMs as well. In order to be able to talk about these new models in the object language, we add operators $[!\varphi]$ and $[\Uparrow \varphi]$. Hence, the full language $\mathcal{L}(K, B^c, \Box, !, \Uparrow)$ has the following BNF:[7]

$$\begin{array}{lll} \varphi & ::= & p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_i\varphi \mid B_i^{\varphi}\varphi \mid \Box_i\varphi \mid [A]\varphi \\ A & ::= & !\varphi \mid \Uparrow \varphi \end{array}$$

The semantics for the two dynamic operators looks as follows:[8]

$$\begin{array}{lll} \mathbb{M}, w \models [!\varphi]\psi & \text{iff} & \text{if } \mathbb{M}, w \models \varphi \text{ then } \mathbb{M}!\varphi, w \models \psi, \\ \mathbb{M}, w \models [\Uparrow \varphi]\psi & \text{iff} & \mathbb{M} \Uparrow \varphi, w \models \psi. \end{array}$$

Finally, dynamic epistemic/doxastic logics are constructed using the well-known *modular* approach: (i) one starts by taking (an axiomatization of) some static base logic (in a sufficiently rich language, so that step (iii) can be done successfully),[9] (ii) then one adds dynamic operators to this logic and (iii) finally, one provides a sound set of reduction axioms, which allow each formula in the dynamic language to be rewritten as an equivalent formula in the static language. Because of this final step, completeness of the dynamified logic is reduced to completeness of the static base logic. It also shows that the dynamic language $\mathcal{L}(K, B^c, \Box, !, \Uparrow)$ is equally expressive as the static language $\mathcal{L}(K, B^c, \Box)$.[10]

[7] Of course, one can also study more restricted languages. BNFs for such restricted languages can easily be obtained from the BNF for the full language.

[8] Note that since public announcement is assumed to be *truthful*, it works with a precondition; this is not the case for radical upgrade.

[9] For example, if the language contains radical upgrade and safe belief operators, then it should also contain the knowledge operator, since the right-hand-side of the reduction axiom for $[\Uparrow \varphi]\Box\psi$ involves the K-operator (cf. infra).

[10] In the remainder of the chapter, we can thus safely focus on (sublanguages of) $\mathcal{L}(K, B^c, \Box)$.

To illustrate this methodology, I will state the most important reduction axioms for public announcement and radical upgrade, viz. those in which the epistemic/doxastic operators are being rewritten:

$$
\begin{array}{lcl}
[!\varphi]K_i\psi & \leftrightarrow & (\varphi \to K_i[!\varphi]\psi), \\
[!\varphi]B_i^{\alpha}\psi & \leftrightarrow & (\varphi \to B_i^{\varphi\wedge[!\varphi]\alpha}[!\varphi]\psi), \\
[!\varphi]\Box_i\psi & \leftrightarrow & (\varphi \to \Box_i[!\varphi]\psi), \\
 & & \\
[\Uparrow \varphi]K_i\psi & \leftrightarrow & K_i[\Uparrow \varphi]\psi, \\
[\Uparrow \varphi]B_i^{\alpha}\psi & \leftrightarrow & \big(\neg K_i\neg(\varphi\wedge[\Uparrow \varphi]\alpha) \wedge B_i^{\varphi\wedge[\Uparrow\varphi]\alpha}[\Uparrow \varphi]\psi\big)\vee \\
 & & \big(K_i\neg(\varphi\wedge[\Uparrow \varphi]\alpha) \wedge B_i^{[\Uparrow\varphi]\alpha}[\Uparrow \varphi]\psi\big), \\
[\Uparrow \varphi]\Box_i\psi & \leftrightarrow & \big(\varphi \to \Box_i(\varphi \to [\Uparrow \varphi]\psi)\big)\wedge \\
 & & \big(\neg\varphi \to (\Box_i(\neg\varphi \to [\Uparrow \varphi]\psi) \wedge K_i(\varphi \to [\Uparrow \varphi]\psi))\big).
\end{array}
$$

4.3 Bisimulations for Epistemic Plausibility Models

We now begin investigating the model theory of epistemic plausibility models. Because van Benthem's notion is the most general, we will start by studying this notion, rather than that of Baltag and Smets. The focus will be on *bisimulation*, which is also *the* central concept in the model theory of Kripke models (Goranko and Otto 2006). In Subsection 4.3.1, I show that the most straightforward generalization of bisimulation fails. In Subsection 4.3.2 I then define various better notions of bisimulations, parametrized by a language $\mathcal{L}$. Furthermore, a Hennessy-Milner type theorem and several (un)definability results are established.

4.3.1 A Straightforward Generalization

Since (van Benthem-type) epistemic plausibility models[11] are obtained from Kripke models by adding the plausibility orders $\leq_{i,w}$, a natural generalization

[11] Since all epistemic plausibility models in Sections 4.3 and 4.4 are assumed to be *van Benthem-type* models, this qualification will often be left implicit in these sections. In Section 4.5, however, van Benthem- and Baltag/Smets-type models will be compared to each other, and then I will always mention explicitly which of the two notions is meant.

of bisimulation is to simply include back- and forth-clauses for these additional relations.[12]

Definition 4.5. Consider two EPMs $\mathbb{M} = \langle W, R_i, \leq_{i,w}, V\rangle_{i\in I}^{w\in W}$ and $\mathbb{M}' = \langle W', R_i', \leq_{i,w'}', V'\rangle_{i\in I}^{w'\in W'}$. Then a relation $Z \subseteq W \times W'$ is said to be a *pseudo-bisimulation* iff

- if $(w, w') \in Z$, then for all atoms p: $w \in V(p) \Leftrightarrow w' \in V'(p)$,
- if $(w, w') \in Z$ and wR_iv, then $\exists v' \in W'\colon (v, v') \in Z$ and $w'R_i'v'$,
- if $(w, w') \in Z$ and $w'R_i'v'$, then $\exists v \in W\colon (v, v') \in Z$ and wR_iv,
- if $(w, w') \in Z$ and wR_iv and $v \leq_{i,w} w$,
  then $\exists v' \in W'\colon (v, v') \in Z$ and $w'R_i'v'$ and $v' \leq_{i,w'}' w'$,
- if $(w, w') \in Z$ and $w'R_i'v'$ and $v' \leq_{i,w'}' w'$,
  then $\exists v \in W\colon (v, v') \in Z$ and wR_iv and $v \leq_{i,w} w$.

This definition inherits the back- and forth-clauses for R_i from the traditional definition for Kripke models, and adds to them the back- and forth-clauses for the plausibility orderings $\leq_{i,w}$ and $\leq_{i,w'}'$. Note that for states w and w' to be pseudo-bisimilar, it suffices that the plausibility orderings *indexed by* w *and* w' satisfy the back- and forth-clauses (rather than *all* plausibility orderings $\leq_{i,t}$ and $\leq_{i,t'}'$, with arbitrary states t and t'). This fits well with the semantics of conditional belief (Definition 4.3): to determine the truth value of a conditional belief formula at a state w, one only needs to take into account (the minimal states according to) the plausibility ordering $\leq_{i,w}$, and not (those according to) other orderings $\leq_{i,t}$.

Despite the simplicity of this definition, it does not capture the right notion of bisimulation between EPMs. Typically, bisimilar model-state pairs are modally equivalent (i.e. satisfy exactly the same formulas). Pseudo-bisimulations, however, are not strong enough to ensure this property. The following example exhibits two model-state pairs which are pseudo-bisimilar, while still differing on some formulas.

[12] The defined notion is called '*pseudo*-bisimulation' because, as will be shown below, it lacks one of the typical properties of bisimulations.

Example 4.1. This example only involves one agent; we can therefore drop agent indices. Define EPMs $\mathbb{M} = \langle \{w, v\}, R, \{\leq_w, \leq_v\}, V\rangle$ and $\mathbb{M}' = \langle \{w', v'\}, R', \{\leq'_{w'}, \leq'_{v'}\}, V'\rangle$, where $R = W \times W$, $R' = W' \times W'$, $V(p) = W$, $V(q) = \{w\}$, $V'(p) = W'$, $V'(q) = \{w'\}$, and

- $\leq_w = \{(w, w), (v, v), (w, v)\} = \leq_v$
- $\leq'_{w'} = \{(w', w'), (v', v')\}$ and $\leq'_{v'} = W' \times W'$

Then one can show that $Z := \{(w, w'), (v, v')\}$ is a pseudo-bisimulation. Still, it holds that $\mathbb{M}, w \models B^p q$, while $\mathbb{M}', w' \not\models B^p q$.

4.3.2 Parametrized Bisimulations

We have just seen that the most straightforward definition of bisimulation for EPMs is unsuccessful, because it fails to guarantee modal equivalence. I therefore propose to consider parametrized bisimulations: for each of the three main operators introduced in Section 4.2, a corresponding notion of bisimulation is introduced. The notions of K- and $\Box$-bisimulation are as expected. The notion of B^c-bisimulation, however, is much more intricate, since it involves universally quantifying over all formulas of the language $\mathcal{L}(B^c)$. The semantic notion of bisimulation thus becomes language- (i.e. syntax-)dependent. I will return to this issue in Section 4.4.

Definition 4.6. Consider two EPMs $\mathbb{M} = \langle W, R_i, \leq_{i,w}, V\rangle_{i\in I}^{w\in W}$ and $\mathbb{M}' = \langle W', R'_i, \leq'_{i,w'}, V'\rangle_{i\in I}^{w'\in W'}$, and a relation $Z \subseteq W \times W'$; then:

1. Z is a *K-bisimulation* iff
   - if $(w, w') \in Z$, then for all atoms p: $w \in V(p) \Leftrightarrow w' \in V'(p)$,
   - if $(w, w') \in Z$ and wR_iv, then $\exists v' \in W'\colon (v, v') \in Z$ and $w'R'_iv'$,
   - if $(w, w') \in Z$ and $w'R'_iv'$, then $\exists v \in W\colon (v, v') \in Z$ and wR_iv;
2. Z is a *$\Box$-bisimulation* iff
   - if $(w, w') \in Z$, then for all atoms p: $w \in V(p) \Leftrightarrow w' \in V'(p)$,
   - if $(w, w') \in Z$ and wR_iv and $v \leq_{i,w} w$,
     then $\exists v' \in W'\colon (v, v') \in Z$ and $w'R'_iv'$ and $v' \leq'_{i,w'} w'$,

- if $(w, w') \in Z$ and $w' R'_i v'$ and $v' \leq'_{i,w'} w'$,
  then $\exists v \in W\colon (v, v') \in Z$ and $w R_i v$ and $v \leq_{i,w} w$;

3. Z is a B^c-*bisimulation* iff

   - if $(w, w') \in Z$, then for all atoms p: $w \in V(p) \Leftrightarrow w' \in V'(p)$,
   - $\forall \alpha \in \mathcal{L}(B^c)$: if $(w, w') \in Z$ and $v \in \mathrm{Min}_{\leq_{i,w}}([\![\, \alpha \,]\!]^{\mathbb{M}} \cap R_i[w])$,
     then $\exists v' \in W'\colon (v, v') \in Z$ and $v' \in \mathrm{Min}_{\leq'_{i,w'}}([\![\, \alpha \,]\!]^{\mathbb{M}'} \cap R'_i[w'])$,
   - $\forall \alpha \in \mathcal{L}(B^c)$: if $(w, w') \in Z$ and $v' \in \mathrm{Min}_{\leq'_{i,w'}}([\![\, \alpha \,]\!]^{\mathbb{M}'} \cap R'_i[w'])$,
     then $\exists v \in W\colon (v, v') \in Z$ and $v \in \mathrm{Min}_{\leq_{i,w}}([\![\, \alpha \,]\!]^{\mathbb{M}} \cap R_i[w])$.

Definition 4.7 formally introduces the notion of $\mathcal{L}$-equivalence (for any language $\mathcal{L}$). Theorem 4.1 shows that the notions introduced in Definition 4.6 are the proper ones: each of them gives rise to the desired 'bisimilarity implies equivalence'-result.

Definition 4.7. Consider two EPMs $\mathbb{M} = \langle W, R_i, \leq_{i,w}, V\rangle_{i\in I}^{w\in W}$ and $\mathbb{M}' = \langle W', R'_i\}, \leq'_{i,w'}, V'\rangle_{i\in I}^{w'\in W'}$, and states $w \in W, w' \in W'$. Fix a language $\mathcal{L}$. We say that $\mathbb{M}, w$ and $\mathbb{M}', w'$ are $\mathcal{L}$-*equivalent* (notation: $\mathbb{M}, w \equiv_{\mathcal{L}} \mathbb{M}', w'$) iff

$$\forall \varphi \in \mathcal{L}\colon \mathbb{M}, w \models \varphi \text{ iff } \mathbb{M}', w' \models \varphi.$$

Theorem 4.1. *Consider two EPMs* $\mathbb{M} = \langle W, R_i, \leq_{i,w}, V\rangle_{i\in I}^{w\in W}$ *and* $\mathbb{M}' = \langle W', R'_i, \leq'_{i,w'}, V'\rangle_{i\in I}^{w'\in W'}$, *states* $w \in W, w' \in W'$, *and a relation* $Z \subseteq W \times W'$. *Suppose that* $(w, w') \in Z$.

1. *If Z is a K-bisimulation, then* $\mathbb{M}, w \equiv_{\mathcal{L}(K)} \mathbb{M}', w'$.
2. *If Z is a $\Box$-bisimulation, then* $\mathbb{M}, w \equiv_{\mathcal{L}(\Box)} \mathbb{M}', w'$.
3. *If Z is a B^c-bisimulation, then* $\mathbb{M}, w \equiv_{\mathcal{L}(B^c)} \mathbb{M}', w'$.

Using these 'separate' notions of bisimulation, we can introduce bisimulations for languages which have more than just one of the operators $K/\Box/B^c$ in a modular way (although conditional belief complicates matters a little bit). Obviously, these combined notions lead to results analogous to Theorem 4.1; three of them are stated as Theorem 4.2, for future reference.

Definition 4.8. Consider two EPMs $\mathbb{M} = \langle W, R_i, \leq_{i,w}, V\rangle_{i\in I}^{w\in W}$ and $\mathbb{M}' = \langle W', R_i', \leq_{i,w'}', V'\rangle_{i\in I}^{w'\in W'}$ and a relation $Z \subseteq W \times W'$.

1. Z is a $\{K, \Box\}$-bisimulation iff Z is a K- and a $\Box$-bisimulation.
2. Z is a $\{K, B^c\}$-bisimulation iff Z is a K- and a B^c-bisimulation, with the universal quantifiers ranging over $\mathcal{L}(K, B^c)$ instead of $\mathcal{L}(B^c)$.
3. Z is a $\{K, \Box, B^c\}$-bisimulation iff Z is a K-, a $\Box$- and a B^c-bisimulation, with the universal quantifiers ranging over $\mathcal{L}(K, \Box, B^c)$ instead of $\mathcal{L}(B^c)$.

Theorem 4.2. *Consider two EPMs* $\mathbb{M} = \langle W, R_i, \leq_{i,w}, V\rangle_{i\in I}^{w\in W}$ *and* $\mathbb{M}' = \langle W', R_i', \leq_{i,w'}', V'\rangle_{i\in I}^{w'\in W'}$*, states* $w \in W, w' \in W'$*, and a relation* $Z \subseteq W \times W'$*. Suppose that* $(w, w') \in Z$*.*

1. *If Z is a $\{K, \Box\}$-bisimulation, then* $\mathbb{M}, w \equiv_{\mathcal{L}(K,\Box)} \mathbb{M}', w'$*.*
2. *If Z is a $\{K, B^c\}$-bisimulation, then* $\mathbb{M}, w \equiv_{\mathcal{L}(K,B^c)} \mathbb{M}', w'$*.*
3. *If Z is a $\{K, B^c, \Box\}$-bisimulation, then* $\mathbb{M}, w \equiv_{\mathcal{L}(K,B^c,\Box)} \mathbb{M}', w'$*.*

Remark 4.1. Unraveling the definitions, it is clear that $\{K, \Box\}$-bisimulations simply are pseudo-bisimulations (as defined in the previous subsection). Item 2 of Theorem 4.2 says that such bisimulations imply $\mathcal{L}(K, \Box)$-equivalence. This is consistent with Example 4.1, since the differentiating formula there did not belong to $\mathcal{L}(K, \Box)$: it contained the conditional belief operator.

One can wonder about the converse direction of theorems such as Theorem 4.2: e.g. if $\mathbb{M}, w \equiv_{\mathcal{L}(K,B^c)} \mathbb{M}', w'$, is there then always a $\{K, B^c\}$-bisimulation $Z \subseteq W \times W'$ such that $(w, w') \in Z$? One of the main results from the model theory of basic modal logic, viz. the Hennessy-Milner theorem (Blackburn et al. 2001, Theorem 2.24) says that this question can be answered positively, at least when the models are assumed to be image-finite. This theorem can easily be generalized to epistemic plausibility models:[13]

Definition 4.9. Consider an EPM $\mathbb{M} = \langle W, R_i, \leq_{i,w}, V\rangle_{i\in I}^{w\in W}$. We say that $\mathbb{M}$ is *image-finite* if for all $i \in I$ and all $w \in W$, the set $R_i[w]$ is finite.

[13] Of course, there are analogues of Theorem 4.3 for all of the languages defined in this chapter.

Theorem 4.3. *Consider two EPMs* $\mathbb{M} = \langle W, R_i, \leq_{i,w}, V\rangle_{i\in I}^{w\in W}$ *and* $\mathbb{M}' = \langle W', R'_i, \leq'_{i,w'}, V'\rangle_{i\in I}^{w'\in W'}$*, and assume that they are image-finite. Then for all states* $w \in W$ *and* $w' \in W'$*, if* $\mathbb{M}, w \equiv_{\mathcal{L}(K,B^c)} \mathbb{M}', w'$*, then* w *and* w' *are* $\{K, B^c\}$*-bisimilar.*

Proof. One can use the proof technique of the Hennessy-Milner theorem for basic modal logic, viz. show that $\equiv_{\{K,B^c\}}$ is itself a $\{K, B^c\}$-bisimulation. □

Now that we have bisimulations which are strong enough to guarantee equivalence, the interdefinability of the three epistemic/doxastic operators can be investigated. Fact 4.1 states that knowledge is already definable in terms of conditional belief (one does not even need safe belief).[14] Our reason for still including knowledge as a separate modality is thus purely practical: EPM semantics can be seen as an extension of Kripke semantics (recall Footnote 2), and therefore it is natural to think of the languages interpretable on EPMs as being extensions of the language $\mathcal{L}(K)$. Propositions 4.1 and 4.2, however, show that the conditional belief and safe belief operators cannot be defined in terms of each other, even in the presence of the knowledge operator.

Fact 4.1. For all EPMs $\mathbb{M}$ it holds that $\mathbb{M} \models K_i\varphi \leftrightarrow B_i^{\neg\varphi}\bot$.

Proposition 4.1. *Conditional belief cannot be defined in terms of knowledge and safe belief.*

Proof. Consider the models defined in Example 4.1. Recall that $\mathbb{M}, w$ and $\mathbb{M}', w'$ are pseudo-bisimilar, and thus $\{K, \Box\}$-bisimilar (recall Remark 4.1). By item 1 of Theorem 4.2 it follows that $\mathbb{M}, w \equiv_{\mathcal{L}(K,\Box)} \mathbb{M}', w'$. Still, Example 4.1 also specified that $\mathbb{M}, w \models B^p q$, while $\mathbb{M}', w' \not\models B^p q$. It follows that the formula $B^p q$ (and thus the conditional belief operator in general) cannot be defined in $\mathcal{L}(K, \Box)$. □

Proposition 4.2. *Safe belief cannot be defined in terms of knowledge and conditional belief.*

Proof. We will work with only one agent, and thus drop agent indices. Define EPMs $\mathbb{M} = \langle\{w, v\}, R, \{\leq_w, \leq_v\}, V\rangle$ and $\mathbb{M}' = \langle\{w', v'\}, R', \{\leq'_{w'}, \leq'_{v'}\}, V'\rangle$, where $R = W \times W$, $R' = W' \times W'$, $V(p) = \{w\}$, $V'(p) = \{w'\}$, and

[14] This definability result was already noted by Baltag and Smets (2008). Fact 4.1 says, however, that this definability result holds not only in Baltag/Smets-type EPMs, but also in van Benthem-type EPMs.

- $\leq_w = \{(w,w),(v,v)\} = \leq_v$,
- $\leq'_{w'} = W' \times W' = \leq'_{v'}$.

One can easily check that $\mathbb{M}, w \models \Box p$, while $\mathbb{M}', w' \not\models \Box p$. However, $Z := \{(w,w'),(v,v')\}$ is a $\{K, B^c\}$-bisimulation (cf. infra), and hence $\mathbb{M}, w \equiv_{\mathcal{L}(K,B^c)} \mathbb{M}', w'$, by item 2 of Theorem 4.2. It follows that the formula $\Box p$ (and thus the safe belief operator in general) cannot be defined in $\mathcal{L}(K, B^c)$.

We now prove the claim that Z is a $\{K, B^c\}$-bisimulation. It is easy to check that Z is a K-bisimulation; we thus focus on the back- and forth-clauses of B^c-bisimulations. Note that for any $x \in \{w, v\}$ and $X \subseteq \{w, v\}$, it holds that $\text{Min}_{\leq_x}(X) = X$ and $R[x] = \{w, v\}$. Hence we get that $y \in \text{Min}_{\leq_x}([\![\alpha]\!]^{\mathbb{M}} \cap R[x])$ iff $y \in [\![\alpha]\!]^{\mathbb{M}}$ iff $\mathbb{M}, y \models \alpha$. Analogously, we show that $y' \in \text{Min}_{\leq'_{x'}}([\![\alpha]\!]^{\mathbb{M}'} \cap R'[x'])$ iff $\mathbb{M}', y' \models \alpha$. The back- and forth-clauses for B^c-bisimulations can thus be written as follows:

1. $\forall \alpha \in \mathcal{L}(K, B^c)$: if $(x, x') \in Z$ and $\mathbb{M}, y \models \alpha$,
   then $\exists y' \in W' : (y, y') \in Z$ and $\mathbb{M}', y' \models \alpha$;
2. $\forall \alpha \in \mathcal{L}(K, B^c)$: if $(x, x') \in Z$ and $\mathbb{M}', y' \models \alpha$,
   then $\exists y \in W : (y, y') \in Z$ and $\mathbb{M}, y \models \alpha$.

Since $Z = \{(w, w'), (v, v')\}$, the conjunction of these two claims can be rewritten as follows:

$$\forall \alpha \in \mathcal{L}(K, B^c) : \big(\mathbb{M}, w \models \alpha \Leftrightarrow \mathbb{M}', w' \models \alpha \text{ and } \mathbb{M}, v \models \alpha \Leftrightarrow \mathbb{M}', v' \models \alpha\big).$$

This is easily proved by induction on the complexity of α. □

4.4 Structural Bisimulations

It was already noted in the previous section that the notion of B^c-bisimulation introduced in Definition 4.6 is much more intricate than the other notions. Therefore this definition is unsatisfactory for both theoretical and practical reasons.

On the *theoretical* level, since the definition of B^c-bisimulation involves universal quantification over $\mathcal{L}(B^c)$, it is not strictly structural. Rather than stating conditions on R_i and $\leq_{i,w}$, it essentially involves truth sets of (arbitrary) formulas. A related issue is that—unlike most definitions of bisimulations—the definition of a B^c-bisimulation for *models* cannot be turned into a definition of a bisimulation for *frames* by simply dropping the atoms clause. After all, this definition depends on truth sets of formulas (viz. $[\![\,\alpha\,]\!]^{\mathbb{M}}$ and $[\![\,\alpha\,]\!]^{\mathbb{M}'}$ for all $\alpha \in \mathcal{L}(B^c)$), and thus also on the concrete valuations of $\mathbb{M}$ and $\mathbb{M}'$.[15]

Practically speaking, the notion of B^c-bisimulation introduced in Definition 4.6 makes it often very difficult to prove that two given EPMs are actually B^c-bisimilar, since it involves a back- and a forth-clause for every single formula $\alpha \in \mathcal{L}(B^c)$. In the appendix of Dégremont and Roy (2010), the authors establish that a relation between two given models is a B^c-bisimulation, and go through the infinitely many back- and forth-clauses by means of induction on the complexity of α (and a cleverly strengthened induction hypothesis). However, this approach is geared towards proving one particular B^c-bisimilarity result about artificially crafted models, and cannot be generalized to the general case. Similar remarks apply to the proof of Proposition 4.2 above. This situation often gives rise to a 'practical circularity', which renders the current notion of B^c-bisimulation practically useless. For example, in Proposition 4.2, we want to show that the states w and w' are $\{K, B^c\}$-bisimilar, and then conclude from this that they are $\mathcal{L}(K, B^c)$-equivalent. However, note that while establishing B^c-bisimilarity, we ended up proving (a strengthened version of) the desired $\mathcal{L}(K, B^c)$-equivalence directly.

I will now propose two different solutions to this problem, and explore and compare their advantages and disadvantages. Both solutions involve reducing conditional belief to other modalities which have more standard (structural) notions of bisimulation. The first approach is presented in Subsection 4.4.1, and involves both extending the language and putting some mild constraints on the epistemic plausibility models. The second approach is discussed in Subsection 4.4.2, and puts more heavy constraints on the models, but does not need to extend the language.

[15] When one is quantifying over all *definable* subsets in a model, and one wants to eliminate reference to the language, this can sometimes be achieved by quantifying over *all* subsets in the model. However, this solution is not available here, since we are 'comparing' sets *across* different models. For a given subset X in $\mathbb{M}$ (where X is meant to generalize a definable subset $[\![\,\alpha\,]\!]^{\mathbb{M}}$), it is not clear what the subset X' in $\mathbb{M}'$ should be (where X' plays the role of $[\![\,\alpha\,]\!]^{\mathbb{M}'}$).

4.4.1 Adding a New Modality

The first approach combines language engineering and putting some mild constraints on the models.[16] These constraints are captured by the following definition:

Definition 4.10. An EPM $\mathbb{M} = \langle W, R_i, \leq_{i,w}, V\rangle_{i\in I}^{w\in W}$ is called *uniform* iff the plausibility relations are uniform within epistemic equivalence classes, i.e. iff for any $i \in I$ and $w, v \in W$: if wR_iv then $\leq_{i,w} = \leq_{i,v}$.

Uniformity has an independent intuitive justification (Aucher 2003, p. 22). Suppose that wR_iv, i.e. agent i cannot epistemically distinguish between states w and v. This means that the same information is available to her in both states. Since agent i is rational, her plausibility ordering (at any given state) is fully determined by the information available to her (at that state). Since the same information is available to agent i in states w and v, her plausibility orderings should be the same in both states, i.e. $\leq_{i,w} = \leq_{i,v}$.[17] Besides being intuitively motivated, uniformity has some nice technical features as well: it leads to a full introspection principle (Fact 4.2), and it is a *dynamically robust* notion, in the sense that uniformity is preserved by the two model update operations studied in this chapter (Fact 4.3).

Fact 4.2. For uniform EPMs $\mathbb{M}$, it is the case that

$$\mathbb{M} \models B_i^\alpha \varphi \to K_i B_i^\alpha \varphi,$$
$$\mathbb{M} \models \neg B_i^\alpha \varphi \to K_i \neg B_i^\alpha \varphi.$$

Fact 4.3. If an EPM $\mathbb{M}$ is uniform, then so are $\mathbb{M}!\varphi$ and $\mathbb{M} \Uparrow \varphi$.

I will now introduce the language extension that is needed for the first approach. For any agent $i \in I$ and state w in a plausibility model, let us abbreviate

$$<_{i,w} \;:=\; \leq_{i,w} \;-\; \geq_{i,w} \,.$$

The language is extended with a modality $[>_i]$ to talk about this strict version of the plausibility order. As in Definition 4.3, the semantics for this modality is relativized to the epistemic equivalence classes:

[16] This approach is based on a suggestion by Johan van Benthem and Davide Grossi. It has also been studied by Girard (2008), whose Fact 3.1.4 is similar to Fact 4.4 established here.

[17] This is a close analogue of the notion of uniformity in probabilistic epistemic models; recall Definition 3.3 on p. 79. Furthermore, note the similarity between Fact 4.2 about introspection for plausibilities (conditional beliefs) and Lemma 3.1 (item 1) about introspection for probabilities.

Definition 4.11. Consider an epistemic plausibility model $\mathbb{M}$ and state w; then

$$\mathbb{M}, w \models [>_i]\varphi \quad \text{iff} \quad \forall v \in R_i[w]\colon v <_{i,w} w \Rightarrow \mathbb{M}, v \models \varphi.$$

Adding this new modality $[>_i]$ as a primitive operator is justified, because it cannot be defined in even the richest language of the previous section:

Proposition 4.3. *The modality* $[>_i]$ *cannot be defined in* $\mathcal{L}(K, B^c, \Box)$.

Proof. We will work with only one agent, and thus drop agent indices. Define EPMs $\mathbb{M} = \langle \{w, v\}, R, \{\leq_w, \leq_v\}, V\rangle$ and $\mathbb{M}' = \langle \{w', v'\}, R', \{\leq'_{w'}, \leq'_{v'}\}, V'\rangle$, where $R = W \times W$, $R' = W' \times W'$, V and V' are irrelevant, and

- $\leq_w = \{(w, w), (v, v), (v, w)\} = \leq_v$
- $\leq'_{w'} = W' \times W' = \leq'_{v'}$

One can easily check that $\mathbb{M}, w \not\models [>]\bot$, while $\mathbb{M}', w' \models [>]\bot$. However, we claim that $\mathbb{M}, w \equiv_{\mathcal{L}(K,\Box,B^c)} \mathbb{M}', w'$ (cf. infra). It follows that the formula $[>]\bot$ (and thus the $[>]$-operator in general) cannot be defined in $\mathcal{L}(K, \Box, B^c)$.

We now show that $\mathbb{M}, w \equiv_{\mathcal{L}(K,\Box,B^c)} \mathbb{M}', w'$. First, we prove an auxiliary claim about the model $\mathbb{M}'$:

$$\forall \varphi \in \mathcal{L}(K, \Box, B^c)\colon \mathbb{M}', w' \models \varphi \Leftrightarrow \mathbb{M}', v' \models \varphi. \tag{4.1}$$

This auxiliary claim is proved by induction on the complexity of φ. Then we go on to prove that

$$\forall \varphi \in \mathcal{L}(K, \Box, B^c)\colon \big(\mathbb{M}, w \models \varphi \Leftrightarrow \mathbb{M}', w' \models \varphi \text{ and } \mathbb{M}, v \models \varphi \Leftrightarrow \mathbb{M}', v' \models \varphi\big). \tag{4.2}$$

This is also proved by induction on the complexity of φ; the auxiliary claim (4.1) is used in the induction cases for safe belief and conditional belief; furthermore, the induction case for safe belief (i.e. $\varphi = B^\alpha\psi$) is proved using a case distinction about whether or not $\mathbb{M}, v \models \alpha$. From (4.2) it follows that $\mathbb{M}, w \equiv_{\mathcal{L}(K,\Box,B^c)} \mathbb{M}', w'$, as required. □

The $[>]$-modality is actually so expressive that, together with the knowledge operator, it is able to define the notion of conditional belief—at least, when we restrict ourselves to *uniform* epistemic plausibility models.

Fact 4.4. For every uniform epistemic plausibility model $\mathbb{M}$, it is the case that

$$\mathbb{M} \models B_i^{\alpha}\varphi \leftrightarrow K_i\big((\alpha \wedge \neg\langle >_i\rangle\alpha) \to \varphi\big).$$

We are now ready to introduce the notion of $[>]$-bisimilarity, which—as desired—is fully structural:

Definition 4.12. Consider two EPMs $\mathbb{M} = \langle W, R_i, \leq_{i,w}, V\rangle_{i\in I}^{w\in W}$ and $\mathbb{M}' = \langle W', R_i', \leq_{i,w'}', V'\rangle_{i\in I}^{w'\in W'}$, and a relation $Z \subseteq W \times W'$. Then Z is a $[>]$-*bisimulation* iff

- if $(w, w') \in Z$, then for all atoms p: $w \in V(p) \Leftrightarrow w' \in V'(p)$,
- if $(w, w') \in Z$ and wR_iv and $v <_{i,w} w$,
  then $\exists v' \in W'$: $(v, v') \in Z$ and $w'R_i'v'$ and $v' <_{i,w'}' w'$,
- if $(w, w') \in Z$ and $w'R_i'v'$ and $v' <_{i,w'}' w'$,
  then $\exists v \in W$: $(v, v') \in Z$ and wR_iv and $v <_{i,w} w$.

As expected, a $\{K, [>]\}$-bisimulation is defined to be a relation that is simultaneously a K- and a $[>]$-bisimulation. Since K- and $[>]$-bisimulations are both structural, $\{K, [>]\}$-bisimulations are structural as well. Item 3 of Theorem 4.4 says that for uniform EPMs, the structural notion of $\{K, [>]\}$-bisimulation suffices to obtain equivalence for the entire language $\mathcal{L}(K, [>], B^c)$, *including* conditional belief.

Theorem 4.4. *Consider two uniform EPMs* $\mathbb{M} = \langle W, R_i, \leq_{i,w}, V\rangle_{i\in I}^{w\in W}$ *and* $\mathbb{M}' = \langle W', R_i', \leq_{i,w'}', V'\rangle_{i\in I}^{w'\in W'}$, *states* $w \in W, w' \in W'$, *and a relation* $Z \subseteq W \times W'$. *Suppose that* $(w, w') \in Z$.

1. *If Z is a* $[>]$*-bisimulation, then* $\mathbb{M}, w \equiv_{\mathcal{L}([>])} \mathbb{M}', w'$.
2. *If Z is a* $\{K, [>]\}$*-bisimulation, then* $\mathbb{M}, w \equiv_{\mathcal{L}(K,[>])} \mathbb{M}', w'$.
3. *If Z is a* $\{K, [>]\}$*-bisimulation, then* $\mathbb{M}, w \equiv_{\mathcal{L}(K,[>],B^c)} \mathbb{M}', w'$.

Proof. Items 1 and 2 are easily proved by induction on the complexity of φ. Item 3 follows immediately from item 2, because Fact 4.4 states that B^c is definable in $\mathcal{L}(K, [>])$. □

Finally, it should be noted that the two model update operations studied in this paper are both *safe* for $\{K, [>]\}$-bisimulation:

Proposition 4.4. *Consider two uniform EPMs* $\mathbb{M} = \langle W, R_i, \leq_{i,w}, V\rangle_{i\in I}^{w\in W}$ *and* $\mathbb{M}' = \langle W', R_i', \leq_{i,w'}', V'\rangle_{i\in I}^{w'\in W'}$*, and states* $w \in W, w' \in W'$*. Suppose that* $\mathbb{M}, w$ *and* $\mathbb{M}', w'$ *are* $\{K, [>]\}$*-bisimilar, and* $\mathbb{M}, w \models \varphi$ *and* $\mathbb{M}', w' \models \varphi$*. Then:*

1. $\mathbb{M}!\varphi, w$ *and* $\mathbb{M}'!\varphi, w'$ *are* $\{K, [>]\}$*-bisimilar.*

2. $\mathbb{M} \Uparrow \varphi, w$ *and* $\mathbb{M}' \Uparrow \varphi, w'$ *are* $\{K, [>]\}$*-bisimilar.*

Proof. 1. If Z is a $\{K, [>]\}$-bisimulation between $\mathbb{M}, w$ and $\mathbb{M}', w'$, then one can easily show that $Z \cap ([\![\varphi]\!]^{\mathbb{M}} \times [\![\varphi]\!]^{\mathbb{M}'})$ is a $\{K, [>]\}$-bisimulation between $\mathbb{M}!\varphi, w$ and $\mathbb{M}'!\varphi, w'$.

2. If Z is a $\{K, [>]\}$-bisimulation between $\mathbb{M}, w$ and $\mathbb{M}', w'$, then one can show that Z itself is still a $\{K, [>]\}$-bisimulation between $\mathbb{M} \Uparrow \varphi, w$ and $\mathbb{M}' \Uparrow \varphi, w'$. For example, let's check the forth-condition for $>$. Consider arbitrary $u, v \in W, u' \in W'$ and suppose that $(u, u') \in Z$ and $uR_i^{\Uparrow\varphi}v$ and $v <_{i,u}^{\Uparrow\varphi} u$. Hence also uR_iv. We now make the following case distinction:

1. $u, v \in [\![\varphi]\!]^{\mathbb{M}}$. Then $v <_{i,u} u$, and hence there exists a $v' \in W'$ such that $(v, v') \in Z$, $u'R_i'v'$, and $v' <_{i,u'}' u'$. Trivially $u'R_i'^{\Uparrow\varphi}v'$. Since $(u, u'), (v, v') \in Z$ and $u, v \in [\![\varphi]\!]^{\mathbb{M}}$, also $u', v' \in [\![\varphi]\!]^{\mathbb{M}'}$. Hence from $v' <_{i,u'}' u'$ it follows that $v' {<_{i,u'}'}^{\Uparrow\varphi} u'$.

2. $u, v \notin [\![\varphi]\!]^{\mathbb{M}}$. Analogous to the previous case.

3. $u \in [\![\varphi]\!]^{\mathbb{M}}, v \notin [\![\varphi]\!]^{\mathbb{M}}$. This case cannot occur.

4. $u \notin [\![\varphi]\!]^{\mathbb{M}}, v \in [\![\varphi]\!]^{\mathbb{M}}$. Since uR_iv and $(u, u') \in Z$, it follows that there exists a $v' \in W'$ such that $(v, v') \in Z$ and $u'R_i'v'$. Since $(u, u') \in Z$ and $u \notin [\![\varphi]\!]^{\mathbb{M}}$, also $u' \notin [\![\varphi]\!]^{\mathbb{M}'}$. Analogously $v' \in [\![\varphi]\!]^{\mathbb{M}'}$. Hence $v' {<_{i,u'}'}^{\Uparrow\varphi} u'$. □

I will finish this subsection by providing an overview of the first strategy to solve the main issue of Section 4.3 (i.e. finding a structural notion of bisimulation that guarantees equivalence for conditional belief), and evaluating its advantages and disadvantages.

This strategy has two components. The first component is to impose an extra condition on epistemic plausibility models, viz. uniformity. I have argued that this is relatively harmless, because uniformity is intuitively plausible and also technically well-motivated. The second component involves what van Benthem calls "redesigning one's language to fit more standard bisimulations" (van Benthem 2002, p. 310): I introduced a new modality $[>]$, and showed that together with knowledge, it can define conditional belief (for uniform models).

The main disadvantage of this approach lies in its second component: the $[>]$-operator was introduced for the sole purpose of defining conditional belief (while maintaining a structural notion of bisimulation). In itself, however, it does not seem to have any intuitive epistemic/doxastic reading. A similar worry is voiced by Baltag and Smets, who write that "[t]he intuitive meaning of these operators [such as $[>]$] is not very clear, but they can be used to define other interesting modalities, capturing various 'doxastic attitudes'." (Baltag and Smets 2008, p. 32).

4.4.2 Assuming Connectedness

The second approach tries to preserve the advantages of the first one, while avoiding its major drawback, viz. the introduction of a new operator that lacks a direct epistemic interpretation. The basic idea is that, with an extra condition on the plausibility models, conditional belief can be reduced to knowledge and safe belief. Hence, the $\Box$-operator will now play the role of the $[>]$-operator, but unlike the $[>]$-operator, it *does* have an intuitive doxastic interpretation (viz. as Stalnakerian 'defeasible knowledge'). The extra condition on the models that we need is local connectedness:

Definition 4.13. An EPM $\mathbb{M} = \langle W, R_i, \leq_{i,w}, V\rangle_{i\in I}^{w\in W}$ is called *locally connected* iff for all agents $i \in I$ and states $w, v \in W$, the following holds:

$$\text{if } wR_iv, \text{ then } w \leq_{i,w} v \text{ or } v \leq_{i,w} w.$$

Whether local connectedness is equally harmless as uniformity is unclear. At least, local connectedness is dynamically robust:

Fact 4.5. If an EPM $\mathbb{M}$ is locally connected, then so are $\mathbb{M}!\varphi$ and $\mathbb{M} \Uparrow \varphi$.

When the models are required to be both uniform and locally connected, then conditional belief can be defined in terms of knowledge and safe belief.[18]

Fact 4.6. For every uniform and locally connected epistemic plausibility model $\mathbb{M}$, it is the case that

$$\mathbb{M} \models B_i^\alpha \varphi \leftrightarrow \big(\hat{K}_i \alpha \to \hat{K}_i(\alpha \wedge \Box_i(\alpha \to \varphi))\big).$$

Using this definability result, we can now immediately prove the analogue of Theorem 4.4; of course, since we did not introduce a new modality and a new notion of bisimulation corresponding to it, only the third item has to be reformulated. The importance of this is that when we restrict ourselves to the class of uniform and locally connected models, we can get equivalence for conditional belief by means of a structural notion of bisimulation, viz. $\{K, \Box\}$-bisimulation.

Theorem 4.5. *Consider two uniform and locally connected EPMs* $\mathbb{M} = \langle W, R_i, \leq_{i,w}, V\rangle_{i\in I}^{w\in W}$ *and* $\mathbb{M}' = \langle W', R_i', \leq_{i,w'}', V'\rangle_{i\in I}^{w'\in W'}$*, states* $w \in W$*,* $w' \in W'$*, and a relation* $Z \subseteq W \times W'$ *such that* $(w, w') \in Z$*. The following holds:*

if Z *is a* $\{K, \Box\}$*-bisimulation, then* $\mathbb{M}, w \equiv_{\mathcal{L}(K,\Box,B^c)} \mathbb{M}', w'$*.*

Proof. This follows by item 1 of Theorem 4.2 and Fact 4.6. □

Finally, it should be noted that the two model update operations studied in this chapter are both *safe* for $\{K, \Box\}$-bisimulation:

Proposition 4.5. *Consider two uniform and locally connected EPMs* $\mathbb{M} = \langle W, R_i, \leq_{i,w}, V\rangle_{i\in I}^{w\in W}$ *and* $\mathbb{M}' = \langle W', R_i', \leq_{i,w'}', V'\rangle_{i\in I}^{w'\in W'}$*, and states* $w \in W$*,* $w' \in W'$*. Suppose that* $\mathbb{M}, w$ *and* $\mathbb{M}', w'$ *are* $\{K, \Box\}$*-bisimilar, and* $\mathbb{M}, w \models \varphi$ *and* $\mathbb{M}', w' \models \varphi$*. Then:*

1. $\mathbb{M}!\varphi, w$ *and* $\mathbb{M}'!\varphi, w'$ *are* $\{K, \Box\}$*-bisimilar.*
2. $\mathbb{M} \Uparrow \varphi, w$ *and* $\mathbb{M}' \Uparrow \varphi, w'$ *are* $\{K, \Box\}$*-bisimilar.*

[18] A similar definition was already proposed in the context of modal conditional logics of normality (Boutilier 1994). Furthermore, Baltag and Smets (2008) propose the same definition. Fact 4.6 shows, however, that this definability result holds not only in Baltag/Smets-type EPMs, but also in uniform and locally connected van Benthem-type EPMs. Finally, note that Fact 4.6 does not contradict Proposition 4.1, since the definability result in Fact 4.6 holds for a *restricted* class of epistemic plausibility models, whereas the undefinability result in Proposition 4.1 holds for the *entire* class of epistemic plausibility models.

Proof. Analogous to the proof of Proposition 4.4. □

Just like in the previous subsection, I will now provide an overview of the second strategy to solve the main issue of Section 4.3. This approach reduced conditional belief to knowledge and safe belief, which are both intuitively clear epistemic/doxastic notions. Therefore, the main issue of the first approach, viz. the *ad hoc* character of its introduction of the $[>]$-operator, is avoided. In order to get the desired results about $\mathcal{L}(K, B^c)$-equivalence, it is required that the epistemic plausibility models are not only uniform, but also locally connected. The uniformity constraint inherits all of its intuitive and technical motivations from the previous subsection. Furthermore, local connectedness is dynamically robust.

Finally, it should be emphasized that the notion of $\{K, \Box\}$-bisimulation is exactly the same as that of pseudo-bisimulation (Definition 4.5), which was introduced in Subsection 4.3.1 as being the most natural notion of bisimulation for EPMs.[19] Hence, when uniformity and local connectedness are imposed, the most natural notion and the technically sound notion coincide. I take this to be an additional justification for imposing these conditions. In sum, then, the second solution seems to be preferable over the first one.

4.5 Some Methodological Reflections

In the previous sections, I have explored the model theory of (van Benthem-type) epistemic plausibility models. Here is a brief overview of the results that have been achieved. The most natural notion of bisimulation for EPMs does not work, unfortunately. Therefore, parametrized notions of bisimulations were introduced. The main problem here was the non-structural notion of B^c-bisimulation. I discussed two ways of solving this, and argued that the best solution involves restricting to the class of uniform and locally connected EPMs, so that the fully structural notion of $\{K, \Box\}$-bisimulation suffices to get equivalence for conditional belief as well. Uniformity and local connectedness have various intuitive and technical motivations. Most importantly, perhaps, by imposing these two conditions, the most natural notion of bisimulation for EPMs and the technically correct one coincide.

[19] Also recall Remark 4.1.

These results also constitute an indirect methodological argument in favor of Baltag and Smets's notion of EPM (and thus, against van Benthem's notion). To see this, note that by imposing the conditions of uniformity and local connectedness on van Benthem-type EPMs, we have actually arrived at a notion which is very close to Baltag/Smets-type EPMs. This connection can be made fully formal.

Definition 4.14. Let $\mathbb{A} = \langle W, R_i, \leq_i, V\rangle_{i\in I}$ be a Baltag/Smets-type EPM, and let $\mathbb{B} = \langle W', R'_i, \leq'_{i,w}, V'\rangle^{w'\in W'}_{i\in I}$ be a van Benthem-type EPM. Then we define:

1. $\sqsubseteq_{i,w} := \leq_i \cap (R_i[w] \times R_i[w])$,
2. $\sqsubseteq'_i := \bigcup_{w'\in W'} \Big(\leq'_{i,w'} \cap (R'_i[w'] \times R'_i[w'])\Big)$,
3. $\mathbb{A}^b := \langle W, R_i, \sqsubseteq_{i,w}, V\rangle^{w\in W}_{i\in I}$,
4. $\mathbb{B}^a := \langle W', R'_i, \sqsubseteq'_i, V'\rangle_{i\in I}$.

Theorem 4.6. *Let $\mathbb{A}$ and $\mathbb{B}$ be as in Definition 4.14. Suppose that $\mathbb{B}$ is uniform and locally connected. Let $w \in W$ be a state in $\mathbb{A}$ and let $w' \in W'$ be a state in $\mathbb{B}$. Then:*

1. *$\mathbb{A}^b$ is a uniform and locally connected van Benthem-type EPM.*
2. *$\mathbb{A}, w \equiv_{\mathcal{L}(K,B^c,\Box)} \mathbb{A}^b, w$.*
3. *$\mathbb{B}^a$ is a Baltag/Smets-type EPM.*
4. *$\mathbb{B}, w' \equiv_{\mathcal{L}(K,B^c,\Box)} \mathbb{B}^a, w'$.*

Proof. Items 1 and 3 are merely a matter of unpacking the definitions. Items 2 and 4 are proved by induction on formula complexity. □

Hence, one can move back and forth between Baltag/Smets-type EPMs and uniform and locally connected van Benthem-type EPMs. Furthermore, these constructions do not affect the logic: even the richest language studied in this chapter cannot distinguish between the model ($\mathbb{A}$ or $\mathbb{B}$) and its companion ($\mathbb{A}^b$ or $\mathbb{B}^a$, respectively). In other words, logically speaking Baltag/Smets-type EPMs can be seen as a particular *subclass* of the class of van Benthem-type EPMs.

Now consider again the results presented in Sections 4.3 and 4.4. In order to obtain a mathematically well-behaved model theory (with only structural bisimulations) for van Benthem-type EPMs, it is necessary to restrict to uniform and locally connected EPMs. Theorem 4.6 says that this particular subclass of van Benthem-type EPMs corresponds exactly with (and cannot be distinguished by the logic from) the class of Baltag/Smets-type models. Loosely speaking: in order to obtain a well-behaved model theory for van Benthem-type EPMs, we need to restrict to Baltag/Smets-type EPMs.

This seems to constitute a methodological argument in favor of Baltag/Smets-type EPMs. As an applied logician, one is broadly motivated by two conflicting desiderata. On the one hand, one looks at the concrete *applications*, and wishes to develop very expressive and general tools suitable for these applications. On the other hand, however, one is a formal logician, and thus a *mathematician*, wishing to develop a mathematically well-behaved metatheory. I have argued that Baltag and Smets's notion of EPMs hits a better equilibrium between these two desiderata than van Benthem's: it is quite expressive and general (while its restrictions are intuitively and technically motivated), but it still allows for the development of a mathematically elegant metatheory (viz. a metatheory with only structural bisimulations, and in which the most natural and the technically sound notion of bisimulation coincide).

4.6 Conclusion

The aim of this chapter has been to explore the model theory of epistemic plausibility models, which has been largely ignored in the literature so far. Because van Benthem's notion of epistemic plausibility model is the most general one, it made sense to start by investigating this type of models (rather than Baltag/Smets-type models). I focused on the notion of bisimulation, and showed that the most natural generalization of bisimulation to epistemic plausibility models fails. I then introduced parametrized bisimulations, and proved various bisimulation-implies-equivalence theorems, a Hennessy-Milner theorem, and several (un)definability results. I discussed the problems related to non-structural bisimulations for conditional belief, and presented and compared two different ways of coping with this issue: adding a modality to the language, and putting extra constraints on the models. I argued that the most successful solution involves restricting to uniform and locally connected models, and showed that such (van Benthem-type) epistemic plausibility models correspond exactly with those defined by Baltag and Smets. This can be seen as a methodological argument favoring Baltag and Smets's definition of epistemic plausibililty model over that of van Benthem.

Part II

Case Studies on the Dynamic Turn

5 The Dynamics of Aumann's Agreement Theorem

5.1 Introduction

The main goal of this chapter is to study Aumann's celebrated 'agreeing to disagree' theorem (Aumann 1976) from the perspective of epistemic logic, in particular *probabilistic dynamic epistemic logic* (PDEL). The agreement theorem, and the related no-trade theorem (Milgrom and Stokey 1982) are of central importance in game theory. Several notions connected to this theorem, such as the common prior assumption, and, especially, the notion of common knowledge, have been studied extensively by game theorists, but also by philosophers, computer scientists and logicians (Lewis 1969, Milgrom and Stokey 1982, Halpern and Moses 1990). This chapter thus establishes a new connection between the epistemic-logical and game-theoretical perspectives on (common) knowledge and related epistemic notions.

This endeavor also has definite advantages for both epistemic logic and game theory as separate disciplines. Probabilistic dynamic epistemic logic is a recent development, and to capture the agreement theorems in this framework, several technical extensions and improvements are necessary. For example, the notion of a well-behaved probabilistic Kripke model (Definition 3.4 on p. 81) was initially singled out because of its great usefulness in applications such as this one. On a more conceptual level, it will be shown how the technical results established in this chapter can be seen as an application of the *dynamic turn* in logic (van Benthem 1996, 2011). The logical perspective on the agreement theorem has definite advantages for game theorists as well, because it offers a new perspective on some methodological issues. In particular, it will be argued that the role of common knowledge is less central to the agreement theorem than is often thought.

Aumann's agreement theorem (and some of its extensions) were first studied from the perspective of dynamic epistemic logic by Dégremont and Roy (2009, 2012). They, however, did not use probabilistic Kripke models, but rather epistemic plausibility models.[1] This shift from a probabilistic to a more qualitative setting has profound consequences for the formulation of the agreement theorem. For example, Dégremont and Roy (2009, 2012)'s agreement theorems depend crucially on the assumption that the agents' plausibility orderings are *well-founded*—an order-theoretic notion that played no role in Aumann's original formulation of the agreement theorem, and that will play no role in this chapter either. [2]

The remainder of the chapter is organized as follows. Section 5.2 provides an introduction to Aumann's original agreement theorem and highlights those features that will become particularly important in later sections. Section 5.3 briefly introduces the semantic setup of probabilistic dynamic epistemic logic. I define (enriched) probabilistic Kripke frames and models, and introduce three ways of updating them: (i) carrying out experiments, (ii) public announcement of a formula φ, and (iii) a *dialogue* about a formula φ, i.e. a sequence of public annoucements that reaches a fixed point after finitely many steps. Section 5.4 contains the key results of this chapter, viz. several (dynamic) agreement theorems for probabilistic Kripke models/frames. Section 5.5 provides characterization results for all conditions of the agreement theorems, and then uses these to obtain a sound and complete dynamic agreement logic. Section 5.6 uses the formal results to show how the dynamic turn in logic can be applied to agreement theorems: I will argue that explicitly representing the dynamics that is behind Aumann's original result leads to important conceptual clarifications, for example, concerning the role and importance of common knowledge in agreement theorems. Finally, Section 5.7 wraps things up and mentions some topics for further research.

[1] See Chapter 4 for an exploration of the model theory of epistemic plausibility models.

[2] A detailed comparison between Dégremont and Roy's qualitative approach and the present probabilistic approach falls outside the scope of this chapter, but can be found in Demey (2010, ch. 6), where I argue that the probabilistic approach is to be preferred on both philosophical and technical grounds.

5.2 Aumann's Original Agreement Theorem

Aumann originally expressed his celebrated 'agreeing to disagree' theorem as follows: "If two people have the same prior, and their posteriors for an event A are common knowledge, then these posteriors are equal." (Aumann 1976, p. 1236). In other words: if two people have the same prior, then they cannot *agree* (have common knowledge of their posteriors) *to disagree* (while these posteriors are not equal). It is clear that, when phrased in this way, the agreement theorem is a *static* result: it is a conditional statement that can be expressed without any dynamic operators:

$$[\mathsf{equalpriors} \wedge C(\mathsf{posteriors})] \rightarrow \mathsf{equalposteriors}. \tag{5.1}$$

Aumann also motivates his theorem by sketching an informal scenario that embodies the intuitions behind it.[3] Roughly speaking, the scenario looks as follows. We are considering two agents, 1 and 2. Initially, they have the same probability distribution ($P_1 = P_2$). Then both agents separate, and each agent performs a (different) experiment. Immediately afterwards, the agents' probability distributions have changed due to the information that they have gained from their experiments. Because the agents performed different experiments, their probability distributions have changed in different ways, and are thus no longer identical. In particular, for some φ it holds that $P_1(\varphi) = a$ and $P_2(\varphi) = b$ (for some $a, b \in [0, 1]$), while $a \neq b$. Furthermore, since agent 1 doesn't know the outcome of agent 2's experiment, she doesn't know how agent 2's probability function has changed. A symmetric argument applies to agent 2. Hence, at this stage it is not common knowledge between both agents that $P_1(\varphi) = a$ and $P_2(\varphi) = b$. Finally, the agents start communicating with each other. Agent 1 tells agent 2 that $P_1(\varphi) = a$; on the basis of this new information, agent 2 changes her probability function, which she, in turn, communicates to agent 1, etc. At a certain point in the conversation, the agents obtain common knowledge of their probabilities. Since both agents had the same prior ($P_1 = P_2$ initially) and their posteriors have become common knowledge, Aumann's theorem now says that these probabilities have to coincide ($P_1(\varphi) = P_2(\varphi)$ in the end).

Although the formal agreement theorem is a static result, the intuitive scenario behind it clearly involves several dynamic phenomena. Two broad types of dynamics can be distinguished: (i) the *experiments* and (ii) the *communication*.

[3] A similar explanatory scenario is described more extensively by Bonanno and Nehring (1997).

This situation seems to be a good illustration of (the strong interpretation of) the recent *dynamic turn* in epistemic logic, which van Benthem (1996, p. 17) has formulated as follows:

> the motivation for standard logics often contains procedural elements present in textbook presentations — and one can make this implicit dynamics explicit.

Of course, this issue defines an entire research agenda: finding extensions (or better: *refinements*) of Aumann's original result, in which the dynamics of the scenario described above is explicitly taken into account. Game theorists such as Geanakoplos and Polemarchakis (1982), Bacharach (1985) and Parikh and Krasucki (1990) have done exactly this, focusing on the communication dynamics. Similarly, Dégremont and Roy (2009, 2012) have formalized a qualitative version of the agreement theorem in dynamic epistemic logic, again focusing on the communication dynamics.

Here, however, I will formalize Aumann's original agreement theorem in probabilistic dynamic epistemic logic, explicitly representing *both* types of dynamics (experimentation *and* communication). Furthermore, I will argue that explicitly representing this dynamics has clear conceptual advantages.[4]

5.3 The General Setup of PDEL

This section introduces the general semantic setup of probabilistic dynamic epistemic logic. This setup will be used in Section 5.4 to formalize and prove various dynamic agreement theorems.

5.3.1 Probabilistic Kripke Models

I first introduce (enriched) probabilistic Kripke frames and models. The focus will be on the two agent-case (this will suffice for the statement of the agreement theorems); generalizations to any (finite) number of agents are straightforward. As usual, we also fix a countably infinite set Prop of atomic propositions.

[4] I will thus not deal with any of the stronger and more general agreement theorems that exist in the game-theoretical literature (Bacharach 1985, Bonanno and Nehring 1997, Feinberg 2000), because the methodological claim about the advantages of explicitly representing the dynamics can already be made for Aumann's original result.

Definition 5.1. An *enriched probabilistic Kripke frame* (for two agents) is a tuple $\mathbb{F} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2 \rangle$, where W is a non-empty finite set of states, R_1, R_2, E_1 and E_2 are equivalence relations on W, and μ_1 and μ_2 assign to each world $w \in W$ a probability mass function $\mu_i(w) \colon W \to [0,1]$ that satisfies the following two conditions:

- $\mu_i(w)(w) > 0$,
- $\mu_i(w)(v) = 0$ for all $v \in W$ such that $(w, v) \notin R_i$.

Definition 5.2. An *enriched probabilistic Kripke model* is a tuple $\mathbb{M} = \langle \mathbb{F}, V \rangle$, where $\mathbb{F}$ is an enriched probabilistic Kripke frame (with set of states W) and $V \colon \mathsf{Prop} \to \wp(W)$ is a valuation.

An enriched probabilistic Kripke frame $\mathbb{F} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2 \rangle$ thus consists of a probabilistic Kripke frame $\mathbb{F}^* = \langle W, R_1, R_2, \mu_1, \mu_2 \rangle$ (see Definition 3.1 on p. 75), together with two equivalence relations E_1 and E_2, whose meaning will be discussed below. Note that the 'base frame' $\mathbb{F}^*$ is required to be well-behaved (see Definition 3.4 on p. 81); I argued in Chapter 3 that the conditions involved in being well-behaved are intuitively plausible and technically well-motivated. Furthermore, in the next subsections, I will introduce several ways of updating probabilistic Kripke models, all of which change the agents' probabilities via Bayesian conditionalization. This requires, however, that $\mu_i(w)(X) > 0$ for several sets $X \subseteq W$. Assuming well-behavedness is an easy way to ensure that $\mu_i(w)(X) > 0$ for all the relevant sets X.

The probabilistic Kripke frames and models that will be used in the remainder of this chapter are always the enriched ones defined above; therefore I will henceforth omit the extra qualifier and simply talk about 'probabilistic Kripke frames/models'.

As usual, R_i is agent i's epistemic accessibility relation: $(w, v) \in R_i$ means that i cannot epistemically distinguish between states w and v. The E_i-relation represents the structure of agent i's experiment: $(w, v) \in E_i$ means that agent i's experiment does not differentiate between w and v. The relation R_i thus captures agent i's information before any dynamics has taken place, while E_i captures the information that she will obtain by carrying out her experiment.[5] Intuitively, one

[5] In game-theoretical contexts, R_i is usually implicitly taken to be the universal relation $W \times W$ (and is therefore often not explicitly mentioned at all), while the equivalence relation E_i is identified with the partition Π_i that it generates. Furthermore, note that if $R_i = W \times W$, then the second well-behavedness condition of Definition 5.1 is vacuously satisfied.

can think of carrying out an experiment as asking a question to nature. This informal analogy carries over to the formal level: the *experiment relations E_i* play the same role in the current framework as the *issue relations* do in dynamic epistemic logics of questions (van Benthem and Minică 2009, 2012).

Similarly, the probability mass function $\mu_i(w)$ represents agent i's subjective probabilities (at state w) before any dynamics has taken place. For example, $\mu_i(w)(v) = a$ means that at state w, agent i assigns subjective probability a to state v being the actual state.

The static language $\mathcal{L}$ is defined by means of the following BNF:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi\wedge\varphi) \mid K_i\varphi \mid R_i\varphi \mid C^{\varphi}\varphi \mid X^{\varphi}\varphi \mid a_1P_i(\varphi_1)+\cdots+a_nP_i(\varphi_n) \geq k$$

where $i \in \{1, 2\}$, $1 \leq n < \omega$ and $a_1, \dots, a_n, k \in \mathbb{Q}$).

As usual, $K_i\varphi$ means that agent i knows that φ. Furthermore, we have the *relativized common knowledge* operator $C^{\varphi}\psi$, which intuitively says that if φ is announced, then it becomes common knowledge (among agents 1 and 2) that ψ was the case before the announcement. The reason for introducing a relativized instead of an ordinary common knowledge operator is well-known: because of its higher expressivity, relativized common knowledge allows for the formulation of a reduction axiom under public announcements (van Benthem et al. 2006).

Knowledge and (relativized) common knowledge have 'post-experimental' counterparts: $R_i\varphi$ and $X^{\varphi}\psi$.[6] First, $R_i\varphi$ says that after carrying out the experiments, agent i will know that φ was the case before the experiments. Second, $X^{\varphi}\psi$ says that after carrying out the experiments, if φ is announced, then it becomes common knowledge (among agents 1 and 2) that ψ was the case before the experiments and the announcement. These operators 'pre-encode' the effects of the experiments in the static language, and will thus enable us to express reduction axioms for the dynamic experimentation operator that will be introduced in the next subsection.

Ordinary (post-experimental) common knowledge can be defined as $C\varphi := C^{\top}\varphi$ and $X\varphi := X^{\top}\varphi$. Furthermore (post-experimental) general knowledge is defined by putting $E\varphi := K_1\varphi \wedge K_2\varphi$ and $F\varphi := R_1\varphi \wedge R_2\varphi$.

[6] Hence there are two R_i's: on the one hand, R_i is agent i's epistemic accessibility relation in a probabilistic Kripke model $\mathbb{M}$; on the other hand, R_i is a unary modal operator of the language $\mathcal{L}$. The main reason for not using another letter for the post-experimental knowledge operator is to ensure uniformity of notation with van Benthem and Minică (2012). I trust that the meaning of R_i will always be clear from the context.

Formulas of the form $a_1 P_i(\varphi_1) + \cdots + a_n P_i(\varphi_n) \geq k$ are called *i-probability formulas*, and have been discussed extensively in Chapter 3. In particular, recall the following definition of conditional probabilities in $\mathcal{L}$:

$$P_i(\varphi \mid \psi) \geq k \;:\equiv\; P_i(\varphi \wedge \psi) - kP_i(\psi) \geq 0.$$

Since the various types of dynamics that will be introduced in the next two subsections all involve Bayesian conditionalization (and thus conditional probabilities), this definition will often be very convenient.

Consider a probabilistic Kripke model $\mathbb{M} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V\rangle$ and a state $w \in W$. Now and in the remainder of this chapter, I will often abbreviate $R := R_1 \cup R_2$, $R^e := (R_1 \cap E_1) \cup (R_2 \cap E_2)$, and $\llbracket \varphi \rrbracket^{\mathbb{M}} := \{v \in W \mid \mathbb{M}, v \models \varphi\}$. Furthermore, for any binary relation $\mathcal{R} \subseteq W \times W$ and state $w \in W$, I abbreviate $\mathcal{R}[w] := \{v \in W \mid (w, v) \in \mathcal{R}\}$, and write $\mathcal{R}^+$ and $\mathcal{R}^*$ for the transitive and reflexive transitive closure of $\mathcal{R}$, respectively. We are now ready to state the semantics of $\mathcal{L}$:

$$\begin{array}{lll}
\mathbb{M}, w \models p & \text{iff} & w \in V(p), \\
\mathbb{M}, w \models \neg\varphi & \text{iff} & \mathbb{M}, w \not\models \varphi, \\
\mathbb{M}, w \models \varphi \wedge \psi & \text{iff} & \mathbb{M}, w \models \varphi \text{ and } \mathbb{M}, w \models \psi, \\
\mathbb{M}, w \models K_i\varphi & \text{iff} & \forall v \in W : (w, v) \in R_i \Rightarrow \mathbb{M}, v \models \varphi, \\
\mathbb{M}, w \models C^{\varphi}\psi & \text{iff} & \forall v \in W : (w, v) \in \big(R \cap (W \times \llbracket \varphi \rrbracket^{\mathbb{M}})\big)^+ \Rightarrow \mathbb{M}, v \models \psi, \\
\mathbb{M}, w \models R_i\varphi & \text{iff} & \forall v \in W : (w, v) \in R_i \cap E_i \Rightarrow \mathbb{M}, v \models \varphi, \\
\mathbb{M}, w \models X^{\varphi}\psi & \text{iff} & \forall v \in W : (w, v) \in \big(R^e \cap (W \times \llbracket \varphi \rrbracket^{\mathbb{M}})\big)^+ \Rightarrow \mathbb{M}, v \models \psi, \\
\mathbb{M}, w \models \sum_\ell a_\ell P_i(\varphi_\ell) \geq k & \text{iff} & \sum_\ell a_\ell \mu_i(w)(\llbracket \varphi_\ell \rrbracket^{\mathbb{M}}) \geq k.
\end{array}$$

Truth and validity at a model $\mathbb{M}$, a frame $\mathbb{F}$, and a class of frames $\mathcal{C}$ are defined as usual:

$$\begin{array}{lll}
\mathbb{M} \models \varphi & \text{iff} & \mathbb{M}, w \models \varphi \text{ for all states } w \text{ in the domain of } \mathbb{M}, \\
\mathbb{F} \models \varphi & \text{iff} & \mathbb{F}, V \models \varphi \text{ for all valuations } V \text{ on the frame } \mathbb{F}, \\
\mathcal{C} \models \varphi & \text{iff} & \mathbb{F} \models \varphi \text{ for all frames } \mathbb{F} \text{ in } \mathcal{C}.
\end{array}$$

5.3.2 Dynamics: The Experimentation Phase

I will now show how to model the first type of dynamics described in Section 5.2, viz. carrying out the experiments. Syntactically, we add a new dynamic operator [EXP] to the language $\mathcal{L}$, thus obtaining the language $\mathcal{L}([\mathrm{EXP}])$. The [EXP]-operator says that both agents perform their experiments; hence, $[\mathrm{EXP}]\,\varphi$ is to be read as: 'after the agents have performed their experiments, φ holds'. The semantic clause for the [EXP]-operator involves going from the model $\mathbb{M}$ to the updated model $\mathbb{M}^e$, which is formally introduced in Definition 5.3:

$$\mathbb{M}, w \models [\text{EXP}]\,\varphi \quad \text{iff} \quad \mathbb{M}^e, w \models \varphi.$$

Definition 5.3. Let $\mathbb{M} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V\rangle$ be an arbitrary probabilistic Kripke model. The updated model $\mathbb{M}^e := \langle W^e, R_1^e, R_2^e, E_1^e, E_2^e, \mu_1^e, \mu_2^e, V^e\rangle$ is defined as follows:

- $W^e := W$,
- $R_i^e := R_i \cap E_i$ and $E_i^e := E_i$ for all $i \in I$,
- $\mu_i^e(w)(v) := \frac{\mu_i(w)(\{v\}\cap E_i[w])}{\mu_i(w)(E_i[w])}$ for all $i \in I$ and $w, v \in W^e$,
- $V^e := V$.

It is easy to check that if $\mathbb{M}$ is a probabilistic Kripke model, then $\mathbb{M}^e$ is a probabilistic Kripke model as well; see Demey (2010, Lemma 9). The model $\mathbb{M}^e$ represents the world and the agents' knowledge and probabilities after the agents have each carried out their experiment.

Recall that I abbreviated $R^e = (R_1 \cap E_1) \cup (R_2 \cap E_2)$ in the previous subsection. Applying Definition 5.3, this can now be rewritten as $R^e = R_1^e \cup R_2^e$, which is structurally analogous to the other abbreviation: $R = R_1 \cup R_2$.

I will now justify the definition of the model update operation $\mathbb{M} \mapsto \mathbb{M}^e$ by explaining the intuitions behind it, and by showing that it leads to the right results in a concrete scenario. Carrying out the experiments does not change the set of possible states. Experiment 1 intersects agent 1's accessibility relation R_1 with the experiment relation E_1, and leaves agent 2's accessibility relation unchanged ($R_1^e = R_1 \cap E_1$). Hence, after carrying out her experiment, agent 1 cannot distinguish between states w and v iff (i) before the experiment, she could not distinguish between those states, *and* (ii) her experiment does not differentiate between them. Dually: after carrying out her experiment, agent 1 knows that φ iff (i) she already knew that φ before the experiment (by perfect recall), *or* she has learned that φ is the case by performing her experiment. Symmetric remarks hold for experiment 2.[7] This closely resembles the description by Bonanno and Nehring (1997) of the experiments as imposing a partition on the model.[8]

Let's now turn to the probabilistic component. The definition of $\mu_i^e(w)$ can be rewritten in terms of conditional probabilities: $\mu_i^e(w)(x) = \mu_i(w)(x \mid E_i[w])$; i.e. agent i conditionalizes on the information that she has gained by performing her experiment (viz. the information that the actual world w belongs to the cell $E_i[w]$ of the partition induced by her experiment). This captures the idea that the agents process new information by means of Bayesian updating.[9]

Example 5.1. Consider the following scenario. Agent 1 doesn't know whether p is the case, i.e. she cannot distinguish between p-states and $\neg p$-states. (At the actual state, p is true.) Furthermore, agent 1 has no specific reason to think that one state is more probable than any other; therefore it is reasonable for her to assign equal probabilities to all states. Finally, although agent 1 does not know whether p is the case, she has an experiment that discriminates between p-states and $\neg p$-states, and that thus, when carried out, will allow her to find out whether p is the case. (Agent 2 does not play a role in this scenario.)

Consider the model $\mathbb{M} := \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V\rangle$, with $W := \{w, v\}$, $R_1 := W \times W$, $E_1 = \{(w,w),(v,v)\}$, $\mu_1(w)(w) = \mu_1(w)(v) = 0.5$, and $V(p) = \{w\}$ (the definitions of $\mu_1(v), R_2, E_2$ and μ_2 are irrelevant). It is easy to see that this model is a faithful representation of the above scenario. Consider, for example:

$$\mathbb{M}, w \models \neg K_1 p \wedge \neg K_1 \neg p \wedge P_1(p) = 0.5 \wedge P_1(\neg p) = 0.5.$$

Now suppose that the agents carry out their experiments, i.e. consider the updated model $\mathbb{M}^e$. Applying Definition 5.3, it is easy to see that

$$\mathbb{M}, w \models [\text{EXP}] \big(K_1 p \wedge P_1(p) = 1 \wedge P_1(\neg p) = 0\big).$$

[7] I already discussed the analogy between carrying out an experiment and asking a question. Modeling the experiments as intersecting R_i with E_i is analogous to the 'resolve' action in the dynamic epistemic logic of questions (van Benthem and Minică 2012, Definition 6): carrying out an experiment means getting an answer to a question posed to nature.

[8] If one assumes that $R_i = W \times W$ (recall Footnote 5), then $R_i^e = E_i$, i.e. agent i's knowledge after the experiments consists entirely of what she has learned from carrying out her experiment. Furthermore, it then holds that $\mu_i^e(w)(v) = \mu_i(w)(v \mid \Pi_i(w))$, where $\Pi_i(w)$ is the cell of the partition Π_i generated by E_i that contains w (obviously, $\Pi_i(w) = E_i[w]$).

[9] Note that μ_i^e is well-defined (no dangerous division by 0): since E_i is an equivalence relation, it holds that $w \in E_i[w]$, and hence $\mu_i(w)(E_i[w]) \geq \mu_i(w)(w) > 0$.

So after carrying out her experiment, agent 1 has come to know that p is in fact the case. She has also adjusted her probabilities: she now assigns probability 1 to p being true, and probability 0 to p being false. These are the results that we would expect intuitively. Therefore, Definition 5.3 seems to be a natural way of representing the experimentation dynamics: it makes the intuitively right 'predictions' about the agents' knowledge and probabilities.

5.3.3 Dynamics: The Communication Phase

I will now show how to model the second type of dynamics described in Section 5.2, viz. the communication phase. Intuitively, the communication protocol will be treated as a *dialogue about* φ, i.e. a sequence in which the agents each repeatedly communicate the subjective probability they assign to φ (at that point in the dialogue). Single steps in the dialogue are modeled as public announcements.

5.3.3.1 Public Announcements

I first introduce single public announcements. Syntactically, we add a new dynamic operator $[!\cdot]$ to the language $\mathcal{L}([\text{EXP}])$, thus obtaining the language $\mathcal{L}([\text{EXP}], [!\cdot])$. The public announcement operator $[!\varphi]$ says that the formula φ is truthfully and publicly announced to all agents. Hence, $[!\varphi]\psi$ is to be read as: 'after the truthful public announcement of φ, it will be the case that ψ'. The truthfulness of the announcement is captured by means of a precondition in the semantic clause; this clause involves going from the model $\mathbb{M}$ to the updated model $\mathbb{M}|\varphi$, which is defined immediately afterwards:

$$\mathbb{M}, w \models [!\varphi]\psi \quad \text{iff} \quad \text{if } \mathbb{M}, w \models \varphi \text{ then } \mathbb{M}|\varphi, w \models \psi.$$

Definition 5.4. Let $\mathbb{M} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V\rangle$ be an arbitrary probabilistic Kripke model and $\varphi \in \mathcal{L}([\text{EXP}], [!\cdot])$ an arbitrary formula such that $[\![\varphi]\!]^{\mathbb{M}} \neq \emptyset$. The updated model $\mathbb{M}|\varphi = \langle W^\varphi, R_1^\varphi, R_2^\varphi, E_1^\varphi, E_2^\varphi, \mu_1^\varphi, \mu_2^\varphi, V^\varphi\rangle$ is defined as follows:

- $W^\varphi := [\![\varphi]\!]^{\mathbb{M}} = \{w \in W \mid \mathbb{M}, w \models \varphi\}$,
- $R_i^\varphi := R_i \cap ([\![\varphi]\!]^{\mathbb{M}} \times [\![\varphi]\!]^{\mathbb{M}})$ for all $i \in I$,
- $E_i^\varphi := E_i \cap ([\![\varphi]\!]^{\mathbb{M}} \times [\![\varphi]\!]^{\mathbb{M}})$ for all $i \in I$,

- $\mu_i^\varphi(w)(v) := \frac{\mu_i(w)(v)}{\mu_i(w)([\![\varphi]\!]^{\mathbb{M}})}$ for all $i \in I$ and $w, v \in W^\varphi$,
- $V^\varphi(p) := V(p) \cap [\![\varphi]\!]^{\mathbb{M}}$ for all $p \in \mathsf{Prop}$.

It is easy to check that if $\mathbb{M}$ is a probabilistic Kripke model, then $\mathbb{M}|\varphi$ is a probabilistic Kripke model as well; see Demey (2010, Lemma 16). This model represents the world and the agents' knowledge and probabilities after the public announcement of φ.

This definition is just a special case of public announcements in well-behaved probabilistic Kripke models (see Definition 3.6 on p. 86), so I will not say much about it. Note that the experiment relations are treated exactly like the epistemic accessibility relations: the main effect of the public announcement of φ is that all $\neg\varphi$-states get deleted, and hence, any R_i- or E_i-links that involved $\neg\varphi$-states are deleted as well.

Definition 5.4 fits well with our intuitive idea of what a public announcement of φ is, and how it influences the agents' knowledge and probabilities. One can easily construct scenarios similar to Example 5.1, in which the 'predictions' about the agents' knowledge and probabilities made by Definition 5.4 match perfectly with our intuitive expectations (for example, see Example 3.1 on p. 87).

5.3.3.2 Dialogues

I will now move from *single* public announcements to *sequences* of public announcements. I will focus on one particular type of such sequences, which will be called a *dialogue about* φ. In a dialogue about φ, each agent repeatedly announces the probability she assigns to φ (at that step in the dialogue). I will show that such dialogues reach a fixed point after finitely many steps.

Consider a probabilistic Kripke model $\mathbb{M} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V\rangle$, a state $w \in W$ and a formula φ. Note that there are unique $a, b \in \mathbb{R}$ such that $\mu_1(w)([\![\varphi]\!]^{\mathbb{M}}) = a$ and $\mu_2(w)([\![\varphi]\!]^{\mathbb{M}}) = b$. The sentence $d(\mathbb{M}, w, \varphi)$ is now defined as follows:[10]

$$d(\mathbb{M}, w, \varphi) := P_1(\varphi) = a \wedge P_2(\varphi) = b.$$

[10] Note that I have tacitly moved outside the official object language here, because the formula $P_1(\varphi) = a \wedge P_2(\varphi) = b$ involves real numbers which might not be rational ($a, b \in \mathbb{R} - \mathbb{Q}$), whereas the official object language only contains *rational* numbers. Technically speaking, this can be 'repaired' (cf. Demey 2010), and it does not matter from a modeling perspective, so I will not dwell on it further.

Note that for any model $\mathbb{M}$, state w of $\mathbb{M}$, and formula φ, it holds—by definition of $d(\mathbb{M}, w, \varphi)$—that

$$\mathbb{M}, w \models d(\mathbb{M}, w, \varphi). \tag{5.2}$$

A single step in the dialogue consists of both agents publicly announcing the probabilities they assign to φ (at that point in the dialogue). In other words, a single step consists of the public announcement of the sentence $P_1(\varphi) = a \wedge P_2(\varphi) = b$, for the unique $a, b \in \mathbb{R}$ that make this sentence true.

For any probabilistic Kripke model $\mathbb{M}$ that contains w, we define $f_{w,\varphi}(\mathbb{M})$ to be the result of publicly announcing the sentence $d(\mathbb{M}, w, \varphi)$ in the model $\mathbb{M}$ (see Definition 5.4). Symbolically:

$$f_{w,\varphi}(\mathbb{M}) := \mathbb{M} \,|\, d(\mathbb{M}, w, \varphi).$$

Since $\mathbb{M}, w \models d(\mathbb{M}, w, \varphi)$, the state w still belongs to $\mathbb{M}|d(\mathbb{M}, w, \varphi)$, and thus $f_{w,\varphi}$ can be applied to this model as well. In general, $f^n_{w,\varphi}(\mathbb{M})$ is a well-defined probabilistic Kripke model for all $n \geq 0$. For example, unraveling the definitions, we see that

$$f^2_{w,\varphi}(\mathbb{M}) = f_{w,\varphi}(f_{w,\varphi}(\mathbb{M})) = \big(\mathbb{M} \,|\, d(\mathbb{M}, w, \varphi)\big) \,|\, d\big((\mathbb{M} \,|\, d(\mathbb{M}, w, \varphi)), w, \varphi\big).$$

The entire dialogue about φ will now be modeled as a sequence in which the agents repeatedly announce the probabilities they assign to φ. Consider a probabilistic Kripke model $\mathbb{M}$ that contains the state w. By repeatedly applying $f_{w,\varphi}$ to $\mathbb{M}$ we obtain a sequence which looks as follows:

$$\mathbb{M} \mapsto f_{w,\varphi}(\mathbb{M}) \mapsto f^2_{w,\varphi}(\mathbb{M}) \mapsto f^3_{w,\varphi}(\mathbb{M}) \mapsto f^4_{w,\varphi}(\mathbb{M}) \mapsto \cdots$$

The following lemma says that the models in this sequence do not continue to change ad infinitum, i.e. the dialogue reaches a *fixed point* after finitely many steps.[11]

Lemma 5.1. *Consider a probabilistic Kripke model* $\mathbb{M}$ *that contains the state* w. *Then there exists an* $n \in \mathbb{N}$ *such that* $f^n_{w,\varphi}(\mathbb{M}) = f^{n+1}_{w,\varphi}(\mathbb{M})$.

[11] Recall that probabilistic Kripke models are assumed to be finite; see Definitions 5.1 and 5.2. If infinite models are allowed as well, then Lemma 5.1 no longer holds. Nevertheless, because the submodels of $\mathbb{M}$ (ordered by the submodel relation) form a chain-complete poset and $f_{w,\varphi}$ is a deflationary map on this poset, the Bourbaki-Witt theorem (Bourbaki 1949) still guarantees that $f_{w,\varphi}$ has a fixed point; however, it might take transfinitely many steps to reach this fixed point. From an application-oriented perspective, such transfinite dialogues make little sense, and I will therefore not pursue this topic any further.

Proof. For any probabilistic Kripke model $\mathbb{K}$, let $|\mathbb{K}|$ denote the number of states in $\mathbb{K}$. For a reductio, suppose that for all $n \in \mathbb{N}$: $f^n_{w,\varphi}(\mathbb{M}) \neq f^{n+1}_{w,\varphi}(\mathbb{M})$. It follows from the definition of $f_{w,\varphi}$ that if $f^n_{w,\varphi}(\mathbb{M}) \neq f^{n+1}_{w,\varphi}(\mathbb{M})$, then $|f^n_{w,\varphi}(\mathbb{M})| > |f^{n+1}_{w,\varphi}(\mathbb{M})|$. Hence we find that

$$|\mathbb{M}| > |f_{w,\varphi}(\mathbb{M})| > |f^2_{w,\varphi}(\mathbb{M})| > |f^3_{w,\varphi}(\mathbb{M})| > \cdots$$

This contradicts the fact that $\mathbb{M}$ has, by definition, only finitely many states. □

I will now provide an exact definition of the communication dynamics. Syntactically, we add the [DIAL($\cdot$)]-operator to the language $\mathcal{L}([\text{EXP}], [!\,\cdot])$, thus obtaining the language $\mathcal{L}([\text{EXP}], [!\,\cdot], [\text{DIAL}(\,\cdot\,)])$; this is the final, and most expressive, language considered in this chapter. The [DIAL(φ)]-operator says that both agents carry out a dialogue about φ, i.e. they repeatedly announce the probabilities they assign to φ, until a fixed point is reached (Lemma 5.1 guarantees that such a fixed point will indeed always be reached after finitely many steps). Hence, [DIAL(φ)]ψ is to be read as: 'after the agents have carried out a dialogue about φ, it will be the case that ψ'.

The semantic clause for [DIAL(φ)] involves going to the fixed point model $\mathbb{M}^{dial_w(\varphi)}$, which is defined immediately afterwards:

$$\mathbb{M}, w \models [\text{DIAL}(\varphi)]\psi \quad \text{iff} \quad \mathbb{M}^{dial_w(\varphi)}, w \models \psi.$$

Definition 5.5. Let $\mathbb{M} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V\rangle$ be an arbitrary probabilistic Kripke model, $w \in W$ an arbitrary state, and φ an arbitrary formula. Then we define $\mathbb{M}^{dial_w(\varphi)} := f^n_{w,\varphi}(\mathbb{M})$, where n is the least number such that $f^n_{w,\varphi}(\mathbb{M}) = f^{n+1}_{w,\varphi}(\mathbb{M})$ (this number is guaranteed to exist by Lemma 5.1).

Remark 5.1. Recall that public announcements are assumed to be *truthful*. Furthermore, a dialogue about φ is modeled as a sequence of public announcements. However, the semantics of [DIAL(φ)] does not involve any preconditions. The reason for this is that the formulas being announced throughout the dialogue sequence are true *by definition*, as was stated in (5.2). Because a dialogue about φ always takes on this form (it will never involve the announcement of other formulas than $d(\mathbb{K}, w, \varphi)$, for probabilistic Kripke models $\mathbb{K}$ that contain the state w), the truth precondition can be safely left out.

Remark 5.2. Given the move from single public announcements to sequences of public announcements that was just described, one might wonder why we considered only single experiments (and no sequences of experiments) in Subsection 5.3.2. However, since experiments typically concern *factual* propositions (Boolean combinations of propositional atoms), the single update $\mathbb{M} \mapsto \mathbb{M}^e$ can be interpreted as 'summarizing' an entire sequence of experiments. Consider, for example, the sequence consisting of a binary experiment 'is p the case?' followed by another binary experiment 'is q the case?'; by performing this sequence of experiments, the agent will first find out whether p is the case, and then whether q is the case. Because the first experiment does not change the truth value of p, this sequence of experiments can be replaced by one single, complex experiment that allows the agent to discover the truth values of p and q simultaneously. Formally, this is achieved by putting $E_i = [\![p \wedge q]\!]^2 \cup [\![p \wedge \neg q]\!]^2 \cup [\![\neg p \wedge q]\!]^2 \cup [\![\neg p \wedge \neg q]\!]^2$.[12]

In principle, one can also compress a sequence of public announcements into a single public announcement. However, such a compression will be much more intricate, because the public announcement of a formula might change the truth value of that formula, and thus influence the set of formulas that can be announced next. (Recall the definitions of $d(\mathbb{M}, w, \varphi)$ and $f_{w,\varphi}$.)

5.4 Agreement Theorems in PDEL

Using the semantic setup introduced in the previous section, I will now formulate and prove various dynamic agreement theorems in probabilistic dynamic epistemic logic. In Subsection 5.4.1, I discuss agreement theorems that make the experimentation dynamics explicit, but still leave the communication implicit. In Subsection 5.4.2, I build on this and formulate agreement theorems that make both the experimentation *and* the communication dynamics explicit.

5.4.1 Only Experimentation

Before turning to the first agreement theorem in probabilistic dynamic epistemic logic, I formulate two easy auxiliary lemmas:

[12] This strategy is no longer available if we restrict to binary experiments (as will be done in Subsection 5.5.1).

Lemma 5.2. *Let* $\mathbb{M} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V\rangle$ *be an arbitrary probabilistic Kripke model and* $w \in W$ *a state of* $\mathbb{M}$. *Then for* $i = 1, 2$, *the set* $R^*[w]$ *can be finitely partitioned in cells of the form* $R_i[v_\ell]$; *i.e. it can be expressed as* $R^*[w] = R_i[v_1] \cup \cdots \cup R_i[v_m]$, *with all the* $R_i[v_\ell]$ *pairwise disjoint.*

Proof. Consider an arbitrary agent $i \in \{1, 2\}$. Since R^* is the reflexive transitive closure of $R = R_1 \cup R_2$ and R_i is reflexive, it holds that $R^*[w] = \bigcup_{x \in R^*[w]} R_i[x]$. Since W is finite, $R^*[w]$ is finite as well and can thus be written as $R^*[w] = \{v_1, \ldots, v_n\}$, and hence $R^*[w] = R_i[v_1] \cup \cdots \cup R_i[v_n]$. Since R_i is an equivalence relation, we know that the $R_i[v_\ell]$ are mutually exclusive and pairwise disjoint. By systematically deleting the 'redundant' $R_i[v_\ell]$ (i.e. if $\ell \neq m$ and $R_i[v_\ell] = R_i[v_m]$, then delete exactly one of $R_i[v_\ell]$ and $R_i[v_m]$; repeat until stabilization), we obtain a (finite) partition of $R^*[w]$ into cells of the form $R_i[v_\ell]$. □

Lemma 5.3. *Let* $\mathbb{M} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V\rangle$ *be an arbitrary probabilistic Kripke model and* $w \in W$ *a state of* $\mathbb{M}$. *Consider sets* $X, Y \subseteq W$ *and a partition* $\{Y_1, \ldots, Y_m\}$ *of* Y. *Furthermore, assume that for each cell* Y_ℓ *of the partition it holds that* $\mu_i(w)(Y_\ell) > 0$ *and that* $\frac{\mu_i(w)(X\cap Y_\ell)}{\mu_i(w)(Y_\ell)} = a$. *Then also* $\mu_i(w)(Y) > 0$ *and* $\frac{\mu_i(w)(X\cap Y)}{\mu_i(w)(Y)} = a$.

Proof. Since $Y_1 \subseteq Y$, it follows that $0 < \mu_i(w)(Y_1) \leq \mu_i(w)(Y)$; this proves the first part. For the second part, note that

$$\begin{aligned}
\mu_i(w)(X \cap Y) &= \mu_i(w)\big(X \cap \textstyle\bigcup_{\ell=1}^m Y_\ell\big) \\
&= \mu_i(w)\big(\textstyle\bigcup_{\ell=1}^m (X \cap Y_\ell)\big) \\
&= \textstyle\sum_{\ell=1}^m \mu_i(w)(X \cap Y_\ell) \\
&= \textstyle\sum_{\ell=1}^m a \cdot \mu_i(w)(Y_\ell) \\
&= a \cdot \textstyle\sum_{\ell=1}^m \mu_i(w)(Y_\ell) \\
&= a \cdot \mu_i(w)\big(\textstyle\bigcup_{\ell=1}^m Y_\ell\big) \\
&= a \cdot \mu_i(w)(Y).
\end{aligned}$$

Since $\mu_i(w)(Y) > 0$, it follows that $\frac{\mu_i(w)(X\cap Y)}{\mu_i(w)(Y)} = a$. □

This brings us to the first agreement theorem:

Theorem 5.1. *Let* $\mathbb{M} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V\rangle$ *be an arbitrary probabilistic Kripke model and* $w \in W$ *a state of* $\mathbb{M}$*. Suppose that the following conditions hold:*

1. $\mu_1(w) = \mu_2(w)$,
2. *for all* $v \in R^*[w]$: $\mu_i(w) = \mu_i(v)$.

Then we have:

$$\mathbb{M}, w \models [\text{EXP}]\, C(P_1(\varphi) = a \wedge P_2(\varphi) = b) \to a = b.$$

Proof. Assume that $\mathbb{M}, w \models [\text{EXP}]\, C(P_1(\varphi) = a \wedge P_2(\varphi) = b)$; it will be shown that $\mathbb{M}, w \models a = b$, i.e. simply that $a = b$.

Applying Lemma 5.2 to $\mathbb{M}^e$ (for agent 1), express $(R^e)^*[w] = R_1^e[v_1] \cup \cdots \cup R_1^e[v_m]$, with all the $R_1^e[v_\ell]$ pairwise disjoint. Now consider an arbitrary cell $R_1^e[v_\ell]$ of this partition ($1 \leq \ell \leq m$). Since R_1^e is reflexive, we have $v_\ell \in R_1^e[v_\ell] \subseteq (R^e)^*[w]$. Since $\mathbb{M}, w \models [\text{EXP}]\, C(P_1(\varphi) = a \wedge P_2(\varphi) = b)$, we get $\mathbb{M}^e, w \models C(P_1(\varphi) = a \wedge P_2(\varphi) = b)$, so $v_\ell \in (R^e)^*[w]$ implies that $\mathbb{M}^e, v_\ell \models P_1(\varphi) = a \wedge P_2(\varphi) = b$. Hence $\mu_1^e(v_\ell)([\![\varphi]\!]^{\mathbb{M}^e}) = a$ (†). Note that $R^e = (R_1 \cap E_1) \cup (R_2 \cap E_2) \subseteq R_1 \cup R_2 = R$, and hence $v \in (R^e)^*[w] \subseteq R^*[w]$, so condition 2 of this theorem applies to v_ℓ, i.e. $\mu_1(w) = \mu_1(v_\ell)$ (‡). We now have:

$$\begin{array}{rcll} a & = & \mu_1^e(v_\ell)([\![\varphi]\!]^{\mathbb{M}^e}) & (\dagger) \\ & = & \frac{\mu_1(v_\ell)([\![\varphi]\!]^{\mathbb{M}^e} \cap E_1[v_\ell])}{\mu_1(v_\ell)(E_1[v_\ell])} & \text{(Definition 5.3)} \\ & = & \frac{\mu_1(v_\ell)([\![\varphi]\!]^{\mathbb{M}^e} \cap E_1[v_\ell] \cap R_1[v_\ell])}{\mu_1(v_\ell)(E_1[v_\ell] \cap R_1[v_\ell])} & \text{(Lemma 3.2, p. 80)} \\ & = & \frac{\mu_1(w)([\![\varphi]\!]^{\mathbb{M}^e} \cap R_1^e[v_\ell])}{\mu_1(w)(R_1^e[v_\ell])}. & (\ddagger) \end{array}$$

(Note that $\mu_1(w)(R_1^e[v_\ell]) = \mu_1(v_\ell)(R_1[v_\ell] \cap E_1[v_\ell]) = \mu_1(v_\ell)(E_1[v_\ell]) > 0$.) As the cell $R_1^e[v_\ell]$ was chosen arbitrarily, this holds for all cells of the partition of $(R^e)^*[w]$. By Lemma 5.3 it now follows that $\mu_1(w)\big((R^e)^*[w]\big) > 0$ and

$$\frac{\mu_1(w)\big([\![\varphi]\!]^{\mathbb{M}^e} \cap (R^e)^*[w]\big)}{\mu_1(w)\big((R^e)^*[w]\big)} = a. \tag{5.3}$$

It is easy to see that the entire argument presented above can also be carried out for agent 2. The conclusion of this second, analogous argument will be that

$$\frac{\mu_2(w)\big(\llbracket \varphi \rrbracket^{\mathbb{M}^e} \cap (R^e)^*[w]\big)}{\mu_2(w)\big((R^e)^*[w]\big)} = b. \tag{5.4}$$

Now recall condition 1 of this theorem: $\mu_1(w) = \mu_2(w)$. Hence (5.3) and (5.4) together imply that $a = b$. □

Remark 5.3. The reader familiar with Aumann (1976) will probably have noticed that the proof of the first agreement theorem in probabilistic dynamic epistemic logic is a straightforward adaptation of Aumann's own proof for his original agreement theorem (but incorporating already the experimentation dynamics, whereas Aumann's theorem is fully static; see Subsection 5.6.1). This shows that probabilistic Kripke models are a natural setting in which to formalize (dynamic) agreement theorems.

I will now comment on the intuitive interpretation of this theorem and on the two assumptions required to prove it. The theorem is essentially a sentence of the formal language $\mathcal{L}([\text{EXP}])$, and says that if after carrying out the experiments, the agents reach common knowledge about their posteriors for φ, then these posteriors have to be identical. Intuitively, this is very close to Aumann's original agreement theorem, but with the experimentation dynamics explicitly represented in the language. Note, however, that this theorem talks about what will be the case *if* the agents reach common knowledge for their posterior about φ, without saying anything about *how* such common knowledge is to be achieved.

The two conditions required to prove the agreement theorem are fairly weak. Condition 1 ($\mu_1(w) = \mu_2(w)$) is an immediate formalization of Aumann's 'common prior' assumption, but localized to the concrete state w. Condition 2 ($\mu_i(w) = \mu_i(v)$ for all $v \in R^*[w]$) is a weakened version of an assumption that is also implicit in Aumann's original setup: Aumann works with structures which have just *one* probability mass function, i.e. he assumes that $\mu_i(x) = \mu_i(y)$ for all states $x, y \in W$. Theorem 5.1 shows that this assumption can be weakened: the local version ($\mu_i(x) = \mu_i(w)$ for all $x \in R^*[w]$) suffices. In Subsection 5.5.2, I will show that under the common prior assumption, common knowledge is not needed to characterize this property: individual knowledge suffices.

It should be noted that Theorem 5.1 is a *local* theorem (about a particular state w) and a theorem about probabilistic Kripke *models*. However, in the proof we nowhere made any use of the concrete valuation. Furthermore, the reference to the concrete state w can also be eliminated by 'de-localizing' the theorem's two assumptions. In this way, we arrive at the following *global frame version* of the first agreement theorem:

Theorem 5.2. *Let* $\mathbb{F} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2 \rangle$ *be an arbitrary probabilistic Kripke frame. Suppose that the following conditions hold:*

1. $\mu_1 = \mu_2$,
2. *for all* $w, v \in W$: *if* $(w, v) \in R^*$, *then* $\mu_i(w) = \mu_i(v)$.

Then we have:

$$\mathbb{F} \models [\text{EXP}]\, C(P_1(\varphi) = a \wedge P_2(\varphi) = b) \rightarrow a = b.$$

Proof. Let $V\colon \mathsf{Prop} \rightarrow \wp(W)$ be an arbitrary valuation on $\mathbb{F}$, and let $w \in W$ be an arbitrary state. Since the conditions of this theorem are simply the 'de-localized' versions of the conditions of Theorem 5.1, it follows immediately by that theorem that $\langle \mathbb{F}, V \rangle, w \models [\text{EXP}]\, C(P_1(\varphi) = a \wedge P_2(\varphi) = b) \rightarrow a = b$. □

5.4.2 Experimentation and Communication

I now turn to the second agreement theorem in probabilistic dynamic epistemic logic, which also explicitly represents the communication dynamics (in contrast with the first agreement theorem).

First, however, one more auxiliary lemma is needed. Intuitively, this lemma says that after a dialogue about φ, the agents' probabilities for φ become common knowledge.

Lemma 5.4. *Let* $\mathbb{M} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V \rangle$ *be an arbitrary probabilistic Kripke model and assume that* $w \in W$. *Then*

$$\mathbb{M}, w \models [\text{DIAL}(\varphi)]\big((P_1(\varphi) = a \wedge P_2(\varphi) = b) \rightarrow C(P_1(\varphi) = a \wedge P_2(\varphi) = b)\big).$$

Proof. Suppose that $\mathbb{M}^{dial_w(\varphi)}, w \models P_1(\varphi) = a \wedge P_2(\varphi) = b$. Hence

$$\delta := d(\mathbb{M}^{dial_w(\varphi)}, w, \varphi) = \big(P_1(\varphi) = a \wedge P_2(\varphi) = b\big).$$

Let n be the least number such that $f^n_{w,\varphi}(\mathbb{M}) = f^{n+1}_{w,\varphi}(\mathbb{M})$ (such a number is guaranteed to exist by Lemma 5.1). Note that $\mathbb{M}^{dial_w(\varphi)} = f^n_{w,\varphi}(\mathbb{M}) = f^{n+1}_{w,\varphi}(\mathbb{M}) = f_{w,\varphi}(f^n_{w,\varphi}(\mathbb{M})) = f_{w,\varphi}(\mathbb{M}^{dial_w(\varphi)})$, so δ is true in *all* states of $\mathbb{M}^{dial_w(\varphi)}$. From this it follows trivially that $\mathbb{M}^{dial_w(\varphi)}, w \models C\delta$, as required. □

This brings us to the second agreement theorem in probabilistic dynamic epistemic logic, which explicitly represents both the experimentation and the communication dynamics:

Theorem 5.3. *Let* $\mathbb{M} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V\rangle$ *be an arbitrary probabilistic Kripke model and* $w \in W$ *a state of* $\mathbb{M}$*. Suppose that the following conditions hold:*

1. $\mu_1(w) = \mu_2(w)$*,*
2. *for all* $v \in R^*[w]\colon \mu_i(w) = \mu_i(v)$*.*

Then we have:

$$\mathbb{M}, w \models [\text{EXP}]\,[\text{DIAL}(\varphi)](P_1(\varphi) = a \wedge P_2(\varphi) = b) \rightarrow a = b.$$

Proof. This proof is structurally completely analogous to that of Theorem 5.1. Making use of Lemma 5.4, we show that

$$\frac{\mu_1(w)\big(\,[\![\,\varphi\,]\!]^{(\mathbb{M}^e)^{dial_w(\varphi)}} \cap \mathcal{R}^*[w]\big)}{\mu_1(w)\big(\mathcal{R}^*[w]\big)} = a \tag{5.5}$$

and that

$$\frac{\mu_2(w)\big(\,[\![\,\varphi\,]\!]^{(\mathbb{M}^e)^{dial_w(\varphi)}} \cap \mathcal{R}^*[w]\big)}{\mu_2(w)\big(\mathcal{R}^*[w]\big)} = b, \tag{5.6}$$

where $\mathcal{R}^*$ is the reflexive transitive closure of $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2$, and $\mathcal{R}_i$ is agent i's epistemic indistinguishability relation in the model $(\mathbb{M}^e)^{dial_w(\varphi)}$. Statements (5.5) and (5.6), together with condition 1 of this theorem, entail that $a = b$.[13] □

The theorem says that after the agents have carried out the experiments, and then carried out a dialogue about φ, their posteriors for φ have to be identical. Intuitively, this is very close to Aumann's original agreement theorem, except that the experimentation and communication dynamics are now explicitly represented in the language.

We again obtain a *global frame version* of the agreement theorem by 'delocalizing' the assumptions:

[13] For a more detailed proof, see Demey (2010, Theorem 38).

Theorem 5.4. *Let $\mathbb{F} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2 \rangle$ be an arbitrary probabilistic Kripke frame. Suppose that the following conditions hold:*

1. $\mu_1 = \mu_2$,
2. *for all $w, v \in W$: if $(w, v) \in R^*$, then $\mu_i(w) = \mu_i(v)$.*

Then we have:

$$\mathbb{F} \models [\mathrm{EXP}]\,[\mathrm{DIAL}(\varphi)](P_1(\varphi) = a \wedge P_2(\varphi) = b) \to a = b.$$

Proof. Let $V\colon \mathsf{Prop} \to \wp(W)$ be an arbitrary valuation on $\mathbb{F}$, and let $w \in W$ be an arbitrary state. Since the conditions of this theorem are simply the 'delocalized' versions of the conditions of Theorem 5.3, it follows by that theorem that $\langle \mathbb{F}, V \rangle, w \models [\mathrm{EXP}]\,[\mathrm{DIAL}(\varphi)](P_1(\varphi) = a \wedge P_2(\varphi) = b) \to a = b$. □

Remark 5.4. The first agreement theorem (Theorems 5.1 and 5.2) states that if the agents have common knowledge of their posteriors, then these posteriors have to be identical. However, it says nothing about how this common knowledge is to be achieved, i.e. it did not say anything about the communication. The second agreement theorem (Theorems 5.3 and 5.4), however, *does* explicitly represent the communication dynamics, and thus no longer needs the assumption of common knowledge: the existence of common knowledge can now be *derived* from the communication protocol (Lemma 5.4).

5.5 Metatheory

I will now develop a sound and complete logic in which the agreement theorem can be derived. Subsection 5.5.1 discusses a technical difficulty related to the syntactic perspective on probabilistic epistemic logic in general, and proposes a solution to it. Subsection 5.5.2 provides characterization results for the conditions of the agreement theorems proved in Section 5.4. These characterization results are then used in Subsection 5.5.3 to obtain various axiomatizations.

5.5.1 A Difficulty about Expressivity

The modeling of the experiments has so far been very general: agent i's experiment corresponds to any equivalence relation E_i (or, equivalently, to any partition of the model) whatsoever. From the syntactic perspective, however, this full generality is difficult to maintain, because it exceeds the expressive powers of the formal language $\mathcal{L}([\text{EXP}])$. I will first give a concrete illustration of this problem and then propose a solution to it.

Recall the semantics for i-probability formulas such as $P_i(\varphi) \geq k$:

$$\mathbb{M}, w \models P_i(\varphi) \geq k \quad \text{iff} \quad \mu_i(w)([\![\, \varphi \,]\!]^{\mathbb{M}}) \geq k.$$

There is a clear asymmetry in expressivity between both sides of this definition. On the left hand side, there is a formula of the formal language $\mathcal{L}([\text{EXP}])$. The Backus-Naur form of this language guarantees that $P_i(\cdot)$ will always receive a formula as its argument. On the right hand side, however, we have the function $\mu_i(w)(\cdot)$, which can receive any set $X \subseteq W$ whatsoever as its argument, even *undefinable* sets (i.e. sets X such that $X = [\![\, \varphi \,]\!]^{\mathbb{M}}$ for no $\mathcal{L}([\text{EXP}])$-formula φ). It may well be the case that $E_i[w]$ is an undefinable set. In that case, several problems of expressivity will arise; for example, the [EXP]-reduction axiom for i-probability formulas will in general not be expressible in $\mathcal{L}([\text{EXP}])$.[14]

To solve the problem, it should thus be ensured that $E_i[w]$ is always definable by means of some formula. One way to ensure this is by restricting to *binary experiments*.[15] The first, *syntactic* step of this strategy is to introduce two new primitive formulas α_1, α_2 into the language. The second, *semantic* step involves assuming that for any probabilistic Kripke frame $\mathbb{F} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2 \rangle$ there exist sets $\mathcal{E}_i^{\mathbb{F}} \subseteq W$ such that $E_i = \big(\mathcal{E}_i^{\mathbb{F}} \times \mathcal{E}_i^{\mathbb{F}}\big) \cup \big((W - \mathcal{E}_i^{\mathbb{F}}) \times (W - \mathcal{E}_i^{\mathbb{F}})\big)$. The third and final step links syntax and semantics, by extending the valuations to the newly introduced α_i's: for any valuation

[14] Here's another way of putting the problem. Both experimentation and public announcement of a formula φ change the probabilistic component of a Kripke model via Bayesian conditionalization: $\mu_i^e(w)(v) = \mu_i(w)(v \mid E_i[w])$ and $\mu_i^{\varphi}(w)(v) = \mu_i(w)(v \mid [\![\, \varphi \,]\!]^{\mathbb{M}})$. In the case of public announcement, this fact can also be expressed in the object language (see Subsection 3.3.3):

$$\varphi \longrightarrow \big([!\varphi]P_i(\psi) = k \leftrightarrow P_i([!\varphi]\psi \mid \varphi) = k\big).$$

In the case of experimentation, however, the fact that the agents' probabilities get updated by means of Bayesian conditionalization cannot be expressed in the object language (because the set $E_i[w]$ might be undefinable).

[15] From a technical perspective, this solution is analogous to the construction of general frames in modal logic (Blackburn et al. 2001, Definition 1.32); also see Footnote 19 on p. 65. An entirely different solution, based on hybrid logic, is explored in detail in Demey (2010). There it is also argued that the 'binary experiments'-solution is preferable on technical as well as methodological grounds.

V on $\mathbb{F}$, we require that $V(\alpha_i) \in \{\mathcal{E}_i^{\mathbb{F}}, W - \mathcal{E}_i^{\mathbb{F}}\}$, and thus obtain:

$$E_i = \big(V(\alpha_i) \times V(\alpha_i)\big) \cup \big((W - V(\alpha_i)) \times (W - V(\alpha_i))\big). \qquad (5.7)$$

It is easy to check that E_i, thus defined, is still an equivalence relation, and that this new definition is 'compatible' with the main types of dynamics discussed in this chapter, in the sense that if a probabilistic Kripke model $\mathbb{M}$ satisfies condition (5.7), then the updated models $\mathbb{M}^e$ and $\mathbb{M}^\varphi$ will satisfy it as well.

Informally, (5.7) says that agent i's experiment only differentiates between α_i-states and $\neg\alpha_i$-states; in other words, it is a 'binary experiment'. Continuing the analogy between experiments and questions, carrying out a binary experiment corresponds to asking a *yes-no question*: 'is α_i the case?'.

In this more restricted setup, it follows easily from condition (5.7) that $E_i[w] = [\![\, \alpha_i \,]\!]^{\mathbb{M}}$ if $\mathbb{M}, w \models \alpha_i$, and $E_i[w] = [\![\, \neg\alpha_i \,]\!]^{\mathbb{M}}$ otherwise. Hence $E_i[w]$ is now always definable: either by α_i or by $\neg\alpha_i$ (depending on whether $\mathbb{M}, w \models \alpha_i$). This definability result will be used extensively in Subsection 5.5.3 (in the [EXP]-reduction axiom for i-probability formulas, but also in other axioms).

5.5.2 Characterization Results

In Section 5.4, I established various dynamic agreement theorems. These theorems required imposing two conditions on probabilistic Kripke models/frames. I will now establish characterization results for (the global frame versions of) these conditions.

I first characterize the common prior assumption, i.e. condition 1 of Theorems 5.2 and 5.4. If φ is a 1-probability formula, I will use $\varphi[P_2/P_1]$ to denote the formula that is obtained by uniformly substituting P_2 for P_1 in φ. It is clear that if φ is a 1-probability formula, then $\varphi[P_2/P_1]$ is a 2-probability formula. Finally, recall that an i-probability formula φ is said to be *atomic* iff it is of the form $\sum_{\ell=1}^{n} a_\ell P_i(p_\ell) \geq k$, i.e. iff the arguments of its probability operators are propositional atoms (rather than arbitrary formulas).

Lemma 5.5. *Let* $\mathbb{F} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2 \rangle$ *be an arbitrary probabilistic Kripke frame. Then we have:*

$$\mu_1 = \mu_2 \quad \textit{iff}$$
$$\textit{for all atomic 1-probability formulas } \varphi\colon\ \mathbb{F} \models \varphi \leftrightarrow \varphi[P_2/P_1].$$

Proof. The left-to-right entailment follows immediately from the semantics; the right-to-left entailment will be proved by contraposition. Suppose that $\mu_1 \neq \mu_2$. Hence there exist states $w, v \in W$ such that $\mu_1(w)(v) \neq \mu_2(w)(v)$. Without loss of generality, assume that $\mu_1(w)(v) < \mu_2(w)(v)$. Since $\mathbb{Q}$ is dense in $\mathbb{R}$, there exists a $k \in \mathbb{Q}$ such that $\mu_1(w)(v) < k < \mu_2(w)(v)$. Now define a valuation $V\colon \mathsf{Prop} \to \wp(W)$ by putting $V(p) := \{v\}$. It follows from this that $\langle \mathbb{F}, V\rangle, w \models P_2(p) \geq k$ and that $\langle \mathbb{F}, V\rangle, w \not\models P_1(p) \geq k$, and hence $\mathbb{F} \not\models P_1(p) \geq k \leftrightarrow P_2(p) \geq k$. □

Remark 5.5. Lemma 5.5 clearly involves a (countably) infinite set of formulas (the same holds for the second characterization result, stated in Lemmas 5.6 and 5.7). Halpern (2002) provides another characterization of the common prior assumption, which also involves a (countably) infinite set of formulas (viz. all instances of a single scheme). Halpern's characterization involves formulas which are strongly related to the agreeing to disagree theorem, and is thus not suitable for our current purposes: in the next subsection, the characterization results established here will be used to provide a complete axiomatization of a logic in which the agreeing to disagree result is formally derivable; if the logic's axiomatization would itself already include (something very close to) the agreement theorem, then this derivation would be trivial. Additionally, the formulas used in Lemma 5.5 seem to be the most straightforward way of formally expressing the common prior assumption: agents 1 and 2 having a common prior means exactly that $P_1(p) \geq 0.5 \leftrightarrow P_2(p) \geq 0.5$, $P_1(p) \leq 0.5 \leftrightarrow P_2(p) \leq 0.5$, etc.

Condition 2 of Theorems 5.2 and 5.4 can be characterized as follows:

Lemma 5.6. *Let $\mathbb{F} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2\rangle$ be an arbitrary probabilistic Kripke frame. Then for $i = 1, 2$ we have:*

$$\textit{for all } w, v \in W\colon \textit{ if } (w, v) \in R^*, \textit{ then } \mu_i(w) = \mu_i(v) \quad \textit{iff}$$
$$\textit{for all atomic } i\textit{-probability formulas } \varphi\colon \mathbb{F} \models \big(\varphi \to C\varphi\big) \wedge \big(\neg\varphi \to C\neg\varphi\big).$$

Proof. The left-to-right entailment follows immediately from the semantics; the right-to-left entailment will be proved by contraposition. Suppose that there exist states $w, v \in W$ such that $(w, v) \in R^*$ and yet $\mu_i(w) \neq \mu_i(v)$. Hence there exists a state $x \in W$ such that $\mu_i(w)(x) \neq \mu_i(v)(x)$. Now define a valuation $V\colon \mathsf{Prop} \to \wp(W)$ by putting $V(p) := \{x\}$. Since $\mu_i(w)(x) \neq \mu_i(v)(x)$, one of the following two cases obtains:

1. $\mu_i(w)(x) > \mu_i(v)(x)$. Since $\mathbb{Q}$ is dense in $\mathbb{R}$, there exists a $k \in \mathbb{Q}$ such that $\mu_i(w)(x) > k > \mu_i(v)(x)$. It follows that $\langle \mathbb{F}, V\rangle, w \models P_i(p) \geq k$ and that $\langle \mathbb{F}, V\rangle, v \not\models P_i(p) \geq k$, and hence $\langle \mathbb{F}, V\rangle, w \not\models C(P_i(p) \geq k)$. Hence $\mathbb{F} \not\models P_i(p) \geq k \rightarrow C(P_i(p) \geq k)$.

2. $\mu_i(w)(x) < \mu_i(v)(x)$. Completely analogously, it follows that there exists a $k \in \mathbb{Q}$ such that $\mathbb{F} \not\models \neg(P_i(p) \geq k) \rightarrow C\neg(P_i(p) \geq k)$. □

Remark 5.6. The condition that $\mu_i(w) = \mu_i(v)$ whenever $(w, v) \in R^*$ is a very heavy constraint to impose on probabilistic Kripke frames: it involves the reflexive transitive closure of R, and might therefore be called 'semi-global'. This aspect is reflected in the characterization result above, which makes use of the common knowledge operator C. However, because frame validity is itself a global notion, it is possible to capture the semi-global frame property involving R^* by means of the more modest general knowledge operator E. This result is still not fully satisfactory, however: the principles that $\varphi \rightarrow E\varphi$ and $\neg\varphi \rightarrow E\neg\varphi$ (for atomic i-probability formulas φ) still require the 'public availability' of agent i's subjective probabilistic setup. However, in frames satisfying the common prior assumption ($\mu_1 = \mu_2$)—and note that all frames used to prove the agreement results indeed satisfy this property!—more plausible 'individual' introspection principles suffice: $\varphi \rightarrow K_i\varphi$ and $\neg\varphi \rightarrow K_i\neg\varphi$ (for atomic i-probability formulas φ). Hence, no notion of social (common/general) knowledge is required to characterize the second assumption of the agreement theorems.

Lemma 5.7. *Let* $\mathbb{F} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2\rangle$ *be an arbitrary probabilistic Kripke frame and suppose that* $\mu_1 = \mu_2$. *Then we have:*

for $i = 1, 2$ *and for all* $w, v \in W$: *if* $(w, v) \in R^*$, *then* $\mu_i(w) = \mu_i(v)$ *iff*
for $i = 1, 2$ *and for all atomic* i*-probability formulas* φ:
$$\mathbb{F} \models \big(\varphi \rightarrow K_i\varphi\big) \wedge \big(\neg\varphi \rightarrow K_i\neg\varphi\big).$$

Proof. Again, the left-to-right entailment follows immediately from the semantics; the right-to-left entailment will be proved directly this time (i.e. not by contraposition). Assume that $\mathbb{F} \models (\varphi \rightarrow E\varphi) \wedge (\neg\varphi \rightarrow E\neg\varphi)$ for all atomic i-probability formulas φ, and call this assumption (†). We now prove for all states $w, v \in W$ that if $(w, v) \in R^*$ then $\mu_i(w) = \mu_i(v)$. Since $R^* = \bigcup_{n\geq 0} R^n$, it suffices to show that for all $n \geq 0$, for all $w, v \in W$: if $(w, v) \in R^n$ then $\mu_i(w) = \mu_i(v)$. This is proved by induction on n. The base case is trivial. For the induction case, consider arbitrary $w, v \in W$ and assume that $(w, v) \in R^{n+1}$. Hence there is a state $u \in W$ such that $(w, u) \in R^n$ and $(u, v) \in R$. Since $(w, u) \in R^n$ it follows by the induction hypothesis that $\mu_i(w) = \mu_i(u)$; I claim that $\mu_i(u) = \mu_i(v)$ as well, and hence it follows that $\mu_i(w) = \mu_i(v)$.

Proof of the claim that $\mu_i(u) = \mu_i(v)$. For a reductio, suppose that $\mu_i(u) \neq \mu_i(v)$. Hence there is a state $x \in W$ such that $\mu_i(u)(x) \neq \mu_i(v)(x)$. Now define a valuation $V\colon \mathit{Prop} \to \wp(W)$ by putting $V(p) := \{x\}$. Since $\mu_i(u)(x) \neq \mu_i(v)(x)$, one of the following two cases obtains:

1. $\mu_i(u)(x) > \mu_i(v)(x)$. Since $\mathbb{Q}$ is dense in $\mathbb{R}$, there exists a $k \in \mathbb{Q}$ such that $\mu_i(u)(x) > k > \mu_i(v)(x)$. It follows that $\langle \mathbb{F}, V\rangle, u \models P_i(p) \geq k$ and that $\langle \mathbb{F}, V\rangle, v \not\models P_i(p) \geq k$, and hence $\langle \mathbb{F}, V\rangle, u \not\models E(P_i(p) \geq k)$. Hence $\mathbb{F} \not\models P_i(p) \geq k \to E(P_i(p) \geq k)$, which contradicts assumption (†).

2. $\mu_i(u)(x) < \mu_i(v)(x)$. Completely analogously, it follows that there exists a $k \in \mathbb{Q}$ such that $\mathbb{F} \not\models \neg(P_i(p) \geq k) \to E\neg(P_i(p) \geq k)$, which again contradicts assumption (†). □

Remark 5.7. Lemma 5.7 is highly similar to the correspondence result for consistency, as stated by Lemma 3.1 on p. 79. Although the correspondence formulas used in both lemmas are the same, the frame properties are quite different: Lemma 5.7 is concerned with a semi-global property (states that are R^*-related), while Lemma 3.1 is concerned with a local property (states that are R_i-related). The lemmas also have different quantificational patterns: Lemma 3.1 establishes the correspondence 'agent per agent', while Lemma 5.7 only holds 'for all agents simultaneously'. Formally, this means that Lemma 3.1 is of the form

$$\forall i \in I\colon \big(\mathit{LHS}(i) \Leftrightarrow \mathit{RHS}(i)\big)$$

and Lemma 5.7 is of the form

$$\big(\forall i \in I\colon \mathit{LHS}(i)\big) \Leftrightarrow \big(\forall i \in I\colon \mathit{RHS}(i)\big).$$

5.5.3 The Logics

I will now define three logics of increasing strength, and prove them to be sound and complete with respect to natural classes of Kripke frames. The second and, especially, the third logic capture the reasoning behind the agreement theorem. For the sake of clarity, these logics are presented in a modular fashion.

Figure 5.1: Componentwise axiomatization of EPEL

1. the propositional component
2. the individual knowledge component
3. the common knowledge component
4. the linear inequalities component
5. the probabilistic component
6. the well-behavedness component
7. the pre-/post-experimental interaction component
8. the α_i-component

The first logic is the *enriched probabilistic epistemic logic* EPEL, which captures the behavior of the epistemic and probabilistic operators. It does not say anything about agreement theorems. Figure 5.1 provides a schematic overview of the logic. I will now discuss each of its components separately.

The propositional, probabilistic and linear inequalities components are exactly as in the axiomatization of basic probabilistic epistemic logic discussed in Chapter 3 (see Figure 3.1 on p. 82), and thus need no further comments. The individual knowledge component says that the individual (pre- and post-experimental) knowledge operators K_i and R_i are S5-modal operators. (Note that Aumann's original result also involved S5-type knowledge.) Similarly, the common knowledge component governs the behavior of (pre- and post-experimental) relativized common knowledge; it consists of the following rules and axioms (van Benthem et al. 2006):

$$\begin{array}{ll}
\text{if} \vdash \psi \text{ then} \vdash C^{\varphi}\psi, & \text{if} \vdash \psi \text{ then} \vdash X^{\varphi}\psi, \\
C^{\varphi}(\psi \to \chi) \to (C^{\varphi}\psi \to C^{\varphi}\chi), & X^{\varphi}(\psi \to \chi) \to (X^{\varphi}\psi \to X^{\varphi}\chi), \\
C^{\varphi}\psi \leftrightarrow E\big(\varphi \to (\psi \wedge C^{\varphi}\psi)\big), & X^{\varphi}\psi \leftrightarrow F\big(\varphi \to (\psi \wedge X^{\varphi}\psi)\big), \\
C^{\varphi}\big(\psi \to E(\varphi \to \psi)\big) \to & X^{\varphi}\big(\psi \to F(\varphi \to \psi)\big) \to \\
\big(E(\varphi \to \psi) \to C^{\varphi}\psi\big), & \big(F(\varphi \to \psi) \to X^{\varphi}\psi\big).
\end{array}$$

The well-behavedness component indicates that the models on which the logic is interpreted are well-behaved; it consists simply of the formulas that correspond to the consistency and liveness properties, which jointly define well-behavedness (see Lemma 3.1 on p. 79):

$$\varphi \to P_i(\varphi) > 0, \qquad K_i\varphi \to P_i(\varphi) = 1.$$

Next, the pre-/post-experimental interaction component describes the influence of the experiments on the agents' (common) knowledge: it says that carrying out the experiments does not make the agents forget anything that they already (commonly) knew before the experiments (principles such as these are sometimes called *perfect recall* principles). Formally:

$$K_i\varphi \to R_i\varphi, \qquad C^{\varphi}\psi \to X^{\varphi}\psi.$$

The final component of EPEL involves the special proposition letters α_i. First of all, there is an axiom which says that the post-experimental knowledge operator R_i can be defined in terms of the usual knowledge operator K_i and these special proposition letters:[16]

$$R_i\varphi \leftrightarrow \Big(\big(\alpha_i \to K_i(\alpha_i \to \varphi)\big) \wedge \big(\neg\alpha_i \to K_i(\neg\alpha_i \to \varphi)\big)\Big).$$

Finally, this component also contains axioms which say that the agents' experiments are *successful*: if α_i is the case, then after carrying out her experiment, agent i will know this; likewise if α_i is not the case. Formally:

$$\alpha_i \to R_i\alpha_i, \qquad \neg\alpha_i \to R_i\neg\alpha_i.$$

This concludes the presentation of enriched probabilistic epistemic logic (EPEL). We now turn to the second logic, viz. *probabilistic epistemic agreement logic* or PEAL. The componentwise axiomatization of PEAL can be found in Figure 5.2. This logic is a simple extension of EPEL: we just add an 'agreement component', which consists of the formulas that characterize the two frame properties needed in the agreement theorems (cf. Lemmas 5.5–5.7).

[16] Given this definability result, it might be asked why R_i is still introduced as a *primitive* operator. The reason for doing this is that this operator is only definable *if* we make use of the special proposition letters α_i; it should be emphasized that these were only introduced at the beginning of Section 5.5, when we shifted from a semantic to a syntactic pespective.

Figure 5.2: Componentwise axiomatization of PEAL

1–8.	the eight components of EPEL	
9.	the agreement component:	
	$\varphi \leftrightarrow \varphi[P_2/P_1]$	(for 1-probability formulas φ)
	$\varphi \rightarrow K_i\varphi$ and $\neg\varphi \rightarrow K_i\neg\varphi$	(for i-probability formulas φ)

Figure 5.3: Componentwise axiomatization of DPEALe

1–9.	the nine components of PEAL
10.	the reduction axioms for [EXP]

I now introduce the third and final logic, viz. *dynamic probabilistic epistemic agreement logic with explicit experimentation* or DPEALe. As can be seen in Figure 5.3, this logic is obtained by simply adding the [EXP]-reduction axioms to PEAL. These reduction axioms are displayed below. Most of them are straightforward; I only emphasize the use of R_i to pre-encode the effects of the experimentation dynamics on K_i (similar remarks apply to common knowledge), and the use of α_i in the reduction axiom for i-probability formulas to avoid non-expressibility (recall Subsection 5.5.1).

$$\begin{array}{lrcll}
1. & [\text{EXP}]\, p & \leftrightarrow & p, & (\text{for } p \in \mathsf{Prop} \cup \{\alpha_1, \alpha_2\}) \\
2. & [\text{EXP}]\, \neg\varphi & \leftrightarrow & \neg\, [\text{EXP}]\, \varphi, & \\
3. & [\text{EXP}](\varphi \wedge \psi) & \leftrightarrow & [\text{EXP}]\, \varphi \wedge [\text{EXP}]\, \psi, & \\
4. & [\text{EXP}]\, K_i\varphi & \leftrightarrow & R_i\, [\text{EXP}]\, \varphi, & \\
5. & [\text{EXP}]\, R_i\varphi & \leftrightarrow & R_i\, [\text{EXP}]\, \varphi, & \\
6. & [\text{EXP}]\, C^{\varphi}\psi & \leftrightarrow & X^{[\text{EXP}]\,\varphi}\, [\text{EXP}]\, \psi, & \\
7. & [\text{EXP}]\, X^{\varphi}\psi & \leftrightarrow & X^{[\text{EXP}]\,\varphi}\, [\text{EXP}]\, \psi, & \\
8. & [\text{EXP}] \sum_\ell a_\ell P_i(\varphi_\ell) \geq k & \leftrightarrow & &
\end{array}$$

$$\begin{cases} \quad \alpha_i \rightarrow \sum_\ell a_\ell P_i([\text{EXP}]\, \varphi_\ell \wedge \alpha_i) \geq k P_i(\alpha_i) \\ \wedge \quad \neg\alpha_i \rightarrow \sum_\ell a_\ell P_i([\text{EXP}]\, \varphi_\ell \wedge \neg\alpha_i) \geq k P_i(\neg\alpha_i). \end{cases}$$

With the logics in place, I now turn to their soundness and completeness. First, consider the classes of frames with respect to which soundness and completeness results will be proved:

Definition 5.6. The class of enriched probabilistic Kripke frames with binary experiments—i.e. satisfying condition (5.7)—will be denoted $\mathcal{PKB}$.

Definition 5.7. Consider an arbitrary frame $\mathbb{F} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2 \rangle \in \mathcal{PKB}$. Then $\mathbb{F}$ is said to be an *agreement frame* iff it satisfies conditions 1 and 2 of Theorems 5.2 and 5.4. The class of agreement frames will be denoted $\mathcal{AGR}$.

Remark 5.8. We immediately obtain:

1. $\mathcal{AGR} \models [\text{EXP}]\, C(P_1(\varphi) = a \wedge P_2(\varphi) = b) \to a = b.$
2. $\mathcal{AGR} \models [\text{EXP}]\, [\text{DIAL}(\varphi)](P_1(\varphi) = a \wedge P_2(\varphi) = b) \to a = b.$

Theorem 5.5. *We have the following soundness and completeness results:*

1. *The logic* EPEL *is sound and complete with respect to* $\mathcal{PKB}$*.*
2. *The logic* PEAL *is sound and complete with respect to* $\mathcal{AGR}$*.*
3. *The logic* DPEALe *is sound and complete with respect to* $\mathcal{AGR}$*.*

Proof. Soundness of each of the three logics is proved by induction on derivation length, as is usual. The completeness proofs of the first two logics involve standard techniques in modal logic, such as a canonical model construction, filtration, etc. (Blackburn et al. 2001). The completeness proof of the third logic relies on standard techniques in dynamic epistemic logic. In particular, the completeness of the dynamic logic DPEALe is 'reduced' to the completeness of its static base logic PEAL via the use of reduction axioms (van Ditmarsch et al. 2007). For full details, see Theorems 56, 57 and 58 of Demey (2010). □

Corollary 5.1. *The logics* EPEL*,* PEAL *and* DPEALe *all have the finite model property.*

Proof. This follows immediately from the completeness results established above, because enriched probabilistic Kripke frames (with binary experiments) are, by definition, finite (recall Definition 5.1). □

Remark 5.9. Combining Theorem 5.5 with Remark 5.8, it immediately follows that DPEALe $\vdash$ $[\text{EXP}]\, C(P_1(\varphi) = a \wedge P_2(\varphi) = b) \to a = b$. The system DPEALe is thus strong enough to derive a dynamic agreement theorem which explicitly represents the experimentation dynamics.

5.6 Agreeing to Disagree and the Dynamic Turn in Epistemic Logic

In this section, I will examine some of the methodological and philosophical implications of the technical results established earlier. Subsection 5.6.1 discusses the importance of explicitly representing the dynamics behind the agreement theorem. Subsection 5.6.2 examines the implications of this for the role and importance of common knowledge in agreement results. Together, these two subsections illustrate how the dynamic turn in logic can be applied to Aumann's agreement theorem.

5.6.1 Static versus Dynamic Agreement Theorems

There is an important dynamic aspect to Aumann's agreement theorem (recall the intuitive scenario described in Section 5.2). However, Aumann does not seem to make this dynamic aspect sufficiently explicit. On the *syntactic* side, his formulation of the theorem is a conditional statement without any dynamic operators; recall (5.1) from Section 5.2. I will now argue that the underlying dynamics is not adequately captured by his *semantic* setup either.

In the approach developed here, we have two 'temporally uniform' models: the model $\mathbb{M} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V\rangle$ represents the agents' knowledge and probabilities *before* the experiments have been carried out, and the model $\mathbb{M}^e = \langle W^e, R_1^e, R_2^e, E_1^e, E_2^e, \mu_1^e, \mu_2^e, V^e\rangle$ represents the agents' knowledge and probabilities *after* the experiments have been carried out. Now contrast this with Aumann's original models. These seem to be 'temporally incoherent': they represent the agents' knowledge *after* the experiments, but their probability distributions *before* the experiments. In the present framework, such a model would look as follows: $\langle W, R_1^e, R_2^e, E_1, E_2, \mu_1, \mu_2, V\rangle$; it is obtained by cutting the (temporally uniform) models $\mathbb{M}$ and $\mathbb{M}^e$ into pieces, and then pasting these pieces back together in a 'temporally incoherent' way.

The situation can be analyzed as follows. The intuitive agreeing to disagree scenario described in Section 5.2 is *intrinsically dynamic*. If one formulates the agreement theorem in a static way (like Aumann did), then one will need to 'smuggle' this dynamics into the semantics somehow, thus obtaining 'temporally incoherent' models.

The approach developed here, however, makes the underlying dynamics fully explicit. On the *semantic* side, we have a probabilistic Kripke model $\mathbb{M}$ which corresponds to the initial stage (before the experiments), a model $\mathbb{M}^e$ which corresponds to the time immediately after the experiments, and finally, a model $(\mathbb{M}^e)^{dial_w(\varphi)}$ which corresponds to the final stage after the communication, at which the agents have reached common knowledge of their posteriors. Hence, there exists a complete structural analogy between the intuitive scenario on the one hand and its model-theoretical formalization on the other. On the *syntactic* side, the agreement theorems proved here (in particular, Theorems 5.3 and 5.4) are formulated using the dynamic $[\text{EXP}]$- and [DIAL(φ)]-operators, and are thus able to talk about this entire sequence of models $\mathbb{M} \mapsto \mathbb{M}^e \mapsto (\mathbb{M}^e)^{dial_w(\varphi)}$. Hence, they can be read as natural and explicit descriptions of the intuitive scenario that was behind the original agreement theorem.

To summarize: the agreement theorems developed in this chapter perfectly illustrate the *dynamic turn* in logic (recall the quote from van Benthem given in Section 5.2). In the next subsection, I will show that this dynamic turn offers a new perspective on the conceptual landscape surrounding the agreement theorem, and, in particular, on the role of common knowledge.

5.6.2 The Role of Common Knowledge

In order to formulate and prove his agreement theorem, Aumann used the notion of *common knowledge*, thus being the first author to introduce this notion in the game-theoretical literature. Therefore, it is widely assumed that common knowledge plays a central role in agreeing to disagree results. Several results established throughout this chapter, however, seem to suggest that the importance of common knowledge is not so central as is often thought.

First of all, in Aumann's original setup, the (common) prior probability distribution is assumed to be common knowledge among the agents. This is reflected in the present framework by the characterization result involving $\varphi \rightarrow C\varphi$ (and $\neg\varphi \rightarrow C\neg\varphi$) for i-probability formulas φ. However, I showed that this can be replaced with the much weaker individual probabilistic-epistemic introspection principle $\varphi \rightarrow K_i\varphi$ (and $\neg\varphi \rightarrow K_i\neg\varphi$) for i-probability formulas φ (see Remark 5.6). In other words, the assumption that the agents' prior probability distributions are common knowledge can be formally captured without making use of the common knowledge operator.

A second, more important observation concerns the role of common knowledge in obtaining consensus (i.e. identical posterior probabilities). Aumann's original theorem says that if after carrying out the experiments, the agents have common knowledge of their posteriors, then these posteriors have to be identical. However, this theorem does not say *how* the agents are to obtain this common knowledge (it just assumes that they have been able to obtain it one way or another). The way to obtain common knowledge is via a certain communication protocol. Once this communication dynamics is made explicitly part of the story (as suggested by the dynamic turn—recall the previous subsection), common knowledge of the posteriors need no longer be *assumed* in the formulation of the agreement theorem (see Remark 5.4), since it will now simply *follow* from the communication protocol (see Lemma 5.4).

Finally, note that these comments on the relative unimportance of common knowledge for agreeing to disagree results are in line with the results by Parikh and Krasucki (1990). They consider groups of more than two agents, in which communication does not occur publicly, but in pairs. They show that, given certain conditions on the communication protocol, the agents will reach a *consensus*—i.e. identical posteriors—, but not *common knowledge* of these posteriors.

5.7 Conclusion

In this chapter I have established various agreement theorems in probabilistic dynamic epistemic logic. In particular, I established model- and frame-based versions of an agreement theorem with experimentation (Theorems 5.1 and 5.2), and of an agreement theorem with experimentation and communication (Theorems 5.3 and 5.4). I developed a sound and complete logical system within which the first agreement result is derivable (Theorem 5.5 and Remark 5.9). Throughout the chapter, I have emphasized that the models and logics are intuitively plausible, and directly connected with Aumann's original agreement result.

I have also discussed how these technical results can be seen as an application of the *dynamic turn* in logic. After showing that Aumann's original result fails to fully capture the essential dynamics behind the agreement theorem (both in its formulation and in its semantic setup), I argued that the agreement theorems established in this chapter *do* succeed in fully capturing this dynamics. In the first place, this means that these agreement theorems can be read as natural and explicit descriptions of the intuitive scenario that was behind Aumann's original result. Moreover, I showed that this perspective has important conceptual consequences, for example, for the role of common knowledge in agreement theorems. Common knowledge and communication seem to be two sides of the same coin: common knowledge is the result of communication, so if the communication dynamics is explicitly represented in the agreement theorem, there is no need anymore to *assume* common knowledge (as this will now *follow* from the communication protocol).

The technical and philosophical results presented in this chapter naturally suggest topics for further research. One issue that might be particularly interesting concerns the *scope* of the logical framework presented here. As I already mentioned earlier, game theorists have continued to work on extensions and refinements of Aumann's original agreement theorem. For example, Parikh and Krasucki (1990) have considered scenarios with different (non-public) communication protocols, and Monderer and Samet (1989) have shown that if common knowledge is replaced with common p-belief, then a weak version of the agreement theorem continues to hold (the agents' posteriors need no longer be identical, but their difference is bounded by a function of the parameter p). Furthermore, note that the current framework assumes that the experimentation and communication dynamics yield *hard* information: they lead to knowledge and full certainty (probability 1); one might wonder how the agreement theorem fares if one or both of these types of dynamics can yield *soft* information (i.e. lead to 'mere' beliefs and probabilities less than 1). It will be interesting to investigate whether such extensions and refinements can also be formalized in the framework of probabilistic dynamic epistemic logic.

6 The Dynamics of the Lockean Thesis

6.1 Introduction

Classical epistemology mainly uses *qualitative* notions, such as knowledge, belief, justification, etc. (Williams 2001). Formal epistemology, however, makes extensive use of *quantitative* notions, such as degrees of belief, coherence, confirmation, etc. (Douven and Meijs 2007, Eells and Fitelson 2000, Huber and Schmidt-Petri 2009). A natural question thus arises: what is the relationship—if any—between the qualitative and the quantitative framework? In particular, one may ask whether there is a relation between *belief* and *degrees of belief*.

A widespread thesis about this issue is that the qualitative notion of belief is reducible to the quantitative notion of degree of belief: believing that φ is defined as having a 'sufficiently high' degree of belief that φ. Foley (1992) argues that this thesis was hinted at by Locke (1975), and therefore labels it the 'Lockean thesis'. The main aim of this chapter is to explore the advantages and disadvantages of this thesis from the perspective of the dynamic turn in epistemic logic. To broaden the scope of the discussion, I will also briefly discuss its relation to the other main tenet of contemporary epistemic logic, viz. the focus on multi-agent operators. Based on these discussions, I will argue that, although the Lockean thesis is quite problematic for *classical* (static, single-agent) epistemic logic, it seems to have a much brighter future in *contemporary* (dynamic, multi-agent) epistemic logic.

The chapter is organized as follows. Section 6.2 compares classical and contemporary epistemic logic, focusing on the singe-agent/multi-agent and static/dynamic distinctions. Furthermore, I will point out that these distinctions interact with each other, and can be found not only in epistemic logic, but also in epistemology. Section 6.3 introduces the formal details of the Lockean thesis. This thesis yields a notion of belief which is not closed under conjunction; I will discuss the relationship of this problem with the well-known lottery paradox. In Section 6.4, I argue that the conjunction problem is typical for *classical* epistemic logic, and propose to reconsider the Lockean thesis from the perspective of *contemporary* epistemic logic. After briefly considering how this thesis can be generalized from single-agent to multi-agent contexts, I focus in Section 6.5 on its dynamic behavior. To this end, I will introduce a system of public announcement logic, enriched with a (qualitative) belief operator, and a system of probabilistic public announcement logic (in which the Lockean thesis can be applied to 'define' a belief operator). I will prove two theorems which say that the Lockean thesis leads to a unified perspective on the dynamic behavior of belief and degrees of belief under public announcements. In Section 6.6, I will explore the conceptual and philosophical consequences of these technical results. The theoretical elegance and practical applications of this unified account can be seen as a methodological argument for the Lockean thesis. Furthermore, I will argue that, when combined with Baltag's so-called 'Erlangen program' in epistemology, the possibility of such a unified account also constitutes a philosophical argument in favor of the Lockean thesis. Section 6.7, finally, summarizes the results obtained in this chapter, and suggests some questions for further inquiry.

6.2 Classical and Contemporary Epistemic Logic

The aim of this section is to introduce and discuss the most important features of contemporary epistemic logic, and to compare them with those of classical epistemic logic. First, however, it should be emphasized that, despite the terminology ('classical'/'contemporary') being used, the distinction being made is in the first place a *conceptual* one, rather than a strictly *historical* one. Most work on classical epistemic logic was being done before the emergence of contemporary epistemic logic, but one can certainly find examples of contemporary epistemic logic as early as the late 1960's—for example, Lewis (1969)—, and conversely, some logicians are still doing (very valuable) work in classical epistemic logic today—for example, Halpern et al. (2009).

The starting point of classical epistemic logic, and of epistemic logic in general, is Hintikka's seminal *Knowledge and Belief* (1962). In this work, knowledge is analyzed as a modal operator, which is given a semantics in terms of Kripke models. The formula $K_i\varphi$ thus means: 'agent i knows that φ'. Hintikka used his framework to gain insight about principles such as $K_i\varphi \to K_iK_i\varphi$ (if agent i knows that φ, does it then follow that she knows that she knows this?). The technical details of (probabilistic extensions of) this framework have already been addressed in great detail earlier in this thesis (in particular, see Sections 3.1 and 3.2.).

For our current purposes, two features are of central importance in this framework. First, the framework is essentially *single-agent*. It is about the knowledge of one single agent, not about the (pieces of) knowledge of several agents, and how these might interact. One can trivially go from one to many agents, by simply 'adding subscripts'; for example, one then gets formulas such as $K_i\varphi\wedge\neg K_j\varphi$ ('agent i knows that φ, but agent j doesn't'). However, in this way, one still cannot obtain the social notions of *common knowledge* and *distributed knowledge*. Syntactically speaking, the common knowledge and distributed knowledge operators are not definable in terms of the individual knowledge operators K_i, and thus have to be added as primitives into the object language, and axiomatized separately. Semantically speaking, the distributed knowledge and common knowledge operators are not interpreted with respect to the individual epistemic accessibility relations R_i, but rather with respect to their intersection and the reflexive transitive closure of their union, respectively.

Second, the framework is *static*. It focuses entirely on an agent's knowledge at a single point in time, without taking into consideration that the agent's knowledge might change over time (e.g. because she learns about new information). For example, Hintikka explicitly rules out occasions "on which people are engaged in gathering new factual information. Uttered on such an occasion, the sentences 'I don't know whether p' and [later] 'I know that p' need not be inconsistent" (1962, p. 7–8).

Contemporary epistemic logic (as this term is used here) can be defined as the opposite of classical epistemic logic with respect to exactly these two key features. In the first place, contemporary epistemic logic is a *multi-agent* enterprise. Because of applications in economics and computer science (distributed systems), the notion of common knowledge has become very important. A typical game-theoretical example of a multi-agent context is Aumann's agreeing to disagree theorem; the role of common knowledge in this context was discussed extensively in Chapter 5. Several characterizations of common knowledge are available; the most important ones are the iterative and the fixed-point characterization (Barwise 1988, Halpern and Moses 1990). Similarly, the notion of distributed knowledge has been studied extensively (van der Hoek et al. 1999, Roelofsen 2007).

In the second place, contemporary epistemic logic focuses on the *dynamics* of knowledge. One typically studies scenarios that involve learning: at first, an agent does not know whether φ; next, φ is (truthfully) announced; then, after the announcement, the agent *does* know that φ. Dynamic epistemic logic can be used to formalize and analyze such scenarios, but also more complicated ones. Several examples (including their probabilistic aspects) were discussed in Chapter 3, such as the Monty Hall puzzle (Example 3.2 on p. 88) and the Picasso scenario (Example 3.4 on p. 98).

It should be noted that these two themes (multi-agent/dynamics) often interact with each other. For example, the distinction between three important types of epistemic dynamics, viz. public announcements, private announcements, and semi-private announcements (Baltag and Moss 2004),[1] only makes sense in a multi-agent setting: in a single-agent setting, these three notions collapse into each other. Another example of the interaction between dynamics and multi-agent operators comes, again, from Aumann's agreeing to disagree theorem: in Section 5.6, I argued that common knowledge and dynamics are two sides of the same coin: if the communication dynamics is explicitly represented in the agreement theorem, common knowledge no longer needs to be assumed, but can rather be derived from the communication protocol.

Finally, it should be noted that this double evolution (from single-agent to multi-agent, from static to dynamic) has taken place not only in epistemic logic, but also in epistemology. Classical epistemology deals with the question whether a single agent, at a given point in time, does or does not possess knowledge concerning some proposition φ (intuitively, think of the Cartesian cogito, sitting quietly next to the fireplace and exploring the contents of its mind). Contemporary epistemologists, however, also deal with *social* (multi-agent) phenomena such as knowledge by testimony, pluralistic ignorance, and the role of experts (Goldman 1999, Pritchard 2004, Lackey and Sosa 2006). Furthermore, formal epistemologists study how new information should be processed, e.g. via Bayesian updat-

[1] The distinction between public and private announcements is illustrated by Examples 1.4 and 1.5 on p. 24. A semi-private announcement is an announcement that is made publicly to a subset of agents $I' \subset I$.

ing or Jeffrey conditionalization (Jeffrey 1983, Lange 2000). These analogous evolutions might have serious methodological consequences, since they seem to suggest a unified perspective on epistemic logic and epistemology.[2] For example, probabilistic public announcement logic and Bayesian epistemology offer similar analyses of learning new information; the subtle connection between these analyses was discussed in Section 3.3.3. Furthermore, various versions of dynamic epistemic logic have been used to analyze social-epistemic phenomena such as testimony and pluralistic ignorance (Holliday 2009, Hendricks 2010, Hansen 2012).

6.3 The Lockean Thesis

Classical epistemology mainly uses the qualitative notion of *belief*, whereas formal epistemology makes extensive use of the quantitative notion of *degrees of belief*. There exists a variety of frameworks in which degrees of belief can be formalized, such as possibility theory and ranking theory (Dubois and Prade 2009, Spohn 2009); the most widespread framework, however, is probability theory. Degrees of belief are then taken to be subjective probabilities, and one works with statements of the form $P(\varphi) = a$ (for some $a \in [0, 1]$), which means that the agent assigns probability a to proposition φ.

A well-known thesis, sometimes called the *Lockean thesis*, is that "it is epistemically rational for us to believe a proposition just in case it is epistemically rational for us to have *sufficiently high* degree of confidence in it" (Foley 1992, p. 111, my emphasis). Formally, this means that in a purely probabilistic framework, one can define ('qualitative') belief as follows:

$$B\varphi \;:\equiv\; P(\varphi) \geq \tau. \tag{6.1}$$

Here, τ is a *threshold*: a degree of belief in the proposition φ is 'sufficiently' high to count as a (qualitative) belief that φ iff that degree of belief is above τ.

There has been a lot of discussion about what the exact value of the threshold τ should be. There seems to be a consensus that τ should be strictly higher than 0.5, since otherwise the resulting notion of belief might violate the requirement that belief be *consistent*:

$$B\varphi \rightarrow \neg B \neg\varphi.$$

[2] Similar remarks are made in Demey (2011b).

For example, if $\tau = 0.4$, then for a proposition φ with $P(\varphi) = 0.45 \geq \tau$, thesis (6.1) yields $B\varphi$, but it also follows that $P(\neg\varphi) = 1 - P(\varphi) = 1 - 0.45 = 0.55 \geq \tau$, and thus also $B\neg\varphi$. On the other hand, if $\tau > 0.5$, this cannot occur: if $B\varphi$, then $P(\varphi) \geq \tau > 0.5$, and thus $P(\neg\varphi) = 1 - P(\varphi) < 0.5 < \tau$, i.e. $\neg B\neg\varphi$.

Some authors have proposed to take $\tau = 1$, but this suggestion seems to be too strong: belief intuitively does not require complete certainty. For $\tau < 1$, however, a well-known problem for the Lockean thesis arises: the resulting notion of belief is not closed under conjunction. For example, suppose that $\tau = 0.6$, consider a fair six-faced die, write p for 'the die will land with 1,2,3 or 4 eyes up' and q for 'the die will land with 3,4,5 or 6 eyes up'; then

$$P(p) = P(q) = 0.66 \geq 0.6, \text{ and } P(p \wedge q) = 0.33 < 0.6$$

and thus (6.1) yields

$$Bp \wedge Bq \wedge \neg B(p \wedge q).$$

One might think that this problem can be solved by taking τ to be increasingly closer to 1, e.g. 0.95. However, consider a fair lottery with 100 tickets (the agent considers all tickets equally likely to win, and exactly one ticket will win) and write p_i for 'ticket i will not win'; then

$$P(p_i) = 0.99 \geq 0.95 \text{ for each } i, \text{ and } P\Big(\bigwedge_{i=1}^{100} p_i\Big) = 0 < 0.95$$

and thus (6.1) yields

$$\bigwedge_{i=1}^{100} Bp_i \wedge \neg B\Big(\bigwedge_{i=1}^{100} p_i\Big).$$

This explains the central role of the lottery paradox in this context: fair lotteries form a canonical class of counterexamples to belief being closed under conjunction. No matter how close to 1 the value of τ is taken to be, one can construct a fair lottery (with a sufficiently large number of tickets) which yields a finite number of propositions that are believed (i.e. their probabilities are $\geq \tau$), while their conjunction is not believed (i.e. its probability is $< \tau$).

Given these considerations, it will be assumed in the remainder of this chapter that the value of the threshold lies in the open interval $(0.5, 1)$, i.e. $0.5 < \tau < 1$.

6.4 The Lockean Thesis from the Perspective of Contemporary Epistemic Logic

In the previous section, I introduced the Lockean thesis, and discussed its most important problem, viz. that it yields a notion of belief which is not closed under conjunction. Recalling the distinction between classical and contemporary epistemic logic from Section 6.2, it is clear that this problem is to be situated in *classical* epistemic logic: it is about a single agent, and it is entirely static (the agent's beliefs are examined at a single point in time).

One can also ask, however, how the Lockean thesis fares from the perspective of *contemporary* epistemic logic. Typical issues that arise in this perspective are:

1. Can the Lockean thesis also be used to define interesting multi-agent notions of belief, such as common belief?

2. Does the Lockean thesis generate interesting behavior under various types of epistemic dynamics?

Game theorists use the Lockean thesis to define a notion of belief they call p-belief (because they usually use the letter p, instead of τ, to denote the threshold value). Just like 'classical' belief (and knowledge) can be used to define common belief (and common knowledge), the notion of p-belief can be used to define a natural notion of *common p-belief.* The formal behavior of common p-belief largely resembles that of 'classical' common belief; for example, it has both an iterative and a fixed-point characterization (Monderer and Samet 1989, Kajii and Morris 1997).

Furthermore, many applications that require the notion of common belief can equally well be modeled using the notion of common p-belief. For example, the agreeing to disagree theorem was first established by Aumann (1976), using the notion of common knowledge. Dégremont and Roy (2009, 2012) prove a logical version of this theorem that only requires the 'classical' (qualitative) notion of common belief; however, Monderer and Samet (1989) already established a version of this theorem using the notion of common *p-belief.* To summarize: both from a theoretical and an application-oriented[3] perspective, the Lockean thesis seems to transfer well from the single-agent to the multi-agent case.

[3] I will return to this application-oriented perspective on the Lockean thesis in Section 6.6.

In sum, then, the first issue (multi-agent contexts) has been studied quite extensively (especially in game theory). The second issue (dynamics), however, has largely been ignored in the literature so far. In the remainder of this chapter, I will explore exactly this. I will study the dynamic behavior of the notion of belief generated by the Lockean thesis, and compare it with the dynamic behavior of a 'classical' qualitative notion of belief. The focus will be on one particular type of dynamics, viz. public announcements.

6.5 A Dynamic Perspective on the Lockean Thesis

In order to formally compare the probabilistic belief operator defined by means of the Lockean thesis and the 'primitive' qualitative belief operator, it is necessary to introduce logical systems in which each of these two operators can be interpreted. Subsection 6.5.1 introduces a system of public announcement logic, enriched with a (qualitative) belief operator. Subsection 6.5.2 introduces probabilistic public announcement logic, in which the Lockean thesis can be applied to define a belief operator. The technical details of these two systems have already been extensively discussed elsewhere in this thesis (in particular, in Chapters 3 and 4), so these two subsections are quite brief, and focus on those aspects that are most relevant for our current purposes. Next, in Subsection 6.5.3, I show that the qualitative belief operator and the probabilistically defined belief operator display the same dynamic behavior with respect to public announcements.

6.5.1 Public Announcement Logic with Beliefs

I will now give a brief overview of a system of public announcement logic, enriched with a belief operator. It is well-known that such systems cannot plausibly be interpreted on Kripke models: if an agent receives a true piece of information φ while previously believing that $\neg\varphi$, then this agent is predicted to go insane and start believing *everything* (rather than performing a realistic process of *belief revision*)—thus contradicting the consistency requirement about belief (van Benthem 2007, Section 3.1). Therefore, systems of public announcement logic with a belief operator have to be interpreted on epistemic plausibility models.

Some of the main authors in this area, in particular van Benthem (2007) and Baltag and Smets (2008), use subtly different notions of epistemic plausibility models. In Chapter 4, I argued that Baltag and Smets's notion is superior over that of van Benthem. Hence, the epistemic plausibility models that will be used here are those defined by Baltag and Smets (2008) (see Definition 4.2 on p. 110 for a formal definition).

The qualitative language $\mathcal{L}_{qual}$ that will be interpreted on such models is defined by means of the following BNF:

$$\varphi \ ::= \ p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid B_i(\varphi \mid \varphi)$$

$B_i(\cdot \mid \cdot)$ is a conditional belief operator; formulas of the form $B_i(\varphi \mid \psi)$ should be read as: 'agent i believes that φ, conditional on ψ'.[4] Recall that the ordinary belief operator can easily be defined in terms of the conditional belief operator, by putting

$$B_i\varphi \ :\equiv \ B_i(\varphi \mid \top). \tag{6.2}$$

The semantics of the conditional belief and ordinary belief operators looks as follows (recall Definition 4.3 on p. 110):

$$\begin{array}{lll} \mathbb{M}, w \models B_i(\varphi \mid \psi) & \text{iff} & \forall v \in \text{Min}_{\leq_i}(R_i[w] \cap [\![\psi]\!]^{\mathbb{M}})\colon \mathbb{M}, v \models \varphi, \\ \mathbb{M}, w \models B_i\varphi & \text{iff} & \forall v \in \text{Min}_{\leq_i}(R_i[w])\colon \mathbb{M}, v \models \varphi. \end{array}$$

As for the dynamics, we focus exclusively on public announcements. We extend $\mathcal{L}_{qual}$ to $\mathcal{L}^!_{qual}$, by adding a public announcement operator $[!\cdot]\cdot$. As usual, formulas of the form $[!\varphi]\psi$ should be read as: 'after any public announcement of φ, it will be the case that ψ'. The dual of $[!\varphi]\psi$ is defined as $\langle !\varphi\rangle\psi :\equiv \neg[!\varphi]\neg\psi$, and should be read as: 'φ can be announced, and afterwards it will be the case that ψ'. Recall that given an epistemic plausibility model $\mathbb{M}$ and a formula $\varphi \in \mathcal{L}^!_{qual}$, we can define the updated epistemic plausibility model $\mathbb{M}|\varphi$ (see Definition 4.4 on p. 111 for details). We can now state the usual semantics for the public announcement operator:

$$\begin{array}{lll} \mathbb{M}, w \models [!\varphi]\psi & \text{iff} & \text{if } \mathbb{M}, w \models \varphi \text{ then } \mathbb{M}|\varphi, w \models \psi, \\ \mathbb{M}, w \models \langle !\varphi\rangle\psi & \text{iff} & \mathbb{M}, w \models \varphi \text{ and } \mathbb{M}|\varphi, w \models \psi. \end{array}$$

The dynamic behavior generated by public announcements is completely described by means of *reduction axioms*. These are biconditional statements which

[4] We usually write $B_i^{\psi}\varphi$ instead of $B_i(\varphi \mid \psi)$. In this chapter, however, I will use the latter notation, to suggest a similarity between conditional belief and conditional probability. Indeed, the entire point of this chapter is that this suggestion has some interesting implications, and should therefore be taken seriously.

allow us to recursively rewrite formulas containing dynamic operators as formulas without such operators; hence, the dynamic language $\mathcal{L}^{!}_{qual}$ is equally expressive as the static $\mathcal{L}_{qual}$, and proving completeness for the dynamic logic can be reduced to that of the static logic (van Ditmarsch et al. 2007). For our current purposes, however, it is more important to note that reduction axioms can also be seen as 'predicting' what will be the case *after* the dynamics has taken place in terms of what is the case *before* the dynamics has taken place.

For expository purposes, I first state the reduction axiom for the ordinary belief operator:

$$[!\varphi]B_i\psi \longleftrightarrow \Big(\varphi \to B_i(\langle !\varphi\rangle\psi \mid \varphi)\Big). \tag{6.3}$$

This illustrates the two perspectives on reduction axioms discussed above. First of all, when (6.3) is read 'from left to right', it states that the public announcement operator $[!\varphi]$ can be 'pushed through' the complex formula $B_i\varphi$: on the right-hand side its scope is just ψ, which has a lower complexity than the original $B_i\psi$. Using the other reduction axioms as well, one can thus rewrite $[!\varphi]B_i\psi$ as a formula that does not involve the public announcement operator at all. Secondly, when (6.3) is read 'from right to left', it 'predicts' that agent i will believe that ψ *after* the public announcement of φ, just in case *before* the announcement, she believed $\langle !\varphi\rangle\psi$, conditional on φ.

Note that (6.3), which is the reduction axiom for the *ordinary* belief operator, requires the *conditional* belief operator to be expressible; this is one of the reasons for introducing this conditional belief operator from the start.[5] The reduction axiom for the conditional belief operator looks as follows:

$$[!\varphi]B_i(\psi \mid \alpha) \longleftrightarrow \Big(\varphi \to B_i(\langle !\varphi\rangle\psi \mid \langle !\varphi\rangle\alpha)\Big). \tag{6.4}$$

If we take $\alpha = \top$ and note that $\langle !\varphi\rangle\top$ is equivalent to φ, it is easy to see that the reduction axiom (6.3) for ordinary belief is just a special case of the reduction axiom (6.4) for conditional belief.

6.5.2 Probabilistic Public Announcement Logic

I will now provide a brief overview of probabilistic public announcement logic. Unlike the system discussed in the previous subsection, this system does not

[5] In other words, this is another example of enriching the static language to ensure the expressibility of all reduction axioms; also see Footnote 9 on p. 112.

contain a ‘primitive’ belief operator; however, since it can express probabilistic information, the Lockean thesis can be applied to it to obtain a ‘defined’ belief operator (this will be discussed in detail in Subsection 6.5.3).

We will work with well-behaved probabilistic Kripke models (see Definitions 3.2, 3.3 and 3.4 on p. 75, 79 and 81 for technical details).

The probabilistic language $\mathcal{L}_{prob}$ that will be interpreted on such models is defined by means of the following BNF:

$$\varphi \ ::= \ p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid a_1 P_i(\varphi) + \cdots + a_n P_i(\varphi) \geq k$$

As was discussed earlier, allowing for linear combinations of probability terms has a technical motivation. In the present context, however, we will usually just be working with formulas of the form $P_i(\varphi) \geq k$, which should be read as: ‘agent i assigns (subjective) probability (i.e. degree of belief) at least k to φ’. The only use we make here of linear combinations of probabilities, is to introduce conditional probabilities into the formal language $\mathcal{L}_{prob}$. Recall that in probability theory, the conditional probability of φ given ψ is defined as follows (provided that $P(\psi) > 0$):

$$P(\varphi \mid \psi) = \frac{P(\varphi \wedge \psi)}{P(\psi)}.$$

It thus makes sense to introduce the following definition in the formal language (recall Equation 3.2 on p. 92):

$$P_i(\varphi \mid \psi) \geq k \ :\equiv \ P_i(\varphi \wedge \psi) - kP_i(\psi) \geq 0. \tag{6.5}$$

Formulas of the form $P_i(\varphi \mid \psi) \geq k$ should be read as ‘agent i assigns conditional probability at least k to φ, conditional on ψ’. Consider the following chain of equivalences:

$$\begin{array}{rcl} P_i(\varphi \mid \top) \geq k & \equiv & P_i(\varphi \wedge \top) - kP_i(\top) \geq 0 \\ & \equiv & P_i(\varphi) - k \cdot 1 \geq 0 \\ & \equiv & P_i(\varphi) \geq k. \end{array}$$

In sum, we thus get:

$$P_i(\varphi) \geq k \ \ \equiv \ \ P_i(\varphi \mid \top) \geq k. \tag{6.6}$$

It should be emphasized that this formula merely expresses an equivalence in probabilistic public announcement logic; it does not *define* the absolute probability on the left in terms of the conditional probability on the right. On the contrary, absolute probabilities are taken as primitive, and conditional probabilities are defined in terms of them by Equation 3.2.

The semantics of i-probability formulas looks as follows (this clause was originally stated on p. 78):

$$\mathbb{M}, w \models \textstyle\sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) \geq k \quad \text{iff} \quad \textstyle\sum_{\ell=1}^{n} a_\ell \mu_i(w)([\![\, \varphi_\ell \,]\!]^{\mathbb{M}}) \geq k.$$

I now turn to the dynamic part, again focusing on public announcements. We extend $\mathcal{L}_{prob}$ to $\mathcal{L}_{prob}^{!}$, by adding a public announcement operator $[!\cdot]\cdot$. The meaning of $[!\varphi]\psi$ and its dual $\langle !\varphi \rangle \psi$ are exactly the same as in the previous subsection. Recall that given a well-behaved probabilistic Kripke model $\mathbb{M}$ and a formula $\varphi \in \mathcal{L}_{prob}^{!}$, we can define the updated well-behaved probabilistic Kripke model $\mathbb{M}|\varphi$ (see Definition 3.6 on p. 86 for details). Again, we can now state the usual semantics for the public announcement operator:

$$\begin{array}{lll} \mathbb{M}, w \models [!\varphi]\psi & \text{iff} & \text{if } \mathbb{M}, w \models \varphi \text{ then } \mathbb{M}|\varphi, w \models \psi, \\ \mathbb{M}, w \models \langle !\varphi \rangle \psi & \text{iff} & \mathbb{M}, w \models \varphi \text{ and } \mathbb{M}|\varphi, w \models \psi. \end{array}$$

To study the dynamic behavior of i-probability formulas under public announcements, we will again look at its reduction axioms. The reduction axiom for the formula $P_i(\psi) \geq k$ reads as follows:[6]

$$[!\varphi]P_i(\psi) \geq k \longleftrightarrow \Big(\varphi \to P_i(\langle !\varphi \rangle \psi \mid \varphi) \geq k\Big). \tag{6.7}$$

Note that to formulate a reduction axiom for the formula $P_i(\varphi) \geq k$, we used conditional probabilities. These can be defined in the formal language, using Equation 6.5. Hence, it is not strictly necessary to provide a separate reduction axiom for them.[7] Still, by unpacking Equation 6.5, we easily obtain a reduction axiom for $P_i(\psi \mid \alpha) \geq k$ as well:

$$[!\varphi]P_i(\psi \mid \alpha) \geq k \longleftrightarrow \Big(\varphi \to P_i(\langle !\varphi \rangle \psi \mid \langle !\varphi \rangle \alpha) \geq k\Big). \tag{6.8}$$

[6] To achieve full generality, one needs to provide a reduction axiom not just for $P_i(\psi) \geq k$, but rather for $\sum_{\ell=1}^{n} a_\ell P_i(\psi_\ell) \geq k$. This can easily be done; however, for our present purposes, it will suffice to focus on the simpler case $P_i(\psi) \geq k$.

[7] In the previous subsection, it *was* necessary to provide a separate reduction axiom for conditional belief, since *that* operator is *not* definable in the formal language.

It should be emphasized that (6.8) is a ‘reduction axiom’ in name only: it can be *derived* from the reduction axiom for absolute probabilities (6.7) and the definition of conditional probabilities in terms of absolute probabilities (6.5).

6.5.3 Unification of the Reduction Axioms

Let’s take stock. In Subsection 6.5.1, I discussed a system of public announcement logic, enriched with a qualitative notion of (conditional) belief. This system gives rise to the reduction axioms (6.3) and (6.4), for belief and conditional belief, respectively. In Subsection 6.5.2, I discussed probabilistic public announcement logic. This system gives rise to the reduction axioms (6.7) and (6.8), for probability and conditional probability, respectively.

Note that the reduction axioms (6.7) and (6.8) hold for any value of k, so in particular also for τ, i.e. the threshold value used in the Lockean thesis. Any concrete instance of (6.7) or (6.8) in which the value of k is τ will be called a τ-instance. For example, if we assume that $\tau = 0.85$ (this satisfies the restriction that $0.5 < \tau < 1$ that was imposed at the end of Section 6.3), then the formula

$$[!p]P_i(q \mid r) \geq 0.85 \longleftrightarrow \Big(p \rightarrow P_i(\langle !p\rangle q \mid \langle !p\rangle r) \geq 0.85\Big)$$

is a τ-instance of (6.8) (with p, q and r for φ, ψ and α, respectively); on the other hand,

$$[!p]P_i(q \mid r) \geq 0.75 \longleftrightarrow \Big(p \rightarrow P_i(\langle !p\rangle q \mid \langle !p\rangle r) \geq 0.75\Big)$$

certainly is an instance of (6.8) (again with p, q and r for φ, ψ and α, respectively), but it is obviously not a τ-instance, since $\tau = 0.85 \neq 0.75$.

Recall that the Lockean thesis says that belief can be defined as ‘high’ probability:

$$B_i\varphi :\equiv P_i(\varphi) \geq \tau.$$

Given the highly similar relationships between belief and conditional belief (Equation 6.2) on the one hand and between probability and conditional probability (Equation 6.6) on the other hand, it makes sense to also consider a slightly more sophisticated version of the Lockean thesis, which says that *conditional* belief can be defined as ‘high’ *conditional* probability:

$$B_i(\varphi \mid \psi) :\equiv P_i(\varphi \mid \psi) \geq \tau.$$

This version of the Lockean thesis can be used to define a mapping π_τ from the qualitative language $\mathcal{L}^!_{qual}$ to the probabilistic language $\mathcal{L}^!_{prob}$.

Definition 6.1. The mapping $\pi_\tau\colon \mathcal{L}^!_{qual} \to \mathcal{L}^!_{prob}$ is defined by induction on formula complexity:

$$\begin{aligned}
\pi_\tau(p) &:= p,\\
\pi_\tau(\neg\varphi) &:= \neg\pi_\tau(\varphi),\\
\pi_\tau(\varphi\wedge\psi) &:= \pi_\tau(\varphi)\wedge\pi_\tau(\psi),\\
\pi_\tau(B_i(\varphi\mid\psi)) &:= P_i(\pi_\tau(\varphi)\mid\pi_\tau(\psi))\geq\tau,\\
\pi_\tau([!\varphi]\psi) &:= [!\pi_\tau(\varphi)]\pi_\tau(\psi).
\end{aligned}$$

The only effect of π_τ is thus that it 'probabilifies' the belief operator according to (the sophisticated version of) the Lockean thesis: each conditional belief formula $B_i(\varphi\mid\psi)$ is mapped to the conditional probability formula in terms of which it is defined according to this thesis, viz. $P_i(\varphi\mid\psi)\geq\tau$. It is easy to see that $\pi_\tau(\varphi\bullet\psi)=\pi_\tau(\varphi)\bullet\pi_\tau(\psi)$ for all binary Boolean connectives, and that $\pi_\tau(\top)=\top$.

Theorem 6.1. *If a formula $\lambda\in\mathcal{L}^!_{qual}$ is an instance of the reduction axiom (6.4) for conditional belief, then $\pi_\tau(\lambda)\in\mathcal{L}^!_{prob}$ is a τ-instance of the reduction axiom (6.8) for conditional probability.*

Proof. If λ is an instance of (6.4), then λ is of the form

$$[!\varphi]B_i(\psi\mid\alpha)\longleftrightarrow\Big(\varphi\to B_i(\langle!\varphi\rangle\psi\mid\langle!\varphi\rangle\alpha)\Big),$$

and hence

$$\begin{aligned}
\pi_\tau(\lambda) &= \pi_\tau\Bigg([!\varphi]B_i(\psi\mid\alpha)\longleftrightarrow\Big(\varphi\to B_i(\langle!\varphi\rangle\psi\mid\langle!\varphi\rangle\alpha)\Big)\Bigg)\\
&= \pi_\tau\Big([!\varphi]B_i(\psi\mid\alpha)\Big)\longleftrightarrow\pi_\tau\Big(\varphi\to B_i(\langle!\varphi\rangle\psi\mid\langle!\varphi\rangle\alpha)\Big)\\
&= [!\pi_\tau(\varphi)]\pi_\tau\big(B_i(\psi\mid\alpha)\big)\longleftrightarrow\Big(\pi_\tau(\varphi)\to\pi_\tau\big(B_i(\langle!\varphi\rangle\psi\mid\langle!\varphi\rangle\alpha)\big)\Big)\\
&= [!\pi_\tau(\varphi)]P_i\big(\pi_\tau(\psi)\mid\pi_\tau(\alpha)\big)\geq\tau\longleftrightarrow\\
&\quad\Big(\pi_\tau(\varphi)\to P_i\big(\pi_\tau(\langle!\varphi\rangle\psi)\mid\pi_\tau(\langle!\varphi\rangle\alpha)\big)\geq\tau\Big)\\
&= [!\pi_\tau(\varphi)]P_i\big(\pi_\tau(\psi)\mid\pi_\tau(\alpha)\big)\geq\tau\longleftrightarrow\\
&\quad\Big(\pi_\tau(\varphi)\to P_i\big(\langle!\pi_\tau(\varphi)\rangle\pi_\tau(\psi)\mid\langle!\pi_\tau(\varphi)\rangle\pi_\tau(\alpha)\big)\geq\tau\Big)
\end{aligned}$$

—which is clearly a τ-instance of (6.8). □

Theorem 6.2. *If a formula $\lambda \in \mathcal{L}^!_{qual}$ is an instance of the reduction axiom (6.3) for belief, then $\pi_\tau(\lambda) \in \mathcal{L}^!_{PKM}$ is a τ-instance of the reduction axiom (6.7) for probability.*

Proof. If λ is an instance of (6.3), then λ is of the form

$$[!\varphi]B_i\psi \longleftrightarrow \Big(\varphi \to B_i(\langle !\varphi\rangle\psi \mid \varphi)\Big),$$

and hence

$$\begin{aligned}
\pi_\tau(\lambda) &= \pi_\tau\bigg([!\varphi]B_i\psi \longleftrightarrow \Big(\varphi \to B_i(\langle !\varphi\rangle\psi \mid \varphi)\Big)\bigg)\\
&\overset{(6.2)}{=} \pi_\tau\bigg([!\varphi]B_i(\psi \mid \top) \longleftrightarrow \Big(\varphi \to B_i(\langle !\varphi\rangle\psi \mid \varphi)\Big)\bigg)\\
&= \pi_\tau\Big([!\varphi]B_i(\psi \mid \top)\Big) \longleftrightarrow \pi_\tau\Big(\varphi \to B_i(\langle !\varphi\rangle\psi \mid \varphi)\Big)\\
&= [!\pi_\tau(\varphi)]\pi_\tau\big(B_i(\psi \mid \top)\big) \longleftrightarrow \Big(\pi_\tau(\varphi) \to \pi_\tau\big(B_i(\langle !\varphi\rangle\psi \mid \varphi)\big)\Big)\\
&= [!\pi_\tau(\varphi)]P_i\big(\pi_\tau(\psi) \mid \pi_\tau(\top)\big) \geq \tau \longleftrightarrow\\
&\quad \Big(\pi_\tau(\varphi) \to P_i\big(\pi_\tau(\langle !\varphi\rangle\psi) \mid \pi_\tau(\varphi)\big) \geq \tau\Big)\\
&= [!\pi_\tau(\varphi)]P_i\big(\pi_\tau(\psi) \mid \top\big) \geq \tau \longleftrightarrow\\
&\quad \Big(\pi_\tau(\varphi) \to P_i\big(\langle !\pi_\tau(\varphi)\rangle\pi_\tau(\psi) \mid \pi_\tau(\varphi)\big) \geq \tau\Big)\\
&\overset{(6.6)}{=} [!\pi_\tau(\varphi)]P_i\big(\pi_\tau(\psi)\big) \geq \tau \longleftrightarrow\\
&\quad \Big(\pi_\tau(\varphi) \to P_i\big(\langle !\pi_\tau(\varphi)\rangle\pi_\tau(\psi) \mid \pi_\tau(\varphi)\big) \geq \tau\Big)
\end{aligned}$$

—which is clearly a τ-instance of (6.7). □

Theorems 6.1 and 6.2 say that the π_τ-translation of the reduction axiom for (conditional) belief is exactly the reduction axiom for high (conditional) probability (where ‘high’ means ‘strictly greater than τ’). In other words: if one accepts the Lockean thesis and its slightly more sophisticated version as probabilistic definitions of (conditional) belief, then the reduction axioms for these probabilistically defined notions are *exactly the same* syntactic expressions as the reduction axioms for the ‘primitive’, qualitative notions of (conditional) belief.

Accepting the Lockean thesis leads to a unified perspective on the dynamic behavior of belief and probabilities (degrees of belief). In the next section, I will discuss the methodological and philosophical importance of this technical observation.

6.6 Methodological and Philosophical Importance

By themselves, Theorems 6.1 and 6.2 are strictly technical results, which are mathematically provable and philosophically 'neutral'. Hence, any methodological and/or philosophical consequences that one wishes to draw from them, will have to be supported by additional (non-formal) argumentation. In this section, I will argue that the availability of these theorems can be seen as a reason to adopt the Lockean thesis after all (despite its problems with being closed under conjunction). To do this, I will discuss three distinct, increasingly strong interpretations of these theorems.

The first reaction one might have about Theorems 6.1 and 6.2 is to regard them as merely technical 'artefacts', which do not have any further conceptual or philosophical implications. However, this interpretation seems to ignore the vast formal and conceptual distance between the reduction axioms for (conditional) belief on the one hand, and for high (conditional) probability on the other. For example, the reduction axioms (6.3–6.4) for (conditional) belief are $\mathcal{L}^!_{qual}$-formulas that are interpreted on (Baltag/Smets-type) epistemic plausibility models. These models are purely *qualitative* entities: (conditional) belief is interpreted by looking at $\leq_i$-minimal states, and the definition of an updated plausibility model is a straightforward extension of the well-known definition of an updated Kripke model. On the other hand, the reduction axioms (6.7–6.8) for (conditional) probability are $\mathcal{L}^!_{prob}$-formulas that are interpreted on well-behaved probabilistic Kripke models. These models have a large *quantitative* (probabilistic) component: (conditional) probability-formulas are interpreted by means of the probability mass functions $\mu_i(w)$, and the definition of an updated probabilistic Kripke model essentially involves the idea of Bayesian conditionalization (including the arithmetic of division, etc.).

Keeping in mind this formal and conceptual distance between the frameworks of qualitative (conditional) belief and high (conditional) probability, it is all the more surprising that the Lockean thesis, together with its more sophisticated version, leads to a precise unification between the reduction axioms of both frameworks. Therefore, the second interpretation takes Theorems 6.1 and 6.2 to constitute a *pragmatic* or *methodological* argument in favor of the Lockean thesis. Accepting this thesis leads to a significant and unexpected unified perspective on the dynamic properties of technically and conceptually very different frameworks, and thus helps to focus on the common purpose of these frameworks (despite their technical differences), viz. providing an account of the agents' 'soft information' and its dynamics.

This is also relevant for the practical or philosophical applications of these frameworks. If a certain application requires that one focuses on the dynamics of belief, but less on its static properties (such as closure under conjunction), then both frameworks described in this chapter are equally applicable, and thus the final decision about which system to use will have to be motivated by other considerations. As a concrete example, Chapter 7 will provide a detailed analysis of the epistemic aspects of surprise, focusing on its dynamic behavior (a typical topic that will be analyzed is the relationship between surprise and belief revision). Given its strong focus on dynamics, this analysis can be carried out in the qualitative framework of epistemic plausibility models—with a primitive notion of (conditional) belief—, but also in the quantitative framework of probabilistic Kripke models—with a notion of (conditional) belief defined according to the Lockean thesis, viz. as high (conditional) probability. Since probabilities are needed in the analysis for other, independent reasons as well (viz. as a quantitative representation of *intensity of surprise*), the methodological interpretation holds that there is no need to introduce 'primitive' belief operators, and that one can thus simply employ probabilistic Kripke models (in particular, see Subsection 7.4.4).

Applications such as these illustrate the *pragmatic* importance of the Lockean thesis. According to a third and final interpretation, however, it might be possible to draw even further *philosophical* conclusions from the technical observations made in Subsection 6.5.3. Baltag (2008, 2011) has argued for an 'Erlangen program' for epistemology, which can be described as follows (Baltag 2011, p. 4):

> in the spirit of Felix Klein's 1862 Erlangen program for mathematics, I argue that 'static' epistemic notions and properties are best characterized in terms of their transformations, their potential dynamics

It was shown above that if one accepts the Lockean thesis (and its more sophisticated version)—if only for methodological or pragmatic reasons, such as in the analysis of surprise—, the epistemic notions of (conditional) belief and high (conditional) probability display exactly the same dynamic behavior (i.e. they have the same reduction axioms) with respect to public announcements. Baltag's Erlangen program for epistemology uses exactly this dynamic behavior to characterize epistemic notions, and therefore classifies (conditional) belief and high (conditional) probability as one and the same epistemic notion. But this exactly means that the Lockean thesis should be accepted, not merely as a practically useful hypothesis, but also as a substantial epistemological claim about the notion of belief.

At this point, it might be objected that belief and high probability really cannot be the same epistemic notion, simply because the former notion is closed under conjunction, whereas the latter isn't (recall Section 6.3). However, from the perspective of Baltag's Erlangen program, this difference is a static difference (not a dynamic one), and should not be accepted as the sole criterium of individuation for epistemic concepts. With respect to dynamic behavior, which is deemed a more relevant individuation criterium in Baltag's Erlangen program, belief and high probability *do* have the same properties. In other words: the difference with respect to closure under conjunction might indicate that belief and high probability are not the same notion *altogether*, but from an *epistemic* perspective, they cannot be distinguished (the difference arises only at a *non-epistemic* level, for example the psychological level).

6.7 Conclusion

In this chapter, I have studied the Lockean thesis about beliefs and degrees of belief from the perspective of contemporary epistemic logic. The main problem of this thesis, viz. that it gives rise to a notion of belief which is not closed under conjunction, is typical for *classical* epistemic logic. I have argued that in *contemporary* epistemic logic, this thesis seems to have a much brighter future.

In the first place, I have briefly pointed out that the Lockean thesis can easily be extended from single-agent to multi-agent settings (via the notion of common p-belief). More importantly, however, I have shown that accepting it (and a more sophisticated version for conditional beliefs) leads to a significant and unexpected unification in the dynamic behavior of (conditional) belief (interpreted on epistemic plausibility models) and high (conditional) probability (interpreted on probabilistic Kripke models) with respect to public announcements. This already constitutes a strong argument for the methodological usefulness of the Lockean thesis. Furthermore, if one accepts Baltag's Erlangen program for epistemology, this technical observation has even stronger philosophical implications: because belief and high probability display the same dynamic behavior, it is very plausible that they are indeed one and the same epistemic notion.

Obviously, much more work needs to be done on this topic. In this chapter, it was shown that belief and high probability have the same dynamic behavior *with respect to public announcements*. However, for Baltag's Erlangen program to reach its full force, it is necessary to show that these two notions have the same dynamic behavior *in general*, i.e. with respect to an entire range of other types of dynamic phenomena. Secondly, there is a more philosophical issue that needs to be addressed. So far, Baltag has only provided a negative motivation for his epistemological Erlangen program: all attempts by classical epistemology to provide static definitions of the main epistemic notions (for example: 'knowledge at time t is defined as justified true belief at time t')[8] have utterly failed, and therefore it seems worthwile to look at an entirely new sort of individuation criterium, viz. sameness of dynamic behavior (this criterium has already proved to be successful in another area, viz. geometry). Still, if this Erlangen program is to develop into a mature epistemological position, much more work will need to be done—in particular, providing a positive motivation.

[8] A notable exception is Goldman (1979)'s 'historical reliabilism'. However, this theory is 'backward-looking' (epistemic states are characterized in terms of how they are generated), whereas Baltag's proposal is 'forward-looking' (epistemic states are characterized in terms of how they can change).

7 The Dynamics of Surprise

7.1 Introduction

The phenomenon of surprise is ubiquitous in everyday life. People get surprised all the time; for example, by an unexpected flash of light, or—more 'down to earth'—about the fact that their local grocery store has run out of milk (after all, the store is usually well-stocked!). The role of surprise in human life has been intensively studied in psychology from cognitive, social, developmental and educational perspectives. Furthermore, computer scientists have implemented the psychological findings about human surprise in artificial agents, and used logical models to describe these agent architectures. Surprise even crops up in various philosophical debates, such as those concerning the role of surprising evidence in Bayesian epistemology, or concerning the so-called surprise examination paradox.[1]

The overarching goal of this chapter is to provide a new analysis of the phenomenon of surprise in the framework of probabilistic dynamic epistemic logic. This account is based on the vast amount of experimental work on surprise in psychology, which should benefit its empirical adequacy. The chapter's main thesis, however, is of a more conceptual nature: surprise is an essentially *dynamic* phenomenon, and any good formal analysis should represent this dynamics explicitly. I will argue that all current formalizations of surprise in artificial intelligence and logic fail to fully capture this dynamics, and show that the framework developed in this paper *is* able to capture it. As an additional benefit, this new framework can be used to analyze some aspects of surprise that could not be analyzed before.

[1] These philosophical debates will not be addressed directly in this chapter; for overviews, the reader can consult Talbott (2008) and Chow (1998), respectively.

This enterprise is motivated by a variety of interrelated issues. In the first place, a logical perspective on surprise can help to elucidate the basic properties of this notion. Starting from the concrete empirical results about surprise, a complete axiomatization is proposed in which the observed behavioral patterns can be derived as theorems. In other words, the fundamental laws of surprise can be 'reverse engineered' out of the concrete behavior that they generate. Secondly, the resulting logical system serves as a highly expressive language to formally specify agent architectures; it belongs to the general framework of (dynamic) epistemic logic, which is becoming a contemporary 'lingua franca' in multi-agent systems (Wooldridge 2002, Shoham and Leyton-Brown 2009). Thirdly, and most importantly, since the conceptual and technical advantages of this system mainly stem from the fact that it explicitly represents the dynamics of surprise, it also constitutes another concrete illustration of the (strong interpretation of the) *dynamic turn* in logic (van Benthem 2011; also see Chapter 1).

The remainder of the chapter is organized as follows. Section 7.2 briefly reviews the literature on surprise in cognitive science, multi-agent systems and logic. In Section 7.3 I argue that two earlier formalizations do not adequately represent the dynamic nature of surprise, and make some suggestions on how this can be achieved. In Section 7.4, then, I show how these suggestions can be developed into a full-fledged dynamic logic of surprise, which can capture several key aspects of surprise, such as its transitory (short-lived) nature and its role in belief revision. Finally, Section 7.5 wraps things up, and discusses some potential lines of further research.

7.2 Three Perspectives on Surprise

This section provides an overview of the literature on surprise in cognitive science, multi-agent systems, and logic, focusing on those topics and debates that are most relevant for our current purposes. For more comprehensive overviews, the reader can fruitfully consult Macedo et al. (2009, 2012) and Reisenzein and Meyer (2009).

7.2.1 Cognitive Science

The emotion of surprise is probably of old phylogenetic origin (Reisenzein et al. 1996). This short-lived state of mind is caused in an agent when she encounters an event that she did not expect. Surprise comes in degrees of intensity, which depend monotonically on the degree of unexpectedness of the surprise-causing event (Stiensmeier-Pelster et al. 1995). Like most emotions, surprise has both phenomenal aspects (there is an experience of 'what it is like to be surprised'; Nagel 1974, Reisenzein 2000) and physical (behavioral/physiological) manifestations, such as a characteristic facial expression (raised eyebrows, opened mouth, etc.) and a decrease in heart rate (Miller 1973, Sokolov et al. 2002).

The cognitive-psychoevolutionary theory of surprise (Meyer et al. 1997) claims that, typically, an unexpected event elicits a sequence of four processes. First, the event is appraised as unexpected, i.e. as conflicting with a previously held belief.[2] Second, if the degree of unexpectedness is sufficiently large, then ongoing processes are interrupted and attention is shifted to the unexpected event. Third, the unexpected event is analyzed and evaluated, which can lead to the fourth process, viz. revision of the relevant beliefs.

The fact that this sequence ends in belief revision helps to explain the transitory (short-lived) character of surprise. When a surprising event occurs again and again, subjects tend to 'get used' to it, and after a few occurrences they do not find it surprising at all anymore (Charlesworth 1964, Experiment II). Initially, the surprising event is unexpected: it conflicts with a previously held belief B. This leads to a process of belief revision, which removes B from the agent's stock of beliefs (and perhaps replaces it with another belief). When the same event happens again, it is no longer surprising, because it no longer conflicts with a previously held belief (in particular, it does not conflict with B anymore).

The third step in the sequence of events triggered by an unexpected event involves *analyzing* that event. One of the features that is typically analyzed, is the event's cause: does it have a 'substantial' cause, or should it be attributed to 'mere chance'? Surprise thus leads to 'causal curiosity': it motivates the agent to inquire about the event's cause (Stiensmeier-Pelster et al. 1995, Meyer et al. 1997).[3] Charlesworth has compared the motivational power of *unexpected (surprising) data* (which conflict with a previously held belief), *expected data* (which are in full agreement with previously held beliefs) and *novel data* (about which

[2] Building upon earlier work on schema theory (Rumelhart 1984, Schank 1986), the cognitive-pyschoevolutionary theory uses the notion of 'schema' rather than 'belief'. This distinction is not relevant for our current purposes, so I will simply use the term 'belief'.

[3] The process of searching for an explanation of an observed event is widely known as *abduction*. Peirce, who coined this term, explicitly refers to surprising events when characterizing abductive reasoning (Peirce 1934, § 189).

the agent had no previous beliefs at all), and his experiments show that surprising data have the highest motivational power, i.e. they trigger further inquiry most frequently (Charlesworth 1964).

I just mentioned Charlesworth's distinction between *unexpected* and *novel* data. For an event to be unexpected, it really has to conflict with a previously held belief; if the agent did not have any beliefs about that (type of) event(s), then the event is not unexpected, but rather novel. The most common perspective is that surprise can only be generated by unexpected data, not by novel data (Charlesworth 1964, Stiensmeier-Pelster et al. 1995, Meyer et al. 1997).[4] However, some theorists maintain that an agent can also be surprised about events that she previously did not have any beliefs about. For example, Ortony and Partridge (1987) distinguish between *actively expected* events and *passively expected* or *assumed* events, and claim that surprise can arise from active expectation failure as well as assumption failure.[5] There is no real contradiction between both perspectives, since Ortony and Partridge maintain that in the case of surprise caused by assumption failure, the agent still has a belief, albeit a 'passive' one (an assumption). For example, if the legs of my chair suddenly break and I fall, I am surprised, not because I actively believed that I would remain seated in the chair, but because I passively expected (i.e. assumed) that the chair's legs are strong enough to support me. The tension between both perspectives can thus be resolved by postulating implicit beliefs with which the novel event conflicts—hence, although the event does not conflict with any explicit (active) beliefs, it *does* conflict with the postulated implicit (passive) belief.[6]

[4] This perspective is also common among philosophers. Davidson, for example, claims that "I could not be surprised [. . .] if I did not have beliefs in the first place. [...] Surprise requires that I be aware of a contrast between what I did believe and what I come to believe" (Davidson 1982, p. 326).

[5] Peirce, too, claims that surprise "has its Active and its Passive variety;—the former when what one perceives positively *conflicts* with expectation, the latter when having no positive expectation but only the absence of any suspicion of anything out of the common something quite unexpected occurs" (Peirce 1958, § 315).

[6] For example, novel events "can also be conceptualised as instances of expectancy disconfirmation: They disconfirm the *implicit*, schema-based belief that the unexpected event was unlikely to occur in the given situation." (Stiensmeier-Pelster et al. 1995, p. 6, my emphasis).

7.2.2 Multi-agent Systems

Since surprise typically leads to processes of learning and belief revision in humans, it is a natural move to endow *artificial agents* with the capability of feeling surprise, which can guide them in their actions. In a recent series of papers, Macedo and Cardoso have done exactly this (Macedo and Cardoso 2001a,b, 2004, Macedo et al. 2004, 2006). This work is based on the cognitive theories of surprise described in the previous subsection (Ortony and Partridge 1987, Meyer et al. 1997), and can thus also be seen as a simulation of the human surprise mechanism (with various simplifications, obviously).

The agent's goal is to explore an unknown and dynamic environment. The agent architecture is similar to the well-known BDI (belief-desire-intention) architecture (Wooldridge 2002), and looks as follows. The agent's *perceptual system* provides (partial) information about the environment, and stores it in *memory*. When new (hypothetical) information comes in, the agent's *surprise-generating module* calculates the intensity of the surprise caused by that piece of information. Finally, the *decision-making module* selects the agent's next action by considering, for every available action α, how surprised the agent would be by the state of the world caused by α, and then selecting the action that maximizes the agent's anticipated surprise. This module thus implements a utility-maximizing function, where the agent's utility is assumed to coincide with her anticipated surprise (more sophisticated architectures also take other emotions into account).

In the simplest model (Macedo and Cardoso 2001b), the anticipated intensity of surprise elicited by a piece of information φ is calculated as follows:[7]

$$S(\varphi) := 1 - P(\varphi). \tag{7.1}$$

The unexpectedness of φ is represented by $1 - P(\varphi)$. Here, $P(\varphi)$ denotes the subjective probability of φ, which is computed based on frequencies stored in the agent's memory. Thus (7.1) clearly shows that the intensity of surprise about φ is a monotone increasing function of the unexpectedness of φ (cf. supra).

[7] There exist more complex (and realistic) proposals for defining surprise in terms of unexpectedness (probability) (Macedo et al. 2004). However, the experimental data do not seem to single out one of these complex definitions over the other ones. Furthermore, the main conceptual points of this chapter (regarding the dynamic nature of surprise) can perfectly be made using (7.1). Therefore, I will stick to the simpler definition.

This work on surprise-based agent architectures fits in the broader field of emotion-based agent architectures (Bates 1994, Macedo et al. 2009, Faghihi et al. 2011). There are also proposals to incorporate the *dynamics* of emotion (El-Nasr 1998, Becker et al. 2004, Marsella and Gratch 2009), but none of them so far make use of the framework of (probabilistic) dynamic epistemic logic.

7.2.3 Logic

Lorini (2008) has argued that researchers attempting to incorporate surprise and other emotions into multi-agent systems can benefit from the accuracy of logical frameworks for the formal specification of emotions. Together with Castelfranchi, he has developed a logical framework for surprise (Lorini and Castelfranchi 2006, 2007). Just like Macedo and Cardoso's, this framework is based on the cognitive theories of surprise described in Subsection 7.2.1 (Ortony and Partridge 1987, Meyer et al. 1997), and can thus be seen as a formal-logical model of human surprise.

I will now discuss the main features of this framework.[8] The base logic is a system of probabilistic epistemic logic with a belief operator B and formulas about (linear combinations of) probabilities, such as $P(\varphi) \geq 0.5$ and $P(\varphi) + 2P(\psi) \geq 0.7$ (Fagin and Halpern 1994). This system is extended with PDL-style dynamic operators (Harel et al. 2000), and two unary operators $Test$ and $Datum$. The formulas $Test(\varphi)$ and $Datum(\varphi)$ are to be read as "the agent is currently scrutinizing φ" and "the agent has perceptual datum φ", respectively. Furthermore, there are actions $observe(\varphi)$ and $retrieve(\varphi)$, which represent observing that φ is the case and retrieving (from memory) that φ. Each of these actions gives rise to a PDL-style dynamic operator. The two most important axioms are:

$$[observe(\varphi)]Datum(\varphi), \tag{7.2}$$

$$[retrieve(\varphi)]Test(\varphi). \tag{7.3}$$

Axiom (7.2) says that after the agent observes that φ, this becomes a perceptual

[8] In this subsection in particular, I will not be able to do justice to all details of the framework under discussion. For example, I will only reason 'within' the logic, and not say anything about its formal semantics. Furthermore, Lorini and Castelfranchi define *two* types of surprise, mismatch-based surprise and astonishment, but I will only discuss the first one, because it is better suited to illustrate the main claims of the next section. (However, one might argue that the notion of surprise defined in Section 7.4 is actually closer to astonishment than to mismatch-based surprise.)

datum; analogously, axiom (7.3) says that after the agent has retrieved φ, this becomes an item under scrutiny.

With these resources, the notion of *mismatch-based surprise* can be defined. This emotion arises when there is a conflict between a perceptual datum ψ and a currently scrutinized belief φ; 'conflict' here means that the agent believes that φ and ψ cannot be jointly true. Furthermore, the *intensity* of a mismatch-based surprise is defined as the probability that the agent assigns to the scrutinized belief φ. Hence, the more confident the agent is in her belief that φ, the more intensely she will be surprised upon receiving a perceptual datum that conflicts with φ (this captures exactly the idea that the intensity of surprise is a monotone function of the degree of unexpectedness). Formally:

$$MismatchS(\psi,\varphi) :\equiv \quad Datum(\psi) \wedge Test(\varphi) \wedge B(\psi \to \neg\varphi), \qquad (7.4)$$

$$IntensityS(\psi,\varphi) = c :\equiv \quad MismatchS(\psi,\varphi) \wedge P(\varphi) = c. \qquad (7.5)$$

7.3 Surprise as a Dynamic Phenomenon

In this section, I will argue that neither Macedo and Cardoso's computational nor Lorini and Castelfranchi's logical models of surprise adequately capture the dynamic nature of surprise. Afterwards I will suggest how the dynamics of surprise *can* adequately be formalized.

7.3.1 Quasi-Static Analyses of Surprise

Let's first fix some terminology. Surprise is caused by an unexpected event. Any mental state (beliefs, desires, emotions, etc.) that the agent had (just) *before* perceiving the unexpected event will be called *'prior'*; any such state that she has (just) *after* perceiving the event will be called *'posterior'*.[9] A statement that involves only prior notions or only posterior notions will be called 'temporally coherent'; a statement that involves both prior and posterior notions will be called 'temporally incoherent'.

[9] This terminology is analogous to the use of 'priors' and 'posteriors' in Bayesian frameworks. However, it should be emphasized that in this chapter, 'prior' and 'posterior' are defined in terms of (being before or after) *perceiving* the unexpected event, while in Bayesian frameworks they are defined in terms of (being before or after) the *probabilistic update* ('probability revision') triggered by that event.

Consider Macedo and Cardoso's analysis of surprise, and recall their Definition (7.1) of surprise intensity as unexpectedness:

$$S(\varphi) = 1 - P(\varphi).$$

The left side contains a posterior notion: the intensity of the surprise felt by the agent after the unexpected event. The right side, however, contains a prior notion: the agent's subjective probability before the unexpected event. Hence, Definition (7.1) is a temporally incoherent statement.

To see this more clearly, note that there are two ways of reading (7.1) as a temporally coherent statement: (i) by considering both S and P to be prior notions, and (ii) by considering both S and P to be posterior notions. For interpretation (i), consider a case where the agent assigns a low (prior) probability to φ; Definition (7.1) then says that she should experience a highly intensive surprise about φ. Under interpretation (i), this surprise is prior; in other words, the agent is highly surprised about an event *before* she has even perceived it—which is clearly absurd. For interpretation (ii), consider a case where the agent is highly surprised after perceiving an occurrence of φ; Definition (7.1) then says that she assigns a low probability to φ. Under interpretation (ii), this probability is posterior; in other words, even after the agent has observed an occurrence of φ, she still assigns a low probability to it—which clearly contradicts the common assumption that agents process new information via Bayesian updating.[10]

I now turn to Lorini and Castelfranchi's analysis of surprise. Let's first consider the qualitative notion of mismatch-based surprise—ignoring, for the moment, surprise intensity. Recall their Definition (7.4):

$$MismatchS(\psi, \varphi) \equiv Datum(\psi) \wedge Test(\varphi) \wedge B(\psi \to \neg\varphi).$$

The left side contains a posterior notion: the agent's mismatch-based surprise after the unexpected event. The right side is more complicated. The first conjunct is posterior: ψ is only a perceptual datum after it has been observed by the agent; this dynamics was explicitly represented in (7.2). The second conjunct is both prior *and* posterior: φ was under scrutiny before the observation of the unexpected event, and remains so afterwards. The third and final conjunct is prior: the agent believed that ψ and φ cannot be jointly true before the unexpected event; typically, she will drop this belief as a result of her surprise

[10] And $P(\varphi|\varphi) = 1$, so after the occurrence of φ, the agent should assign probability 1 to it.

(recall from Subsection 7.2.1 that surprise typically leads to a process of belief revision). Thus, in total, Definition (7.4) is temporally incoherent.[11]

Finally, let's consider the quantitative aspects of Lorini and Castelfranchi's system. Recall their Definition (7.5) of surprise intensity:

$$IntensityS(\psi, \varphi) = c \equiv MismatchS(\psi, \varphi) \wedge P(\varphi) = c.$$

The left side contains a posterior notion: the intensity of the agent's mismatch-based surprise after she has perceived the unexpected event. The right side is, again, more complicated. The first conjunct—which was also the left side of (7.4)—is posterior: the agent experiences mismatch-based surprise only after perceiving the unexpected event. The second conjunct, however, involves a prior notion, viz. the probability that the agent assigns to the scrutinized item φ before perceiving the unexpected event. Hence, Definition (7.5) is temporally incoherent as well.[12]

An intuitively correct principle about surprise should look somewhat like this: if the agent has a (prior) belief that ψ and φ are incompatible, and assigns (prior) probability c to φ, then after retrieving φ and observing an occurrence of ψ, she will experience a (posterior) mismatch-based surprise with intensity c. Formally, this looks as follows:[13]

$$\begin{aligned}&\big(B(\psi \rightarrow \neg\varphi) \wedge P(\varphi) = c\big) \longrightarrow \\ &\qquad [retrieve(\varphi); observe(\psi)]IntensityS(\psi, \varphi) = c.\end{aligned} \tag{7.6}$$

However, to derive (7.6) in Castelfranchi and Lorini's system, one needs principles such as (7.7) and (7.8), which link the agent's prior and posterior states by

[11] Again, there are two ways of reading (7.4) as a temporally coherent statement: by considering all notions that appear in it to be prior, or by considering all those notions to be posterior. It is easy to see, however, that both interpretations quickly lead to counterintuitive consequences. Similar remarks apply to (7.5), which will be discussed next.

[12] It should be emphasized that the assessment of Lorini and Castelfranchi's analysis as temporally incoherent is only valid on the assumption that the terms 'prior' and 'posterior' are defined relative to the moment of *perceiving* the unexpected event, as specified at the beginning of this subsection (also recall Footnote 9). In particular, if these terms are defined relative to the moment of *recognizing the mismatch* between the datum and the scrutinized expectation—which is the viewpoint taken by Lorini and Castelfranchi themselves—, then this analysis *is* temporally coherent.

[13] As usual, ';' denotes ordinary PDL sequential composition (Harel et al. 2000); this operation on actions is allowed in Lorini and Castelfranchi's system.

claiming that her observation of the occurrence of ψ does not change her relevant beliefs and probabilities in any way. This is highly counterintuitive: both (7.7) and (7.8) go entirely against the idea that surprise triggers a process of belief revision. Additionally, (7.8) clearly contradicts the common assumption that agents process new information via Bayesian conditionalization.

$$B(\psi \to \neg\varphi) \quad \to \quad [observe(\psi)]B(\psi \to \neg\varphi), \tag{7.7}$$

$$P(\varphi) = c \quad \to \quad [observe(\psi)]P(\varphi) = c. \tag{7.8}$$

7.3.2 Towards a Fully Dynamic Analysis of Surprise

I have shown that both Macedo and Cardoso's definition of surprise intensity (7.1) and Lorini and Castelfranchi's definitions of mismatch-based surprise and its intensity (7.4–7.5) are temporally incoherent (but recall Footnote 12). There is a uniform explanation for this: surprise is an essentially dynamic phenomenon, but none of these authors explicitly represents this dynamics, so they have to 'smuggle' it into their systems—which thus end up being temporally incoherent.[14]

Before moving on, we need to clarify the relationship between the systems discussed above and the system that will be developed in this chapter. The problem of temporal incoherence is situated at the *conceptual*, rather than at the *empirical* level, and is therefore largely independent of the original motivations behind the systems discussed above. For example, Lorini and Castelfranchi's goal is first and foremost to propose a cognitively realistic model of surprise; although their analysis is largely static, it certainly achieves its main goal, since it is highly successful at capturing various experimentally observed properties of surprise. In contrast, the main motivation behind this chapter is to propose a temporally coherent model of surprise (using the framework of dynamic epistemic logic). Looking ahead, this means that the major advantage of the new account of surprise that will be developed in Section 7.4 will not be so much its level of empirical adequacy—which, I will argue, is more or less comparable to that of the other accounts—, but rather the fact that it is temporally coherent, and thus better able to capture the dynamic nature of surprise.

Now that this methodological issue has been clarified, we are ready to move on. To obtain a temporally coherent definition of surprise, which respects the dif-

[14] This analysis is highly similar to that of Aumann's agreement theorem in Chapter 5; in particular, see Subection 5.6.1.

ferent 'stages' (before vs. after perceiving the unexpected event), the dynamics of surprise needs to be represented explicitly. I will use a public announcement operator $[!\varphi]$ for this purpose (technical details will be discussed in the next section). Whether a certain notion is to be interpreted as prior or as posterior, is now encoded directly in the syntax of the language: if the notion is within the scope of a dynamic operator, it is posterior, otherwise it is prior. For example, $P(\varphi) = 0.2$ means that the agent's *prior* probability of φ is 0.2, while $[!\varphi]P(\varphi) = 0.2$ means that her *posterior* probability of φ is 0.2.

We will work with a simple measure of surprise intensity S, based on Macedo and Cardoso's (7.1).[15] When the surprise dynamics is explicitly represented, (7.1) is transformed into the following:

$$[!\varphi]S(\varphi) = c \longleftrightarrow P(\varphi) = 1 - c. \tag{7.9}$$

This principle says that the agent will be surprised about φ with intensity c after the unexpected event iff she assigns probability $1 - c$ to φ before the unexpected event. It thus says exactly the same as (7.1), but now in a temporally coherent way: both sides of (7.9) are prior statements.[16] Furthermore, note that the right-to-left direction of (7.9) is similar in spirit to (7.6), which was very intuitive, but which was only derivable using additional implausible principles such as (7.7–7.8).

7.4 Modeling Surprise in Probabilistic DEL

In the previous section, I made some suggestions on how the dynamics of surprise can be represented explicitly. In this section, these suggestions will be developed into a full-fledged logical system. I will also show how this system can naturally capture several important properties of surprise, and how it can be used to define a qualitative notion of surprise.

[15] Recall Footnote 7.

[16] The left formula *as a whole* is prior; the *subformula* $S(\varphi) = c$ occurs inside the scope of the $[!\varphi]$-operator, and is thus posterior. In other words, principle (7.9) is able to express a connection between the agent's *prior* probability and her *posterior* surprise intensity in a temporally coherent way, by making use of the dynamic $[!\varphi]$-operator.

7.4.1 Semantic Setup

Given the dynamic nature of surprise, and its connection with epistemic states and processes (beliefs, unexpectedness, belief revision, etc.), it is natural to work in the general framework of dynamic epistemic logic. This framework is rapidly becoming a 'lingua franca' or 'universal toolbox', which has been applied to problems in game theory, philosophy, artificial intelligence, etc. (Fagin and Halpern 1994, Kooi 2003, van Ditmarsch et al. 2007).

As usual, we fix a countable set Prop of proposition letters. In this chapter, we will only work with a single agent, so it is not necessary to introduce agent indices. The formal language $\mathcal{L}$ is given by the following BNF:

$$\varphi \ ::= \ p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K\varphi \mid \sum_{i=1}^{n} a_i P(\varphi) \geq b \mid S(\varphi) \geq b \mid S(\varphi) \leq b \mid [!\varphi]\varphi$$

where $p \in \mathsf{Prop}$, $1 \leq n < \omega$, and $a_1, \dots, a_n, b \in \mathbb{Q}$. The informal interpretation of this language has largely been discussed in Chapter 3. The only new formulas are those of the form $S(\varphi) \geq b$ (resp. $S(\varphi) \leq b$), which says that the agent is surprised about φ with intensity at least b (resp. at most b). Recall that for probability formulas, only the $\geq$-form is taken as primitive, and the $\leq$-form is then defined in terms of it, via multiplication with -1 (see p. 66). There is also an alternative definition, which relies on the probabilistic law that $P(\neg\varphi) = 1 - P(\varphi)$ (see p. 68). However, for surprise formulas, we do *not* have linear combinations, *nor* an analogous law that links surprise about φ and surprise about $\neg\varphi$,[17] and therefore the $\geq$- and $\leq$-forms both have to be taken as primitive. One can then define $S(\varphi) < b$ as $\neg(S(\varphi) \geq b)$, etc.

Public announcement is usually explicated in terms of rational communication, but actually, almost any public event can be modeled using public announcements; for example, a strike of lightning can be modeled as a public announcement of the proposition 'lightning occurs (at time t and location ℓ)'.[18] It thus makes perfect sense to represent an unexpected event (whatever its exact

[17] This will be further clarified in Footnote 20 and Lemma 7.4.

[18] Van Benthem, Gerbrandy and Kooi make a similar comment: "While much of the theory has been developed with conversation and communication in mind, it is important [...] to stress that we are not doing some sort of formal linguistics. The formal systems we will be dealing with apply just as well to observation, experimentation, learning, or any sort of information-carrying scenario." (van Benthem et al. 2009, p. 71).

nature) as a public announcement.[19]

I now introduce the models on which this language will be interpreted:

Definition 7.1. A *surprise model* is a tuple $\mathbb{M} := \langle W, R, \mu, \sigma, V\rangle$, where W is a non-empty finite set of states, R is an equivalence relation on W, and $V\colon Prop \to \wp(W)$ is a valuation function. Furthermore, μ assigns to every state $w \in W$ a probability mass function $\mu(w)\colon W \to [0,1]$ that satisfies the following two conditions:

- $\mu(w)(w) > 0$,
- $\mu(w)(v) = 0$ for all $v \in W$ such that $(w, v) \notin R$.

Finally, σ assigns to every state $w \in W$ a surprise measure, i.e. a partial function $\sigma(w)\colon \wp(W) \rightharpoonup [0,1]$.

Definition 7.2. The class of all surprise models will be denoted $\mathcal{C}_S$. Furthermore, $\mathcal{C}_S^*$ is the class of all surprise models whose surprise measures are entirely undefined, i.e. such that $\sigma(w)(X)$ is undefined for all $w \in W$ and $X \subseteq W$.

A surprise model is thus just a (single-agent) well-behaved probabilistic Kripke model $\langle W, R, \mu, V\rangle$ (see Definitions 3.2 and 3.4 on p. 75 and 81 for details), with an additional component σ. Recall that $\mu(w)(v) = c$ means that at state w, the agent assigns probability c to v being the actual state. Similarly, $\sigma(w)(X) = c$ means that at state w, the agent experiences surprise with intensity c about X (i.e. about one of the states in X being the actual state). Note the following differences between $\mu(w)$ and $\sigma(w)$ (for any state $w \in W$):

- $\mu(w)$ is a total function, so $\mu(w)(v)$ is defined for every state $v \in W$ (this simplifying assumption was discussed after Definition 3.2 on p. 75); in contrast, $\sigma(w)$ is a partial function, so it is allowed that $\sigma(w)(X)$ is undefined for some sets $X \subseteq W$,
- $\mu(w)$ is required to satisfy the liveness and consistency conditions, whose motivations were discussed extensively after Definition 3.3 and Lemma 3.1

[19] This also resolves a terminological tension in the literature on surprise. Agents are surprised *about* some *propositional* content (a piece of information), but their surprise is *caused by* some (non-propositional) *event*. In the new system, the propositional content of the surprise is formalized as the proposition φ, while its cause is formalized as the *public announcement of* that proposition. In short: φ is a proposition, but $!\varphi$ is an event.

on p. 79ff.; in contrast, $\sigma(w)$ is not required to satisfy any additional conditions whatsoever,

- $\mu(w)$ is defined on individual states, and can additively be lifted to sets of states: $\mu(w)(X) = \sum_{x \in X} \mu(w)(x)$ (this essentially reflects the finite additivity of probabilities); in contrast, $\sigma(w)$ is defined directly on sets of states, so it might happen that $\sigma(w)(\{x, y\}) \neq \sigma(w)(\{x\}) + \sigma(w)(\{y\})$.

These differences show that unlike the well-behaved epistemological notion of probability (degree of belief), the psychological notion of (degree of) surprise satisfies no static regularities whatsoever. This is a clear manifestion of the essentially dynamic nature of surprise in the definition of surprise models.[20]

I now turn to the logic's semantics. This is entirely as expected, so Definition 7.3 only provides the formal clause for surprise formulas. This clause holds for $\gtreqless \in \{\geq, \leq\}$; I will return to it later (see Lemma 7.2). As usual, interpreting a formula of the form $[!\varphi]\psi$ on a surprise model $\mathbb{M}$ requires that the subformula ψ be interpreted on the updated model $\mathbb{M}|\varphi$, which is well-defined because of Definition 7.4 and Lemma 7.1. Finally, Definition 7.5 states the usual definition of semantic validity.

Definition 7.3. Consider a surprise model $\mathbb{M}$, a state w in $\mathbb{M}$, and $\varphi \in \mathcal{L}$. Then:

$$\mathbb{M}, w \models S(\varphi) \gtreqless c \qquad \text{iff} \qquad \begin{cases} \sigma(w)([\![\varphi]\!]^{\mathbb{M}}) \gtreqless c & \text{if } \sigma(w)([\![\varphi]\!]^{\mathbb{M}}) \text{ is defined,} \\ c = 0 & \text{otherwise.} \end{cases}$$

Definition 7.4. Consider an arbitrary surprise model $\mathbb{M} = \langle W, R, \mu, \sigma, V \rangle$ and formula $\varphi \in \mathcal{L}$, and suppose that $\mathbb{M}, w \models \varphi$ for some $w \in W$. Then the *updated model* $\mathbb{M}|\varphi := \langle W^\varphi, R^\varphi, \mu^\varphi, \sigma^\varphi, V^\varphi \rangle$ is defined as follows:

- $W^\varphi := [\![\varphi]\!]^{\mathbb{M}} = \{w \in W \mid \mathbb{M}, w \models \varphi\}$,
- $R^\varphi := R \cap ([\![\varphi]\!]^{\mathbb{M}} \times [\![\varphi]\!]^{\mathbb{M}})$,

[20] One might consider adding the requirements that if $X \subseteq Y \subseteq W$, then $\sigma(w)(X) \geq \sigma(w)(Y)$ and $\sigma(w)(W - X) = 1 - \sigma(w)(X)$, in analogy to the well-known Kolmogorov axioms for probability. However, the only motivation for such requirements seems to be the observation that 'surprise is inversely correlated with probability', which is only plausible on the temporally incoherent reading in which 'surprise' is taken to be a posterior notion and 'probability' a prior notion. I will return to this suggestion after the dynamics has been formally introduced (see Lemma 7.4).

- $\mu^{\varphi}(w)(v) := \frac{\mu(w)(v)}{\mu(w)([\![\varphi]\!]^{\mathbb{M}})}$ for all $w, v \in W^{\varphi}$,
- $\sigma^{\varphi}(w)(X) := 1 - \mu(w)(X)$ for all $w \in W^{\varphi}, X \subseteq W^{\varphi}$,
- $V^{\varphi}(p) := V(p) \cap [\![\varphi]\!]^{\mathbb{M}}$ for all $p \in \mathsf{Prop}$.

Definition 7.5. For any formula $\varphi \in \mathcal{L}$ and class of models $\mathcal{C}$, we say that $\mathcal{C} \models \varphi$ iff $\mathbb{M}, w \models \varphi$ for all models $\mathbb{M} \in \mathcal{C}$ and states w in $\mathbb{M}$.

Lemma 7.1. *The class $\mathcal{C}_S$ is closed under public announcements, i.e. if $\mathbb{M} \in \mathcal{C}_S$, then also $\mathbb{M}|\varphi \in \mathcal{C}_S$ (for any formula $\varphi \in \mathcal{L}$). This does* not *hold for $\mathcal{C}_S^*$.*

Proof. The $\mathcal{C}_S$ case is trivial: for the non-surprise components, see Lemma 3.3 on p. 86, and since Definition 7.1 does not require the surprise measures to satisfy any additional requirements, there is nothing else to prove. For $\mathcal{C}_S^*$, note that by Definition 7.4, the updated surprise measures are total functions, even if the original surprise measures were entirely undefined. □

Definition 7.4 is a special case of public announcements in well-behaved probabilistic Kripke models (see Definition 3.6 on p. 86), so I will not comment upon it, except for re-emphasizing that the probability functions are changed by Bayesian conditionalization on the announced proposition φ. More importantly, note that the updated surprise measure $\sigma^{\varphi}(w)$ is defined in terms of the original probability function $\mu(w)$. This is the only substantial property of surprise that is assumed in the logic's semantic setup; it is clearly of a dynamic nature (linking the original and the updated model).

Even though the surprise measures $\sigma(w)$ are allowed to be partial, Lemma 7.2 below shows that this does not lead to any truth value gaps in the semantics. When we are modeling concrete scenarios, we typically want to assume that the agent initially (i.e. before any unexpected events have taken place) experiences no surprise. Lemma 7.2 therefore justifies the following heuristic rule (HEUR) :

> When modeling a scenario, it can always be assumed that the 'initial' model $\mathbb{M}$ (which represents the situation before any unexpected events have taken place) leaves all surprise measures undefined, or formally: $\mathbb{M} \in \mathcal{C}_S^*$.

Lemma 7.2. *Consider an arbitrary surprise model $\mathbb{M} = \langle W, R, \mu, \sigma, V\rangle$ and formula $\varphi \in \mathcal{L}$, and suppose that $\sigma(w)([\![\varphi]\!]^{\mathbb{M}})$ is undefined. Then $\mathbb{M}, w \models S(\varphi) = 0$.*

Proof. Since $\sigma(w)(\llbracket \varphi \rrbracket^{\mathbb{M}})$ is undefined, it follows by the semantic clause for $S(\varphi) \geq c$ that $\mathbb{M}, w \models S(\varphi) \geq 0$ (and $\mathbb{M}, w \not\models S(\varphi) \geq c$ for all $c \neq 0$). Entirely analogously, $\mathbb{M}, w \models S(\varphi) \leq 0$ (and $\mathbb{M}, w \not\models S(\varphi) \leq c$ for all $c \neq 0$). □

The following lemma states that the language $\mathcal{L}$ contains no redundancies. In particular, the surprise operator cannot be defined in terms of the other available operators.

Lemma 7.3. *There exists no formula $\varphi \in \mathcal{L} - \{S\}$ such that*

$$\mathcal{C}_S \models \varphi \leftrightarrow S(p) \geq 0.5.$$

Proof. Consider the surprise models $\mathbb{M}_1$ and $\mathbb{M}_2$, defined as follows:

- $\mathbb{M}_1 = \langle W_1, R_1, \mu_1, \sigma_1, V_1 \rangle, W_1 = \{w_1\}, R_1 = \{(w_1, w_1)\}, \mu(w_1)(w_1) = 1, \sigma_1(w_1)(X) = 0.6$ for all $X \subseteq W_1$, and $V_1(p) = W_1$,
- $\mathbb{M}_2 = \langle W_2, R_2, \mu_2, \sigma_2, V_2 \rangle, W_2 = \{w_2\}, R_2 = \{(w_2, w_2)\}, \mu_2(w_2)(w_2) = 1, \sigma_2(w_2)(X) = 0.4$ for all $X \subseteq W_2$, and $V_2(p) = W_2$.

One can show by induction on the complexity of φ that

$$\text{for all } \varphi \in \mathcal{L} - \{S\}\colon \mathbb{M}_1, w_1 \models \varphi \text{ iff } \mathbb{M}_2, w_2 \models \varphi.$$

But it also holds that $\mathbb{M}_1, w_1 \models S(p) \geq 0.5$, while $\mathbb{M}_2, w_2 \not\models S(p) \geq 0.5$. □

The distinction between the original and the updated model corresponds exactly to the distinction between prior and posterior notions that was introduced in the previous section. Hence, the definition $\sigma^{\varphi}(w)(X) = 1 - \mu(w)(X)$ defines *posterior* surprise in terms of *prior* probability. As a consequence, all the properties of probability are manifested in the posterior surprise measure (recall Footnote 20):

Lemma 7.4. *Consider an arbitrary surprise model $\mathbb{M} = \langle W, R, \mu, \sigma, V \rangle$ and formula $\varphi \in \mathcal{L}$, and suppose that $\mathbb{M}, w \models \varphi$ for some $w \in W$. For all $w \in W^{\varphi}$ and $X \subseteq Y \subseteq W^{\varphi}$, it holds that $\sigma^{\varphi}(w)(X) \geq \sigma^{\varphi}(w)(Y)$ and that $\sigma^{\varphi}(w)(W - X) = 1 - \sigma^{\varphi}(w)(X)$.*

Proof. Both items follow from the definition of σ^{φ} and the fact that $\mu(w)$ is a probability function. For example, if $X \subseteq Y$, then $\mu(w)(X) \leq \mu(w)(Y)$, and hence $\sigma^{\varphi}(w)(X) = 1 - \mu(w)(X) \geq 1 - \mu(w)(Y) = \sigma^{\varphi}(w)(Y)$. □

Before moving to the logic's proof theory, I will illustrate and justify its semantics by discussing a simple example in full detail.

Example 7.1. Consider the following scenario. Mary does not know whether it is currently snowing. In fact, it is indeed currently snowing, but since Mary does not yet know about this, she experiences no surprise about it whatsoever. Furthermore, since it is July and Mary knows that snow in July is very rare at her current location, she considers it very unlikely that it is currently snowing. This example can be formalized using the following surprise model: $\mathbb{M} = \langle W, R, \mu, \sigma, V \rangle, W = \{w, v\}, R = W \times W, \mu(w)(w) = \mu(v)(w) = 0.05, \mu(w)(v) = \mu(v)(v) = 0.95$, $V(p) = \{w\}$, and $\sigma(w)(X)$ and $\sigma(v)(X)$ undefined for all $X \subseteq W$. (Note that we have followed the heuristic rule HEUR discussed above.) The proposition letter p represents 'it is snowing'; the state w represents the actual world. This model is a faithful representation of the scenario described above; for example:

$$\mathbb{M}, w \models \neg Kp \wedge \neg K\neg p \wedge P(p) = 0.05 \wedge P(\neg p) = 0.95 \wedge S(p) = 0.$$

Now suppose that Mary goes outside and sees that it is actually snowing. This can be modeled as a public announcement of p (recall Footnote 18). Applying Definition 7.4, we obtain the updated model $\mathbb{M}|p$, with $W^p = \{w\}, R = \{(w, w)\}$,

$$\mu^p(w)([\![p]\!]^{\mathbb{M}|p}) = \mu^p(w)(w) = \frac{\mu(w)(w)}{\mu(w)([\![p]\!]^{\mathbb{M}})} = \frac{\mu(w)(w)}{\mu(w)(w)} = 1,$$

$$\sigma^p(w)([\![p]\!]^{\mathbb{M}|p}) = \sigma^p(w)(\{w\}) = 1 - \mu(w)(\{w\}) = 1 - 0.05 = 0.95.$$

Using this updated model $\mathbb{M}|p$, we find that

$$\mathbb{M}, w \models [!p]\big(Kp \wedge P(p) = 1 \wedge P(\neg p) = 0 \wedge S(p) = 0.95\big).$$

So after going outside, Mary comes to know that it is in fact snowing. She also adjusts her probabilities: she is now certain that it is snowing, i.e. she assigns probability 1 to p being true and probability 0 to p being false. These are the main *cognitive* effects of Mary's observation that it is snowing. However, on the *emotional* side, she is also highly surprised to find out that it is snowing, because she initially considered this highly unlikely. These are the results that one would intuitively expect, so the semantic setup introduced above seems to yield an adequate representation of (the interactions between) the cognitive (epistemic and probabilistic) and emotional (surprise) effects of a public announcement.

7.4.2 Axiomatization

I now turn to the logic's proof theory. We can make use of the well-known strategy of adding reduction axioms to a static base logic (see Subsection 3.3.2 for details). The reduction axioms for all operators of $\mathcal{L} - \{S\}$ are well-known; see items 1–5 of Definition 7.6 below. What about reduction axioms for S? Recall that in Subsection 7.3.2 I suggested a dynamified (and temporally coherent!) version (7.9) of Macedo and Cardoso's original (7.1). With only minor modifications,[21] this suggestion can be turned into reduction axioms for S; see items 6–7 below.

Definition 7.6. The reduction axioms for public announcement:

$$\begin{array}{lrcll}
1. & [!\varphi]p & \longleftrightarrow & \varphi \to p, & (\text{for } p \in \mathsf{Prop}) \\
2. & [!\varphi]\neg\psi & \longleftrightarrow & \varphi \to \neg[!\varphi]\psi, & \\
3. & [!\varphi](\psi_1 \wedge \psi_2) & \longleftrightarrow & [!\varphi]\psi_1 \wedge [!\varphi]\psi_2, & \\
4. & [!\varphi]K\psi & \longleftrightarrow & \varphi \to K[!\varphi]\psi, & \\
5. & [!\varphi]\sum c_i P(\psi_i) \geq c & \longleftrightarrow & \varphi \to \sum c_i(\langle !\varphi\rangle\psi) \geq cP(\varphi), & \\
6. & [!\varphi]S(\psi) \geq c & \longleftrightarrow & \varphi \to P(\langle !\varphi\rangle\psi) \leq 1 - c, & \\
7. & [!\varphi]S(\psi) \leq c & \longleftrightarrow & \varphi \to P(\langle !\varphi\rangle\psi) \geq 1 - c. &
\end{array}$$

We are now ready to provide an axiomatization. As is shown in Figure 7.1, the logic SURPRISE can be axiomatized in a highly modular fashion. The first five components are exactly the same as their namesakes in the componentwise axiomatization of EPEL (see Figure 5.1 on p. 158). The reduction axioms component consists of the reduction axioms stated in Definition 7.6. Finally, the surprise component comprises some static axioms and rules for the surprise operator S:

- $S(\varphi) \geq 0$,
- $S(\varphi) > 0 \to S(\varphi) \leq 1$,
- $\neg\big(S(\varphi) \leq k \wedge S(\varphi) \geq k'\big)$ for all $k < k'$,

[21] Trivial modifications are that the statement about = needs to be 'split out' into statements about $\leq$ and $\geq$, and that in the reduction axioms the argument of S should be an arbitrary formula ψ, and not just φ itself. A more serious modification is that the right sides of the reduction axioms should not contain simply $P(\psi)$, but rather $P(\langle !\varphi\rangle\psi)$, to 'pre-encode' the effect of the public announcement of φ on ψ.

Figure 7.1: Componentwise axiomatization of SURPRISE

1. the propositional component
2. the individual knowledge component
3. the linear inequalities component
4. the probabilistic component
5. the well-behavedness component
6. the surprise component
7. the reduction axioms component

- $S(\varphi) > 0 \to (S(\varphi) \geq k \vee S(\varphi) \leq k)$,
- if $\vdash \varphi \leftrightarrow \psi$ then $\vdash S(\varphi) \gtrless c \leftrightarrow S(\psi) \gtrless c$ (for $\gtrless \in \{\geq, \leq\}$).

Note that these static axioms for surprise are all concerned with the technical details of this particular formalization of surprise, rather than with any substantial properties of surprise itself. The only substantial axioms for surprise are thus its reduction axioms (items 6–7 of Definition 7.6), which together constitute a dynamified version of Macedo and Cardoso's original definition (7.1). I take this to be a clear manifestion of the essentially dynamic nature of surprise in the axiomatization of the logic.

I will finish this subsection by showing that the logic's semantics and axiomatization are in perfect harmony: the axiom system is sound and complete with respect to the semantics.

Theorem 7.1. SURPRISE *is (weakly) sound and complete with respect to* $\mathcal{C}_S$.

Proof. As usual, soundness is proved by induction on derivation length. It is easy to check that all axioms of SURPRISE are semantically valid on $\mathcal{C}_S$, and that all of its rules are $\mathcal{C}_S$-validity-preserving.

Completeness can also be proved using standard techniques. First of all, because the reduction axioms allow us to rewrite any formula as an equivalent formula without any dynamic operators, it suffices to prove completeness for the static fragment of the logic. This is done using a filtration of a canonical model over a set of formulas Σ which is finite and closed under subformulas. These methods are well-known in probabilistic epistemic logic (Fagin and Halpern 1994), so I will only discuss the surprise component.

The following can easily be proved for maximally consistent sets $\Gamma \subseteq \Sigma$:

- for all $\chi \in \Sigma$, there exists a number $\alpha_{\Gamma,\chi} \in [0,1] \cap \mathbb{Q}$ such that the formula $S(\chi) = \alpha_{\Gamma,\chi}$ is consistent with Γ,
- for all $\chi, \chi' \in \Sigma$, if $\vdash \chi \leftrightarrow \chi'$, one can always choose $\alpha_{\Gamma,\chi} = \alpha_{\Gamma,\chi'}$.

The canonical model $\mathbb{M}^c$ has states $W^c = \{\Gamma \subseteq \Sigma \mid \Gamma \text{ is maximally consistent}\}$; its surprise function is defined as follows: for all $\Gamma \in W^c$ and $X \subseteq W^c$, put

$$\sigma^c(\Gamma)(X) := \begin{cases} \alpha_{\Gamma,\chi} & \text{if } \exists \chi \in \Sigma : X = \{\Delta \in W^c \mid \chi \in \Delta\}, \\ 0 & \text{otherwise.} \end{cases}$$

The truth lemma can now easily be extended to the case of surprise formulas. Suppose, for example, that the formula $S(\chi) \geq c$ belongs to Σ; then χ itself also belongs to Σ, and by the definition of σ^c, showing that $\mathbb{M}^c, \Gamma \models S(\chi) \geq c$ iff $S(\chi) \geq c \in \Gamma$ boils down to showing that $\alpha_{\Gamma,\chi} \geq c$ iff $S(\chi) \geq c \in \Gamma$. The latter follows from the fact that the formula $S(\chi) = \alpha_{\Gamma,\chi}$ is consistent with Γ. □

Corollary 7.1. SURPRISE *has the finite model property.*

Proof. Trivial, since surprise models are, by definition, finite. □

7.4.3 Some Interesting Modeling Results

I will now show that the logical system developed above is able to capture several properties of surprise. However, there is one technical caveat. Recall that φ can only be publicly announced if φ is true *before* the announcement. It is natural to assume that φ will still be true *after* the announcement. However, because public announcements take into account higher-order information, it might happen that φ, simply by being announced, becomes false. A typical example is $\varphi = p \wedge \neg Kp$.[22] If no such 'self-falsifying' effects occur, φ is called successful:

Definition 7.7. A formula $\varphi \in \mathcal{L}$ is called *successful* iff $\mathcal{C}_S \models [!\varphi]\varphi$.

When modeling 'real-life' scenarios in a single-agent setting, formulas typically do not involve higher-order information,[23] so at least from this modeling

[22] See Subsection 3.3.3 for a more detailed discussion.

[23] In a single-agent setting one is typically surprised about 'facts of nature', not about one's *epistemic attitudes about* such facts. In a multi-agent setting, however, it would be natural to have scenarios like "Alice was surprised when finding out that *Bob knows that* φ".

perspective, the assumption of successfulness in many of the propositions below is quite harmless.[24] I now turn to the first concrete result.

Proposition 7.1. *The following formula is satisfiable:*

$$\begin{array}{lrlll} \varphi \wedge & \neg K\varphi & \wedge\ P(\varphi)=0.2 & \wedge\ S(\varphi)=0 \\ \wedge & \langle !\varphi\rangle\big(K\varphi & \wedge\ P(\varphi)=1 & \wedge\ S(\varphi)=0.8\big) \\ \wedge & \langle !\varphi\rangle\langle !\varphi\rangle\big(K\varphi & \wedge\ P(\varphi)=1 & \wedge\ S(\varphi)=0\big). \end{array}$$

Proof. Consider $\mathbb{M} := \langle W, R, \sigma, \mu, V\rangle$, with $W = \{w, v\}$, $R = W \times W$, $V(p) = \{w\}$, $\mu(w)(w) = 0.2$, $\mu(w)(v) = 0.8$ and $\sigma(w)(X)$ and $\sigma(v)(X)$ undefined for all $X \subseteq w$ (all components which have not been mentioned are irrelevant, and can thus be assigned values at random). One can easily check that this is indeed a surprise model, and that the formula stated above (with φ instantiated to p) is indeed true at $\mathbb{M}, w$. Finally, note that $\mathbb{M} \in \mathcal{C}_S^*$, i.e. we have followed the heuristic rule HEUR. □

Proposition 7.1 shows that the logic is capable of doing what it was designed to do, viz. explicitly representing surprise dynamics. It describes the following scenario. Initially, φ is true, but the agent does not know this. Furthermore, she assigns rather low prior probability to it (and thus does not expect its announcement). However, because she does not yet know that φ is actually true, she experiences no surprise about it whatsoever. Next, the unexpected announcement of φ occurs, and three things happen: (i) the agent comes to know that φ, (ii) she processes this new information by Bayesian conditionalization and thus assigns probability 1 to it, and (iii) she experiences a very high degree of surprise about φ (inversely correlated to the low probability that she initially assigned to it). After another announcement of φ, the agent's knowledge and probabilities are not changed; however, because this second announcement was no longer unexpected (after all, in the meanwhile she has come to know that φ), her surprise about φ drops again to 0. The formula in Proposition 7.1 captures this scenario in a very natural way, using nested public announcement operators to explicitly represent the successive layers of surprise dynamics.

[24] Next to the 'standard' unsuccessful formulas involving knowledge ($p \wedge \neg Kp$, van Ditmarsch et al. 2007) and probability ($p \wedge P(p) < 1$, Kooi 2003), one can also define unsuccessful formulas involving the surprise operator S, e.g. $P(p) > 0 \wedge S(p) \geq 1$. Clearly, these formulas all have the same underlying syntactic structure. For technical results on the syntactical characterization of successful and unsuccessful formulas, see Holliday and Icard III (2010).

At this point, it should be pointed out that not all scenarios described by satisfiable formulas are equally plausible. In particular, it is easy to check that formulas of the form $S(\varphi) > 0 \wedge P(\varphi) < 1$ are satisfiable, although the scenario described by such formulas sounds highly counterintuitive. The first conjunct says that the agent experiences some surprise about φ, which normally only happens after a public announcement of φ; however, this announcement should also have led the agent to become certain about φ (i.e. to assign probability 1 to it), which contradicts the second conjunct. To rule out such scenarios, one might consider adding an axiom of the form $S(\varphi) > 0 \rightarrow P(\varphi) = 1$ to the SURPRISE system. However, this ignores the fact that there exist formulas φ to which the agent does *not* assign probability 1 after they have been announced.[25] Furthermore, even if such formulas are excluded—for example, by only considering (Boolean combinations of) propositional atoms—, then it is still the case that $\mathcal{C}_S \not\models S(p) > 0 \rightarrow P(p) = 1$. However, it *does* hold that $\mathcal{C}_S^* \models S(p) > 0 \rightarrow P(p) = 1$. In other words, even though the formula $S(p) > 0 \wedge P(p) < 1$ is satisfiable in a $\mathcal{C}_S$-model, it is *not* satisfiable in a $\mathcal{C}_S^*$-model. Hence, when we are modeling concrete scenarios (and following the heuristic rule HEUR), the entire problem does not arise.

We now turn to Proposition 7.2 below. This says that an occurrence of φ can lead to surprise about φ itself, but also about all of its consequences. For example, it follows from items 1 and 2 that if an agent assigns probability 0.2 to $p \wedge q$, then after the announcement of this conjunction, she is surprised with intensity 0.8 about $p \wedge q$, but also about p and q individually. Items 3 and 4 are trivial consequences of 1 and 2; they are mentioned to highlight the subtleties of unsuccessful formulas: if φ is not assumed to be successful, then 4 continues to hold, but 3 doesn't.

Proposition 7.2. *Assume that $\varphi \in \mathcal{L}$ is successful, and that $\models \varphi \rightarrow \psi$. Then:*

1. $\mathcal{C}_S \models P(\varphi) \geq c \rightarrow [!\varphi]S(\psi) \leq 1 - c$,
2. $\mathcal{C}_S \models P(\varphi) \leq c \rightarrow [!\varphi]S(\psi) \geq 1 - c$,
3. $\mathcal{C}_S \models P(\varphi) \geq c \rightarrow [!\varphi]S(\varphi) \leq 1 - c$,
4. $\mathcal{C}_S \models P(\varphi) \leq c \rightarrow [!\varphi]S(\varphi) \geq 1 - c$.

[25] For example, let $\varphi := p \wedge P(p) = 0.05$, and let $\mathbb{M}$ be the surprise model defined in Example 7.1; it is now easy to check that $\mathbb{M}, w \models \varphi \wedge [!\varphi]P(\varphi) = 0$.

Proof. Straightforward applications of the semantics. □

The fact that an occurrence of φ can lead to surprise about its consequences presupposes that the agent is actually able to *draw* those consequences (if the agent did not realize that ψ is a logical consequence of φ, then an unexpected occurrence of φ would cause her to be surprised about φ, but not about ψ). In other words, Proposition 7.2 shows that the logical system assumes the agent to be logically omniscient.[26] An even clearer illustration of this assumption is provided by item 1 of Proposition 7.3 below, which says that the agent is never surprised about semantic validities. Similarly, items 2 and 3 say that if an agent already knows φ, or assigns probability 1 to it, then she will not be surprised about it. These principles are clearly false for actual human beings, which are not logically omniscient, and can thus e.g. be genuinely surprised upon learning (that some formula is actually) a semantic validity; rather, the main importance of item 1 is that it elucidates Wittgenstein's famous anti-psychologistic claim that "there can never be surprises in logic" (Wittgenstein 1922, Proposition 6.1251).

Proposition 7.3. *Assume $\varphi \in \mathcal{L}$ is successful. Then:*

1. *if $\mathcal{C}_S \models \varphi$, then $\mathcal{C}_S \models [!\varphi]S(\varphi) = 0$,*
2. *$\mathcal{C}_S \models P(\varphi) = 1 \rightarrow [!\varphi]S(\varphi) = 0$,*
3. *$\mathcal{C}_S \models K\varphi \rightarrow [!\varphi]S(\varphi) = 0$.*

Proof. Straightforward applications of the semantics. □

I will finish this subsection by proving two more substantial results, both of which illustrate how important empirical properties of surprise can be obtained as semantic validities of the logical system.

Proposition 7.4. *Assume $\varphi \in \mathcal{L}$ is successful. Then for all $n \geq 2$, we have:*[27]

$$\mathcal{C}_S \models [!\varphi]^n S(\varphi) = 0.$$

[26] This also illustrates the thoroughly *epistemic* character of surprise: the problem of logical omniscience is originally a problem for epistemic logic, but it automatically carries over into the surprise component.

[27] $[!\varphi]^n$ is defined inductively: $[!\varphi]^0\psi := \psi$, and $[!\varphi]^{n+1}\psi := [!\varphi][!\varphi]^n\psi$.

Proof. First of all, note that since φ is successful, it holds that $\models \varphi \leftrightarrow \langle !\varphi\rangle\varphi$; call this principle (†). Consider an arbitrary surprise model $\mathbb{M} = \langle W, R, \mu, \sigma, V\rangle$ and state w, and assume that $\mathbb{M}, w \models \varphi$. For any $n \geq 0$, we abbreviate

$$\langle W^n, R^n, \mu^n, \sigma^n, V^n\rangle = \mathbb{M}|n := (\cdots(\mathbb{M}\underbrace{|\varphi)|\varphi\cdots)|\varphi}_{n \text{ times}}.$$

Let's now show that $\mathbb{M}, w \models [!\varphi]^{n+1}P(\varphi) = 1$ for all $n \geq 0$. This follows directly from the following calculation:

$$\begin{array}{rcll} \mu^{n+1}(w)([\![\varphi]\!]^{\mathbb{M}|n+1}) & = & \mu^{n+1}(w)([\![\langle !\varphi\rangle\varphi]\!]^{\mathbb{M}|n}) & \\ & = & \mu^{n+1}(w)([\![\varphi]\!]^{\mathbb{M}|n}) & (\dagger) \\ & = & \mu^{n}(w)([\![\varphi]\!]^{\mathbb{M}|n} \mid [\![\varphi]\!]^{\mathbb{M}|n}) = 1. & \end{array}$$

We now use this to justify the (‡)-labeled step in the following calculation:

$$\begin{array}{rcll} \sigma^{n+2}(w)([\![\varphi]\!]^{\mathbb{M}|n+2}) & = & \sigma^{n+2}(w)([\![\langle !\varphi\rangle\varphi]\!]^{\mathbb{M}|n+1}) & \\ & = & \sigma^{n+2}(w)([\![\varphi]\!]^{\mathbb{M}|n+1}) & (\dagger) \\ & = & 1 - \mu^{n+1}(w)([\![\varphi]\!]^{\mathbb{M}|n+1}) & \\ & = & 1 - 1 = 0. & (\ddagger) \end{array}$$

This shows that $\mathbb{M}, w \models [!\varphi]^{n+2}S(\varphi) = 0$ for all $n \geq 0$. □

Informally speaking, Proposition 7.4 says that after two public announcements of φ, the agent is no longer surprised about φ. It thus nicely captures the transitory nature of surprise, which was discussed in Subsection 7.2.1. Furthermore, the proof closely resembles the informal explanation which was given there: the first announcement of φ causes the agent to update her probabilities and to assign probability 1 to φ, so that the second (and subsequent) announcement is no longer unexpected, and thus no longer surprising.[28]

Finally, Proposition 7.5 says that if an occurrence of (a public announcement of) φ leads an agent to change her probability of ψ from a to b in a non-

[28] The fact that surprise intensity drops to 0 after only two announcements is no problem for Proposition 7.4, even though for most real subjects this drop happens more gradually and requires several more repetitions (Charlesworth 1964). The more gradual decrease in surprise intensity is the consequence of personal and coincidental factors, such as intelligence and fatigue. Both the informal explanation in Subsection 7.2.1 and Proposition 7.4 make abstraction of such factors, and predict that the drop in surprise intensity will already happen after the second repetition.

trivial[29] fashion, then she will experience at least *some* surprise about ψ. In other words: surprise is a *necessary condition* for belief revision (in the current framework: probability revision).[30] This is perfectly in line with the cognitive-psychoevolutionary theory of surprise described in Subsection 7.2.1, which holds that surprise is part of a sequence of processes triggered by an unexpected event; the final stage of this sequence is typically a process of belief revision.

Proposition 7.5. *Consider $\varphi, \psi \in \mathcal{L}$ and suppose that $\models \neg\psi \to [!\varphi]\neg\psi$. Then*

$$\mathcal{C}_S \models \big(P(\psi) = a \ \wedge \ [!\varphi]P(\psi) = b \ \wedge \ a \neq b\big) \to [!\varphi]S(\psi) > 0.$$

Proof. Consider an arbitrary surprise model $\mathbb{M} = \langle W, R, \mu, \sigma, V\rangle$ and state w, and assume that the antecedent of the formula above is true at $\mathbb{M}, w$. For a reductio, assume that $\mathbb{M}, w \not\models [!\varphi]S(\psi) > 0$. Then it follows that

$$0 = \sigma^{\varphi}(w)(\llbracket \psi \rrbracket^{\mathbb{M}|\varphi}) = \sigma^{\varphi}(w)(\llbracket \langle !\varphi\rangle\psi \rrbracket^{\mathbb{M}}) = 1 - \mu(w)(\llbracket \langle !\varphi\rangle\psi \rrbracket^{\mathbb{M}}),$$

and thus $\mu(w)(\llbracket \langle !\varphi\rangle\psi \rrbracket^{\mathbb{M}}) = 1$. From the assumption that $\models \neg\psi \to [!\varphi]\neg\psi$ in the statement of the proposition, it follows that $\llbracket \langle !\varphi\rangle\psi \rrbracket^{\mathbb{M}} \subseteq \llbracket \psi \rrbracket^{\mathbb{M}}$, and thus

$$1 = \mu(w)(\llbracket \langle !\varphi\rangle\psi \rrbracket^{\mathbb{M}}) \leq \mu(w)(\llbracket \psi \rrbracket^{\mathbb{M}}) = a,$$

so $a = 1$. Since $\models \langle !\varphi\rangle\psi \to \varphi$, we similarly get that $\mu(w)(\llbracket \varphi \rrbracket^{\mathbb{M}}) = 1$, and hence

$$b = \mu^{\varphi}(w)(\llbracket \psi \rrbracket^{\mathbb{M}|\varphi}) = \mu^{\varphi}(w)(\llbracket \langle !\varphi\rangle\psi \rrbracket^{\mathbb{M}}) = \frac{\mu(w)(\llbracket \langle !\varphi\rangle\psi \rrbracket^{\mathbb{M}})}{\mu(w)(\llbracket \varphi \rrbracket^{\mathbb{M}})} = \frac{1}{1} = 1.$$

We thus have $a = 1 = b$, which contradicts the assumption that $a \neq b$. □

[29] This non-triviality requirement is captured by the condition that $\models \neg\psi \to [!\varphi]\neg\psi$, i.e. the public announcement of φ should not turn any $\neg\psi$-states into ψ-states. In other words, the change of $P(\psi)$ from a to b is non-trivial if $\llbracket \psi \rrbracket^{\mathbb{M}}$ does not grow. (If $\llbracket \psi \rrbracket^{\mathbb{M}}$ grows, then it is trivial that the value of $P(\psi)$ might change: if $A \subseteq B$, then $P(A) \leq P(B)$.) Intuitively, exactly the same argument can be made about $\llbracket \psi \rrbracket^{\mathbb{M}}$ shrinking rather than growing (i.e. about the requirement that $\models \psi \to [!\varphi]\psi$), but it turns out that this second requirement is technically speaking not necessary for Proposition 7.5 to hold. This disanalogy is similar to the disanalogy between items 3 and 4 of Proposition 7.2.

[30] I use the term 'belief revision' as synonymous to 'probability revision' here, and do not mean to suggest any straightforward technical connection with AGM-style theories of belief revision (Alchourrón et al. 1985, Gärdenfors 1988).

Corollary 7.2. *For any $\varphi \in \mathcal{L}$, it holds that*

$$\mathcal{C}_S \models \big(P(\varphi) = a \wedge [!\varphi]P(\varphi) = b \wedge a \neq b\big) \rightarrow [!\varphi]S(\varphi) > 0.$$

Proof. It always holds that $\models \neg\varphi \rightarrow [!\varphi]\neg\varphi$, so by putting $\psi = \varphi$, the condition of Proposition 7.5 is always satisfied. □

7.4.4 A Lockean Thesis for Surprise

The current framework allows us to express statements such as 'the agent is surprised about φ with intensity 0.8'. In many natural cases, however, we might want to say that the agent is surprised about φ, without wishing to commit ourselves to some particular value for her surprise intensity. This is entirely analogous to the epistemic cases, where we might sometimes want to say that the agent believes that φ, without committing ourselves to some particular degree of belief.

A widespread proposal is to define 'belief' as 'sufficiently high degree of belief'; this proposal is called the *Lockean thesis*, and was studied in Chapter 6. Formally, the Lockean thesis for belief looks as follows:

$$B\varphi \;:\equiv\; P(\varphi) \geq \tau \tag{7.10}$$

where $\tau \in (0.5, 1)$ is some threshold value. In Chapter 6, I also introduced a more sophisticated version that defines conditional belief in terms of high conditional probability, but the basic version (7.10) will suffice for our current purposes.

Because of the high similarity between the epistemic case and the surprise case, it seems natural to apply the Lockean thesis also to surprise. In other words, we will introduce a 'qualitative' surprise operator by saying that the agent is surprised about φ iff she is surprised about φ with some sufficiently high intensity. I will argue below that the most natural choice for the value of the surprise intensity threshold is τ, i.e. the same value as the degree of belief threshold. Formally, the Lockean thesis for surprise thus looks as follows:

$$S\varphi \;:\equiv\; S(\varphi) \geq \tau \tag{7.11}$$

Principles (7.10) and (7.11) allow us to talk about an agent's 'qualitative' beliefs and surprises. Furthermore, since both principles make use of the same threshold value τ, there is a natural connection between the two operators they define. Proposition 7.6 says that after an announcement of φ, the agent will be surprised about ψ iff (assuming that φ is true) she initially believed that ψ would be false then. This qualitative observation is in line with the psychoevolutationary theory of surprise described in Subsection 7.2.1, which holds that surprise stems from a conflict between unexpected data and a previously held belief.

Proposition 7.6. $\mathcal{C}_S \models [!\varphi]S\psi \longleftrightarrow \varphi \to B[!\varphi]\neg\psi$.

Proof. Consider the reduction axiom for surprise formulas:

$$[!\varphi]S(\psi) \geq \tau \longleftrightarrow \varphi \to P(\langle !\varphi \rangle \psi) \leq 1 - \tau. \tag{7.12}$$

We have the following chain of SURPRISE-equivalences:

$$\begin{aligned} P(\langle !\varphi \rangle \psi) \leq 1 - \tau &\leftrightarrow 1 - P(\langle !\varphi \rangle \psi) \geq \tau \\ &\leftrightarrow P(\neg \langle !\varphi \rangle \psi) \geq \tau \\ &\leftrightarrow P([!\varphi]\neg\psi) \geq \tau \end{aligned}$$

and thus (7.12) can be rewritten as

$$[!\varphi]S(\psi) \geq \tau \longleftrightarrow \varphi \to P([!\varphi]\neg\psi) \geq \tau.$$

Applying (7.11) and (7.10) to the left- and right-hand sides, respectively, yields the desired result. □

The Lockean theses for belief (7.10) and surprise (7.11) thus allow us to simulate qualitative notions of belief and surprise in a probabilistic framework. The interaction between these notions, as described in Proposition 7.6, is in line with the cognitive-psychoevolutionary theory of surprise. Note that this theory is not primarily concerned with the static properties of belief and/or surprise (such as being closed under conjunction), but rather with their dynamic properties. Technical results such as Propositions 7.4, 7.5 and 7.6 show that the framework developed here is able to capture, in particular, these dynamic properties. In sum, then, the fruitfulness of applying the Lockean thesis inside this framework can be seen as a further illustration of the pragmatic argument in favor of this thesis (see Section 6.6).

7.5 Conclusion

In this chapter, I have presented a new analysis of surprise in the framework of probabilistic dynamic epistemic logic. This analysis is based on current psychological theories, and as a result, several experimentally observed aspects of surprise can be derived as theorems within the logical system (recall, for example, Proposition 7.5 on the role of surprise in belief revision). Furthermore, being based on the contemporary 'lingua franca' of (dynamic) epistemic logic, it offers a natural, well-understood and highly expressive language for the formal description of agent architectures (cf. Proposition 7.1). The framework also allows us to define qualitative surprise and belief operators, and study their interaction (cf. Proposition 7.6).

Most importantly, however, the analysis naturally captures the dynamic nature of surprise. This is clearly manifested in the logic's semantics (the surprise measures $\sigma(w)$ are not required to satisfy any static properties) as well as in its proof theory (the only substantial axioms for surprise are its reduction axioms). These reduction axioms jointly constitute a temporally coherent definition of surprise, in contrast to earlier, temporally incoherent formalizations such as Macedo and Cardoso's and Lorini and Castelfranchi's. This temporal coherence has several advantages. First and foremost, by explicitly distinguishing between prior and posterior notions, the proposed analysis is able to reach a high level of *conceptual hygiene* (recall the methodological remark at the beginning of Subsection 7.3.2). This conceptual advantage also yields additional *empirical* benefits: the new analysis can capture important aspects of surprise that are not covered by earlier frameworks, such as its transitory nature (cf. Proposition 7.4).[31]

Several questions are left for further research. For example, I intend to explore what happens with the propositions mentioned in Subsection 7.4.3 when the assumption of successfulness is lifted (unsuccessful formulas require higher-order information, and thus seem to arise most naturally in multi-agent scenarios; recall Footnote 23). Another topic involves adding awareness to the logic, which would greatly increase its empirical adequacy (cf. Proposition 7.3).

[31] Unsurprisingly, the aspect of transitoriness is itself of a highly dynamic character, involving repeated occurrences of the unexpected event.

Part III

Dynamic Epistemic Logic and Logical Geometry

8 Aristotelian Diagrams for Dynamic Epistemic Logic

8.1 Introduction

The Aristotelian square of oppositions is a historically rich and interesting subject in philosophical logic. It provides a compact way of representing various logical relations between formulas, and thus serves as an illustration of the underlying logic's expressive and deductive powers. Therefore, throughout the history of logic, many authors have found it worthwhile to show that the logics they were studying gave rise to square-like diagrams. Typical examples include the construction of Aristotelian squares for deontic logic (McNamara 2010) and modal logic (Fitting and Mendelsohn 1998).[1]

In this chapter, however, I will continue studying the (family of) logic(s) that has been the main topic of this thesis so far, viz. dynamic epistemic logic. Lenzen, one of the main epistemic logicians of the 20th century (Lenzen 1978, 1980), has shown how squares can be constructed for notions such as knowledge and belief (Lenzen 2012). He thus connected 'classical' epistemic logic and 'classical' squares of oppositions. In recent years, however, both topics have been developing rapidly. On the one hand, squares of oppositions have been generalized, mainly by Moretti (2009a) and Smessaert (2009), to larger and more complex Aristotelian diagrams; these generalizations are systematically studied in *logical geometry*.[2] On the other hand, epistemic logic has been dynami-

[1] A more elaborate overview of the square's history and applications is given in Section 9.2.

[2] The term 'logical geometry' dates back to Smessaert's PhD thesis (Smessaert 1993). Moretti (2009a) prefers the terms 'n-opposition theory', and, more recently, 'oppositional geometry' (Moretti 2012a). There are subtle conceptual and technical differences between Moretti's ap-

fied, leading to various versions of *dynamic epistemic logic* (van Ditmarsch et al. 2007, van Benthem 2011).

The main purpose of this chapter is to study dynamic epistemic logic from the perspective of logical geometry. In other words, I will establish a connection between 'contemporary' epistemic logic and 'contemporary' generalizations of the Aristotelian square of oppositions.[3,4]

Establishing such a connection has many advantages for both of the frameworks involved. On the one hand, dynamic epistemic logic can benefit from the representational powers of logical geometry. Epistemic dynamic phenomena (such as public announcements) have various structural and epistemic properties. Because of the expressivity of dynamic epistemic logic, especially the latter tend to become quite subtle. The tools of logical geometry allow us to visualize this vast amount of information in a clear and compact way. They can thus help us to understand the subject more thoroughly, and even gain new insights.[5] On the other hand, there are also several advantages for logical geometry itself. Dynamic epistemic logic provides new examples of the use of logical geometry, thus broadening its scope of applicability. More importantly, this chapter shows that not only classical static modalities, but also dynamic modalities give rise to oppositional phenomena, thus broadening the scope of logical geometry in a more fundamental way. Finally, the technical properties of such dynamic modalities turn out to be directly relevant for the philosophical foundations of logical geometry.

The remainder of this chapter is organized as follows. Section 8.2 introduces one particular type of epistemic dynamics, viz. public announcements,[6] and dis-

proach and Smessaert's approach; however, these are irrelevant for our current purposes, so I will henceforth uniformly use the term 'logical geometry'.

[3] Moretti (2009a, p. 306–308) has already shown how Lenzen's original squares for epistemic logic can be generalized to larger Aristotelian diagrams (in particular, tetraicosahedra). In other words, he has used the *contemporary* machinery of logical geometry to look at *classical* epistemic logic.

[4] Note that the terms 'classical epistemic logic' and 'contemporary epistemic logic' are used here in the same sense as in Subsection 6.2; however, the focus here is exclusively on the static/dynamic distinction (not on the single-agent/multi-agent distinction).

[5] This heuristic role is not often acknowledged. A notable exception are Davey and Priestley, who write that "Diagram-drawing is as much an art as a science [. . .] good diagrams can be a real asset to understanding and to theorem-proving." (Davey and Priestley 2002, p. 11).

[6] For ease of exposition, I focus on public announcement logic, rather than dynamic epistemic logic in general (with product updates). However, all the results obtained in this chapter straight-

cusses some of its structural and epistemic properties. Section 8.3 initiates the investigation of Aristotelian diagrams in public announcement logic, focusing on the *structural* properties of public announcements. I show how an Aristotelian square and hexagon can be constructed in a non-trivial fashion, and argue that dynamic modalities fit well in the structuralist philosophy of logical geometry. In Section 8.4, we turn to the *epistemic* properties of public announcements. I show how to construct an octagon and a rhombic dodecahedron for public announcement logic, and compare these results to those of Smessaert (2009) for the modal logic $\mathsf{S5}$. Section 8.5, finally, wraps things up by summarizing the results obtained in this chapter.

8.2 A Brief Overview of Public Announcement Logic

This section provides a brief overview of public announcement logic, focusing on those aspects that are most important for the further development of the chapter. This introduction is quite brief, since an informal discussion of public announcement logic was already given in Subsection 1.2, and the technical details of its probabilistic extensions have already been discussed extensively earlier in this thesis (in particular, see Section 3.3).

As usual, we fix a finite set I of agents and a countably infinite set Prop of atomic propositions. The models used in public announcement logic are multi-agent Kripke models $\mathbb{M} = \langle W, R_i, V\rangle_{i\in I}$, where R_i is an equivalence relation on W. These models are thus probabilistic Kripke models (see Definition 3.2 on p. 75) without the probabilistic components μ_i. The formal language of public announcement logic is defined by means of the following BNF:

$$\varphi \ ::= \ p \mid \neg\varphi \mid (\varphi\wedge\varphi) \mid K_i\varphi \mid [!\varphi]\varphi$$

—where $p \in \mathsf{Prop}$ and $i \in I$. Note that this is just the language of probabilistic epistemic logic (see its BNF on p. 77) without i-probability formulas, and thus its intuitive interpretation and formal semantics do not need much additional explanation. Below, I will just highlight a few aspects that are particularly relevant for our current purposes.

Recall that the public announcement operator has a dual, which is defined as $\langle !\varphi\rangle\psi :\equiv \neg[!\varphi]\neg\psi$. Furthermore, recall that the standard meaning of $[!\varphi]\psi$ is

forwardly generalize from public announcement logic to product update logic.

that after a public announcement of φ (assuming it can be publicly announced at all), it will be the case that ψ. Similarly, $\langle !\varphi\rangle\psi$ means that φ can actually be publicly announced, and after this public announcement of φ, it will be the case that ψ. These two operators can thus be seen as quantifying over the set of public announcements of φ: $[!\varphi]\psi$ means that ψ holds after *all* public announcements of φ,[7] and $\langle !\varphi\rangle\psi$ means that ψ holds after *at least one* public announcement of φ (see Subsection 8.3.3 for a more abstract discussion of this quantificational interpretation of the public announcement operators).

Recall the formal semantics of $[!\varphi]\psi$ and $\langle !\varphi\rangle\psi$:

$$\begin{array}{lll} \mathbb{M}, w \models [!\varphi]\psi & \text{iff} & \text{if } \mathbb{M}, w \models \varphi \text{ then } \mathbb{M}|\varphi, w \models \psi, \\ \mathbb{M}, w \models \langle !\varphi\rangle\psi & \text{iff} & \mathbb{M}, w \models \varphi \text{ and } \mathbb{M}|\varphi, w \models \psi. \end{array}$$

These clauses involve going to the updated model $\mathbb{M}|\varphi$. This is defined by means of the well-known 'state-deleting' process, just as in the definition of updated well-behaved probabilistic Kripke models (see Definition 3.6 on p. 86). The following definitions are entirely standard:

Definition 8.1. Consider an arbitrary formula φ. Then

- φ is said to be *PAL-valid* (notation: $\models \varphi$) iff $\mathbb{M}, w \models \varphi$ for all Kripke models $\mathbb{M}$ and states w,[8]
- φ is said to be *PAL-contingent* iff $\not\models \varphi$ and $\not\models \neg\varphi$.

Usually, the 'PAL' qualifier will be omitted, and we will simply talk about 'valid' or 'contingent' formulas.

[7] Recall that if φ is false, then it cannot be publicly announced at all (because of the truthfulness assumption of public announcements). This means that the set of all public announcements of φ is empty, and hence, the formula $[!\varphi]\psi$, which involves universally quantifying over this set, is vacuously true.

[8] Public announcement logic also has a stronger notion of validity. A formula φ is said to be *schematically valid* iff all of its substitution instances $\varphi[\psi/p]$ are valid (where $\varphi[\psi/p]$ is the formula that results from uniformly replacing every occurrence of p in φ with an occurrence of ψ). Obviously, schematic validity implies validity. In logics which have the *uniform substitution* property (if φ is valid, then every substitution instance $\varphi[\psi/p]$ is valid), validity also implies schematic validity, and thus the two notions coincide. Public announcement logic, however, does not have the uniform substitution property, and hence, schematic PAL-validity is strictly stronger than 'ordinary' PAL-validity (van Benthem 2006a,b). Also see Footnote 10.

I will now discuss some important formal properties of public announcements, which will be useful in Sections 8.3 and 8.4. We distinguish between two types of properties: *structural* properties, which describe the general structure of public announcement as a dynamic phenomenon, and *epistemic* properties, which describe the interaction between public announcements and knowledge.

The structural properties of public announcements are summarized by the following lemma:[9]

Lemma 8.1. *The following hold:*

1. $\models [!\varphi](\psi \to \chi) \to ([!\varphi]\psi \to [!\varphi]\chi)$,
2. *if* $\models \psi$ *then* $\models [!\varphi]\psi$,
3. $\models \varphi \leftrightarrow \langle !\varphi \rangle \top$,
4. *if* φ *is contingent, then* $\not\models \langle !\varphi \rangle \top$,
5. *if* φ *is contingent, then* $\not\models [!\varphi]\psi \to \langle !\varphi \rangle \psi$,
6. $\models \langle !\varphi \rangle \psi \to [!\varphi]\psi$.

Items 1 and 2 say that the public announcement operator $[!\varphi]$ satisfies distributivity and the 'necessitation' rule, which is often summarized by saying that it is a *normal* modal operator.[10] Item 3 says that announceability equals truth: a formula can be announced iff it is true. Since not all formulas are true, it follows that not all formulas can be announced, which is what item 4 says. This property is called the *partiality* of public announcements. Item 5 says that $[!\varphi]\psi$ does not require that φ can be announced, whereas $\langle !\varphi \rangle \psi$ does require this. Finally, item 6 says that public announcement is *functional*: if a formula φ is announced in identical states (in which it *can* be announced, i.e. in which it is true), this will always lead to identical 'outcome states'. In model-theoretic terms: the updated model $\mathbb{M}|\varphi$ is uniquely determined by $\mathbb{M}$ and φ.

We now turn to the epistemic properties of public announcements:

Lemma 8.2. *The following hold:*

[9] For proofs of Lemmas 8.1 and 8.2, see van Ditmarsch et al. (2007, chapter 4).

[10] However, it should be emphasized that public announcement logic, as a logical system, is not a normal modal logic, because it does not have the uniform substitution property. For example, it holds that $\models [!p]p$, but $\not\models [!(p \wedge \neg K_i p)](p \wedge \neg K_i p)$ (van Ditmarsch et al. 2007, p. 106).

1. $\models \langle !\varphi\rangle K_i\psi \rightarrow K_i[!\varphi]\psi$*,*

2. $\models K_i[!\varphi]\psi \rightarrow [!\varphi]K_i\psi$*,*

3. $\models \langle !\varphi\rangle \neg K_i\psi \rightarrow \neg K_i[!\varphi]\psi$*,*

4. $\models \neg K_i[!\varphi]\psi \rightarrow [!\varphi]\neg K_i\psi$*.*

Item 1 says that if φ can be announced and after that announcement agent i knows that ψ is the case, then she already knows 'now' (i.e. before any announcements) that ψ will be the case after any announcement of φ. Similarly, item 2 says that if i knows 'now' that ψ will be the case after any announcement of φ, then after any announcement of φ she will know that ψ is the case. Finally, items 3 and 4 are merely the contrapositives of items 2 and 1, respectively, so they will not be discussed separately.

The general idea behind these principles is quite clear: they all state sufficient or necessary conditions for an agent to know something *after* a public announcement in terms of what she knows *before* the announcement. The more fine-grained distinctions between them (in particular, whether $[!\varphi]$ or $\langle !\varphi\rangle$ is used) illustrate the subtlety of the interactions between public announcement and knowledge. The Aristotelian diagrams developed in Sections 8.3 and 8.4 visualize networks of subtle interactions such as these ones.

8.3 Structural Oppositions for Public Announcements

This section initiates our investigation of Aristotelian diagrams for public announcement logic. For now, I will focus on the *structural* properties of public announcements (as listed in Lemma 8.1), and thus consider only 'structural oppositions'. Subsection 8.3.1 introduces some key notions from logical geometry, loosely based on Smessaert (2009) and Smessaert and Demey (2013b). These notions are used in Subsection 8.3.2 to construct an Aristotelian square and hexagon. Finally, in Subsection 8.3.3, I make some remarks about the role of partial functionality, and connect this with the structuralist positions held by many authors in logical geometry.

8.3.1 Some Notions from Logical Geometry

The fundamental building blocks of the traditional square of oppositions, and any other Aristotelian diagrams, are the Aristotelian relations:

Definition 8.2. The *Aristotelian relations* are binary relations between formulas, defined as follows:

1. φ and ψ are *contradictory* iff $\models \neg(\varphi \wedge \psi)$ and $\models \varphi \vee \psi$

   (i.e. they cannot be true together and they cannot be false together),

2. φ and ψ are *contrary* iff $\models \neg(\varphi \wedge \psi)$ and $\not\models \varphi \vee \psi$

   (i.e. they cannot be true together, but they can be false together),

3. φ and ψ are *subcontrary* iff $\not\models \neg(\varphi \wedge \psi)$ and $\models \varphi \vee \psi$

   (i.e. they can be true together, but they cannot be false together),

4. φ and ψ are in *subalternation* iff $\models \varphi \rightarrow \psi$ and $\not\models \psi \rightarrow \varphi$

   (φ is called the *superaltern* and ψ the *subaltern*).

It is easy to check that the first three of these relations are symmetrical, and that the fourth one is (by definition) asymmetrical. Furthermore, Lemma 8.3 says that these four relations are mutually exclusive (at least when one restricts to contingent formulas). The logical and informational properties of these relations will be discussed in much more detail in Chapter 9.

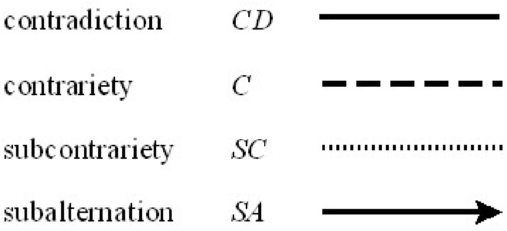
Lemma 8.3. *The Aristotelian relations are mutually exclusive for contingent formulas; i.e. if two contingent formulas φ and ψ stand in one of these four relations, then they cannot stand in any of the other three.*

Proof. First of all, note that it follows trivially from the definitions that the three symmetrical Aristotelian relations are mutually exclusive (one does not even need to restrict to contingent formulas). It thus suffices to show that two contingent formulas cannot simultaneously be in subalternation and in one of the three other Aristotelian relations. For this purpose, let φ and ψ be contingent formulas, and suppose that φ and ψ are in subalternation (so $\models \varphi \rightarrow \psi$). Then:

1. Suppose, towards a contradiction, that φ and ψ are contradictories. Hence $\models \neg(\varphi \wedge \psi)$. From $\models \varphi \rightarrow \psi$ it follows that $\models \varphi \leftrightarrow (\varphi \wedge \psi)$, and hence $\models \neg\varphi$, which contradicts the contingency of φ.

2. Suppose, towards a contradiction, that φ and ψ are contraries. By an analogous argument it follows that $\models \neg\varphi$, which contradicts the contingency of φ.

Figure 8.1: Code for visually representing the Aristotelian relations



contradiction	CD	━━━━━━
contrariety	C	▬ ▬ ▬ ▬ ▬ ▪
subcontrariety	SC	··························
subalternation	SA	⟶

3. Suppose, towards a contradiction, that φ and ψ are subcontraries. Hence $\models \varphi \vee \psi$. Combining this with $\models \varphi \rightarrow \psi$ it follows that $\models \psi$, which contradicts the contingency of ψ. □

We are now ready to define the general notion of an Aristotelian diagram:

Definition 8.3. An *Aristotelian diagram (for public announcement logic)* is a diagram that visualizes a labeled graph G. The vertices of G are contingent and pairwise non-equivalent formulas $\varphi_1, \ldots, \varphi_n$;[11] the edges of G are the Aristotelian relations between those formulas, i.e. if φ_i and φ_j stand in any Aristotelian relation, then this is visualized in the diagram, according to the code in Figure 8.1.[12]

Note that the formulas appearing in Aristotelian diagrams are assumed to be contingent and pairwise non-equivalent. These restrictions are motivated by historical as well as systematical reasons; see the discussion ensuing Definition 9.2 on p. 248 for details. Furthermore, note that Definition 8.3 does not mention visual simplicity, two- or three-dimensionality, symmetry, compactness, elegance, etc. Although logical geometry is primarily concerned with diagrams which have qualities such as these, it is difficult to capture them in formal, precise definitions. (Recall Footnote 5 about diagram-drawing as a science versus an art.)[13]

[11] So $\not\models \varphi_i$, $\not\models \neg\varphi_i$, and $\not\models \varphi_i \leftrightarrow \varphi_j$, for $1 \leq i \neq j \leq n$.

[12] Note that the three symmetrical relations are represented by *non-directed* edges, whereas the asymmetrical subalternation relation is represented by *directed* edges going *from* the superaltern *to* the subaltern.

[13] This does not mean that nothing at all can be said about these properties. For example, in on-

A powerful technique for generating new Aristotelian diagrams from already existing ones is that of taking *Boolean closures*. I first define this notion for sets of formulas, and then for Aristotelian diagrams.

Definition 8.4. A set S of contingent formulas is called *Boolean closed* iff for any $\varphi, \psi \in S$, the following holds:

1. if $\not\models \neg(\varphi \wedge \psi)$, then there is a $\theta \in S$ such that $\models \theta \leftrightarrow (\varphi \wedge \psi)$,
2. there is a $\theta \in S$ such that $\models \theta \leftrightarrow \neg\varphi$.

Definition 8.5. Let S be a set of contingent formulas. Then the *Boolean closure of S* is the set B such that (i) $S \subseteq B$, (ii) B is Boolean closed, and (iii) for any set B' such that $S \subseteq B'$ and B' is Boolean closed, it holds that $B \subseteq B'$.

Definition 8.4 is semantic in nature: if $\varphi \in S$, then this definition does not require that S contains $\neg\varphi$ *itself*, but at least a formula that is *equivalent* to $\neg\varphi$ (similarly for $\wedge$).[14] Note, furthermore, that the $\wedge$-closure condition is also subject to a consistency requirement;[15] for example, if a set S contains the contingencies p and $\neg p$, then it is not required to also contain $p \wedge \neg p$ (or any other equivalent formula), since the latter is not consistent. Finally, note that it is clear from Definition 8.5 that if S is already Boolean closed, then it is its own Boolean closure.

We can now lift these notions from sets of formulas to Aristotelian diagrams:

Definition 8.6. An Aristotelian diagram S is called *Boolean closed* iff the set of formulas appearing as vertices in S is Boolean closed (according to Definition 8.4).

going work with Hans Smessaert, I am exploring the symmetry properties of certain n-dimensional diagrams by viewing them as vertex-first projections of $(n+1)$-dimensional hypercubes (Demey and Smessaert 2013a). We also have some promising group-theoretical results about the relation between the visual properties of a given Aristotelian diagram and the logical/combinatorial properties of the formulas that appear in it (Demey and Smessaert 2013b).

[14] It suffices to consider the connectives $\wedge$ and $\neg$, since these two are functionally complete, i.e. any other truth-functional connective (of any arity) can be written in terms of them (van Dalen 2004, p. 24–25).

[15] This consistency requirement does not need to be stated explicitly for $\neg$-closure, because the negation of a contingent formula is itself always consistent.

Definition 8.7. Let S be an Aristotelian diagram. Let F be the set of formulas appearing as vertices in S, and let F' be the Boolean closure (according to Definition 8.5) of this set. Then the *Boolean closure of S* is the Aristotelian diagram which has as its vertices the formulas of F'.

Definition 8.6 is the natural generalization of Definition 8.4. Note that in Definition 8.7, the set F contains only contingent formulas, so that it indeed makes sense to take its Boolean closure F'. Finally, note that these definitions yield some easy consequences. First of all, the Boolean closure of an Aristotelian diagram is itself always Boolean closed (as an Aristotelian diagram). Secondly, if an Aristotelian diagram is already Boolean closed, then it is its own Boolean closure.

8.3.2 From Square to Hexagon

I will now start putting the tools of logical geometry to use, by constructing Aristotelian diagrams for the structural properties of public announcement (which describe the general structure of public announcement as a dynamic phenomenon).

Despite its dynamic nature, the public announcement operator $[!\varphi]$ is essentially a *modal* operator; recall items 1 and 2 of Lemma 8.1. Modal operators come in dual pairs; for example in alethic modal logic (necessary/possible), epistemic logic (knowing/holding possible), deontic logic (obligatory/permissible), and temporal logic (always/sometimes). The first item of each of these dual pairs is given a universal reading, and the second one an existential reading (for example: 'in *all* possible worlds'/'in *some* possible worlds'). Furthermore, in the squares of oppositions the universal notions occupy the two upper corners, and the existential notions occupy the two lower corners, so that each universal notion and the corresponding existential notion stand in subalternation.[16]

The public announcement operator comes in a dual pair as well: $[!\varphi]$ and $\langle !\varphi\rangle$. Furthermore, recall from Section 8.2 that the first item of this pair can informally be read as a universal quantifier ('after *all* public announcements of φ'), and the second one as an existential quantifier ('after *at least one* public announcement of φ'). Hence, a natural idea might be to construct a square of oppositions for public announcement logic with the formulas $[!p]q$ and $[!p]\neg q$

[16] Note that I do not consider Aristotelian squares for the first-order quantifiers themselves, to avoid getting caught up in the issue of *existential import*. Although this issue has been hotly debated, it is not relevant for our current purposes; also see Footnote 12 on p. 249.

Figure 8.2: Square of oppositions for the structural properties of public announcement

$$\begin{tikzcd}
\langle !p\rangle q \arrow[r, dash, dashed] \arrow[d] \arrow[dr, dash] & \langle !p\rangle\neg q \arrow[d] \arrow[dl, dash] \\
{[!p]q} \arrow[r, dash, dotted] & {[!p]\neg q}
\end{tikzcd}$$

in the upper corners, and the formulas $\langle !p\rangle q$ and $\langle !p\rangle\neg q$ in the lower corners. However, it follows from item 5 of Lemma 8.1 that $\not\models [!p]q \rightarrow \langle !p\rangle q$,[17] and hence, the proposed square cannot be constructed.

Because of the specific dynamic structure of public announcements, however, it is still possible to obtain a valid square of oppositions. The key idea is to *reverse* the direction of the subalternation relation. It then suffices to verify that this reversing operation does not mess up the other Aristotelian relations:

Theorem 8.1. *The square in Figure 8.2 is a valid Aristotelian diagram.*

Proof. It suffices to check that the Aristotelian relations represented in Figure 8.2 (according to the code of Figure 8.1) do indeed hold; Lemma 8.3 then guarantees that there are no other Aristotelian relations that have been 'left out'.

Given Definition 8.2, checking that these relations hold involves showing that certain formulas are valid and that certain others are invalid. For example, checking that $\langle !p\rangle q$ and $\langle !p\rangle\neg q$ are contraries involves checking that

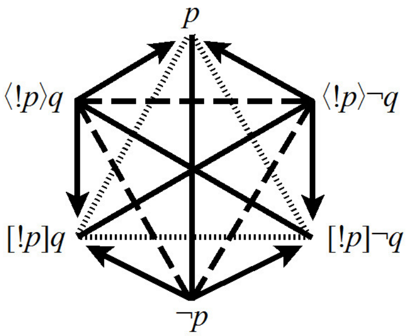
$$\models \neg(\langle !p\rangle q \wedge \langle !p\rangle\neg q) \quad \text{and} \quad \not\models \langle !p\rangle q \vee \langle !p\rangle\neg q.$$

Proving the validities is an easy exercise in reasoning in public announcement logic. Similarly, showing that certain formulas are invalid involves constructing models that falsify them; because these formulas are syntactically very simple, the countermodels can be 'read off' almost directly from them. □

Lemma 8.4. *The Aristotelian square in Figure 8.2 is not Boolean closed.*

[17] Similar remarks apply to $[!p]\neg q$ and $\langle !p\rangle\neg q$.

Figure 8.3: Hexagon of oppositions for the structural properties of public announcement



Proof. Consider, for example, the conjunction of the formulas in the two lower corners, $\chi := [!p]q \wedge [!p]\neg q$. It is easy to check that χ is consistent, but not equivalent to any of the four formulas appearing in the square. □

Because the Aristotelian square in Figure 8.2 is not Boolean closed, we can extend it to a new Aristotelian diagram by constructing its Boolean closure. This turns out to be the hexagon in Figure 8.3.

Theorem 8.2. *The Boolean closure of the Aristotelian square in Figure 8.2 is the Aristotelian hexagon in Figure 8.3.*

Proof. Let S be the set of formulas appearing in the square in Figure 8.2, and let H be the set of formulas appearing in the hexagon in Figure 8.3. Note that $S \subseteq H$; furthermore, the following tables show that H is Boolean closed:[18]

	$\langle !p\rangle q$	$\langle !p\rangle \neg q$	$[!p]q$	$[!p]\neg q$	p	$\neg p$
$\neg$	$[!p]\neg q$	$[!p]q$	$\langle !p\rangle \neg q$	$\langle !p\rangle q$	$\neg p$	p

[18] The format $\begin{array}{c|c} & \alpha \\ \hline \neg & \beta \end{array}$ means that $\models \beta \leftrightarrow \neg\alpha$. Similarly, the format $\begin{array}{c|c} \wedge & \beta \\ \hline \alpha & \gamma \end{array}$ means that $\models \gamma \leftrightarrow (\alpha \wedge \beta)$. Because of the commutativity of $\wedge$, it suffices to state only the upper right half of the table.

$\wedge$	$\langle !p\rangle q$	$\langle !p\rangle\neg q$	$[!p]q$	$[!p]\neg q$	p	$\neg p$
$\langle !p\rangle q$	$\langle !p\rangle q$	$\bot$	$\langle !p\rangle q$	$\bot$	$\langle !p\rangle q$	$\bot$
$\langle !p\rangle\neg q$		$\langle !p\rangle\neg q$	$\bot$	$\langle !p\rangle\neg q$	$\langle !p\rangle\neg q$	$\bot$
$[!p]q$			$[!p]q$	$\neg p$	$\langle !p\rangle q$	$\neg p$
$[!p]\neg q$				$[!p]\neg q$	$\langle !p\rangle\neg q$	$\neg p$
p					p	$\bot$
$\neg p$						$\neg p$

It is easy to check that there is no set of formulas X such that $S \subseteq X \subset H$ and X is Boolean closed. This shows that H is indeed the Boolean closure of S (recall Definition 8.5). Finally, showing that the hexagon in Figure 8.3 is itself an Aristotelian diagram is entirely analogous to the proof of Theorem 8.1. □

The Aristotelian hexagon in Figure 8.3 thus arises as the Boolean closure of the Aristotelian square in Figure 8.2. In logical geometry, this type of hexagon is called a *(strong) Sesmat-Blanché hexagon*, which is named after the first logicians to study such Aristotelian diagrams (Sesmat 1951, Blanché 1952, 1953, 1957, 1966).[19] This observation is more general than the framework of public announcement logic for which it was proved here: *all* of the traditional squares of oppositions gives rise to a (strong) Sesmat-Blanché hexagon *via* their Boolean closure.[20]

I will finish this subsection by making some remarks on how the Aristotelian hexagon in Figure 8.3 represents the structural properties of public announcements. First of all, it shows that public announcement operators are modal operators and thus, like any other modal operator, give rise to Aristotelian diagrams. More importantly, however, the subalternations from $\langle !p\rangle(\neg)q$ to $[!p](\neg)q$ capture the *functionality* of public announcements, while the subalternations involving $(\neg)p$ capture the *partiality* of public announcements (after all, the precon-

[19] The distinction between *strong* and *weak* Sesmat-Blanché hexagons was introduced by Pellissier (2008). A Sesmat-Blanché hexagon contains three pairs of contrary formulas; by the definition of contrariety, the *pairwise* disjunctions of these formulas are not valid. If the *simultaneous* disjunction of all three formulas is valid, the hexagon is called 'strong', otherwise it is called 'weak'. The Sesmat-Blanché hexagon in Figure 8.3 is strong, since $\models \neg p \vee \langle !p\rangle q \vee \langle !p\rangle\neg q$.

[20] In the specific framework of public announcement logic, the hexagon is a bit more elegant than in other frameworks, because the conjunction $[!p]q \wedge [!p]\neg q$ and the disjunction $\langle !p\rangle q \vee \langle !p\rangle\neg q$ 'reduce to' the simpler formulas $\neg p$ and p, respectively. However, when we turn to Aristotelian diagrams for the epistemic properties of public announcements in Section 8.4, we will not be able to maintain this simplicity, and we *will* have to work with syntactically complex formulas.

dition for a public announcement to be executable is the truth of the announced formula).

8.3.3 The Role of Partial Functionality

I have just argued that partial functionality (items 4–6 of Lemma 8.1) plays a key role in the Aristotelian diagrams for the structural properties of public announcements. However, this observation is not restricted to the particular context of public announcement logic. I will now examine the role of partial functionality from a more abstract perspective.

For *any* 'process' or 'operation' (epistemic or otherwise), it makes sense to ask whether it is partially functional. Consider an arbitrary process π (π can be an epistemic process, such as the public announcement of some formula, but it can also be a non-epistemic process, such as the execution of a computer program). Then the following questions arise:

1. Is π functional?[21] In other words, if π is executed in identical states (in which it *can* be executed), will it always lead to identical 'outcome states'?

2. Is π partial? In other words, are there states where π cannot be executed at all? And if so, can the states where π *is* executable be characterized by means of a 'precondition' $pre(\pi)$, so that π is executable in a state iff $pre(\pi)$ is true in that state?

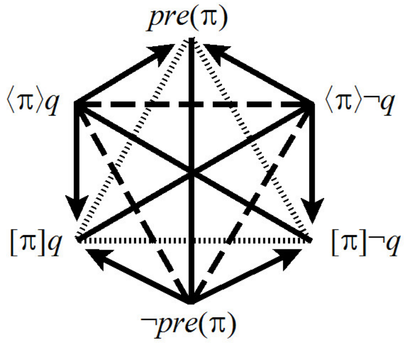
The logical properties of such arbitrary processes are studied by *propositional dynamic logic* (Kozen and Parikh 1981, Harel et al. 2000). This framework represents a process π by means of dynamic modal operators $[\pi]$ and $\langle\pi\rangle$. As is to be expected, $[\pi]\psi$ means that after *all* executions of π it will be the case that ψ, whereas $\langle\pi\rangle\psi$ means that after *some* execution of π it will be the case that ψ.

It is easy to see that a process π is functional iff $\models \langle\pi\rangle\psi \rightarrow [\pi]\psi$, and partial iff $\not\models [\pi]\psi \rightarrow \langle\pi\rangle\psi$.[22] Furthermore, the partiality of π can be further specified by means of its precondition: π can be executed iff (i.e. in exactly those states where) $pre(\pi)$ is true; formally: $\models pre(\pi) \leftrightarrow \langle\pi\rangle\top$. There is an alternative characterization of the partiality of a process π in terms of its precondition $pre(\pi)$: the process π is partial iff $\not\models pre(\pi)$.

[21] Computer scientists prefer the term 'deterministic'.

[22] In this subsection (and in this subsection *only*), I use $\models$ to denote validity in propositional dynamic logic, rather than in public announcement logic.

Figure 8.4: Aristotelian hexagon for a partially functional process π

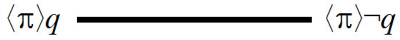


With the same reasoning as in Subsection 8.3.2, we are now able to show that any partially functional process π (whose executability can be captured by means of a precondition $pre(\pi)$) gives rise to an Aristotelian hexagon as in Figure 8.4. In fact, if we take π to be the concrete process $!p$ (a public announcement of p), and recall that the precondition of the public announcement of p is the truth of the announced formula p (i.e. $pre(!p) = p$), then the Aristotelian diagram in Figure 8.3 turns out to be merely a particular instance of that in Figure 8.4.

A question that arises naturally at this point is what happens when we move from *partial* functionality to *total* functionality, i.e. when we focus on processes that are still functional, but that can be executed in *all* states. (Note that in this context the notion of a precondition is meaningless, since a totally functional process has as its precondition always $\top$.) For totally functional processes π, we do not only have $\models \langle\pi\rangle q \rightarrow [\pi]q$, but also the stronger $\models \langle\pi\rangle q \leftrightarrow [\pi]q$.

Since Aristotelian diagrams cannot contain equivalent formulas, this means that the two left formulas of (the square part of) Figure 8.4 ($\langle\pi\rangle q$ and $[\pi]q$)

Figure 8.5: Degenerate Aristotelian diagram for a totally functional process π



collapse, and similarly that the two right formulas ($\langle\pi\rangle\neg q$ and $[\pi]\neg q$) collapse. Hence, for totally functional processes the square of oppositions collapses into a single binary opposition (viz. a contradiction relation), as in Figure 8.5. Furthermore, it is easy to see that this Aristotelian diagram is already Boolean closed, and is thus its own Boolean closure.

In sum: from the perspective of logical geometry, partially functional processes are more interesting than totally functional ones, because the latter only give rise to highly degenerate Aristotelian diagrams (Figure 8.5), whereas the former give rise to rich, non-degenerate Aristotelian diagrams, viz. (strong) Sesmat-Blanché hexagons (Figure 8.4).

I will finish this subsection by showing how the partial functionality of dynamic modal operators nicely corroborates the structuralist foundations of logical geometry (Moretti 2009a). According to the structuralist viewpoint, Aristotelian diagrams are in the first place determined by their constituent *relations*, rather than by their constituent *formulas*. The identity and properties of the concrete formulas only matter insofar as they stand in one of the Aristotelian relations.[23]

An important argument for this claim stems from the fact that Aristotelian diagrams can be constructed for a wide variety of logics. For ease of exposition, let us focus on Aristotelian *squares* (generalizations to other Aristotelian diagrams are straightforward). There are squares for Aristotelian syllogistics, alethic modal logic, epistemic logic, deontic logic, temporal logic, etc. All of these logics have formulas of very different kinds ('all As are Bs', 'it is necessary that φ', 'agent i knows that φ', 'it is obligatory that φ', 'it is always the case that φ', etc.), yet all of these formulas enter into the same kinds of Aristotelian relations (*in casu*: a contradiction relation, a contrariety relation, and (being the superaltern in) a subalternation relation). Hence, the 'relational' properties of these formulas are more important than the 'subject matter' that they speak about (knowledge/time/etc.).

Opponents of structuralism grant all of this, but they maintain that at a sufficiently high level of abstraction, the formulas still have 'independent' (non-relational) properties that determine their relational properties. For example, Parry and Hacker (1991, p. 157) *define* the contrariety relation as holding be-

[23] Note that I implicitly subscribed to this structuralist philosophy when I claimed in Subsection 8.3.1 that the 'fundamental building blocks' of Aristotelian diagrams are the Aristotelian relations, rather than the concrete formulas that stand in those relations.

tween universal formulas of different quality, and the subcontrariety relation as holding between existential formulas of different quality.[24] Hence, the formulas' independent properties (universality/existentiality) are still essential to determine their relational properties.

Parry and Hacker are right that in all traditional Aristotelian squares, the contrariety relation holds between 'universal' formulas of different quality (for example, between $\Box q$ and $\Box\neg q$) and the subcontrariety relation holds between 'existential' formulas of different quality (for example, between $\Diamond q$ and $\Diamond\neg q$); also see the remarks at the beginning of Subsection 8.3.2. The Aristotelain squares obtained from partially functional dynamic modal operators, however, are counterexamples to Parry and Hacker's claims: in these diagrams, the formulas in the contrariety relation are *existential* ($\langle\pi\rangle q$ and $\langle\pi\rangle\neg q$), and the formulas in the subcontrariety relation are *universal* ($[\pi]q$ and $[\pi]\neg q$).

This clearly shows that the independent properties of the formulas (universality/existentiality) do *not* suffice to determine their relational properties. In other words, the Aristotelian relations cannot be reduced to the properties of independent formulas, and thus still need to be taken as primitive—i.e., we have arrived back at the original structuralist position.

8.4 Epistemic Oppositions for Public Announcements

This section continues our investigation of Aristotelian diagrams in public announcement logic. We shift our attention from the structural to the epistemic properties of public announcements, thus arriving at 'epistemic oppositions'. In Subsection 8.4.1, I show that this naturally leads to the construction of an Aristotelian octagon. In Subsection 8.4.2 this octagon is further generalized to a three-dimensional Aristotelian diagram, viz. a rhombic dodecahedron. Subsection 8.4.3 provides a comparison between the results obtained here and those of Smessaert (2009).

[24] Parry and Hacker (1991, p. 161) *do* state the characterizations of contrariety and subcontrariety in terms of 'being able to be true/false together' (as in our Definition 8.2), but only as further *lemmas* about these relations, not as their *definitions*. Similar remarks apply to de Pater and Vergauwen (2005, p. 100).

8.4.1 Adding Knowledge: From Hexagon to Octagon

In Subsection 8.3.3, I argued that the essence of the Aristotelian hexagon in Figure 8.3 was the partial functionality of public announcements, and noted that this can be generalized to *any* partially functional process π (see the Aristotelian hexagon in Figure 8.4). In Sections 8.1 and 8.2, however, I stated that this chapter would be concerned not just with dynamics in general, but with one particular kind of dynamics, viz. public announcements.

The main difference between public announcements and other (partially functional) processes is that public announcements are *epistemic* processes: they interact with the agents' knowledge. For example, after the public announcement of a formula φ, it (usually)[25] becomes common knowledge between the agents that φ is the case. The most important of these interactions were stated in Lemma 8.2. I will now use this lemma to construct Aristotelian diagrams for the epistemic properties of public announcements (i.e. for public announcements *qua* epistemic processes, not simply *qua* partially functional processes).

I will first informally describe how to build these new, 'epistemic' Aristotelian diagrams as conservative extensions of the original hexagon (Figure 8.3). The first step is to replace the ontic formula q (which describes what is the case after the public announcement) with the epistemic formula Kq.[26] For example, the formula in the upper left corner is now no longer $\langle !p\rangle q$, but rather $\langle !p\rangle Kq$. On the left side, we thus obtain the subalternation $\langle !p\rangle Kq \to [!p]Kq$. By inserting the formula $K[!p]q$, this subalternation can be broken up into *two* subalternations. (Note that this new formula no longer speaks about what will be the case *after* any/some public announcement of p; rather it says something about (what is known in) the *present*.) Similar remarks apply to the subalternation on the right side (inserting the formula $\neg K[!p]q$).

These two new formulas also enter into Aristotelian relations with each other, and with all but two of the other formulas already present in the original hexagon. This leads to the octagon in Figure 8.6.

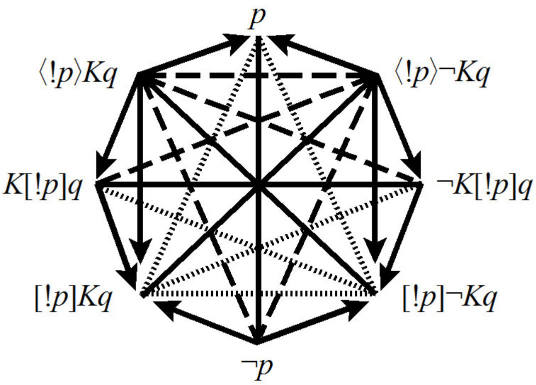
Theorem 8.3. *The octagon in Figure 8.6 is a valid Aristotelian diagram.*

Proof. Similar to the proof of Theorem 8.1. □

[25] Modulo the existence of unsuccessful formulas; see Definition 7.7 on p. 206.

[26] In the remainder of this section, I will drop agent indices, since they are not crucial.

Figure 8.6: Octagon of oppositions for the epistemic properties of public announcement



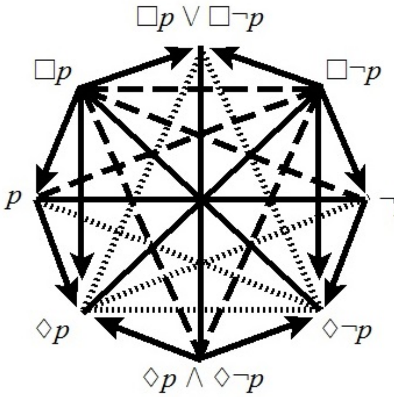
The key idea to obtain this octagon of oppositions was to split the subalternation $\langle !p\rangle(\neg)Kq \to [!p](\neg)Kq$ into two subalternations, by inserting the new formula $(\neg)K[!p]q$. Béziau (2003) applied the same idea to the Sesmat-Blanché hexagon for the S5 modal operators: the subalternation $\Box(\neg)p \to \Diamond(\neg)p$ is then split into two subalternations, by inserting the new formula $(\neg)p$. The additional Aristotelian relations that arise in this way are exactly the same as those we obtained in Theorem 8.3. In other words, the Aristotelian octagon for the epistemic properties of public announcement (Figure 8.6) is essentially the same as Béziau's Aristotelian octagon for the S5 modal operators (Figure 8.7).[27]

8.4.2 A Three-dimensional Aristotelian Diagram

In the previous subsection, I established an analogy between the Aristotelian octagon for the epistemic properties of public announcement and Béziau's octagon for modal S5. In this subsection, I will further exploit this analogy. Béziau

[27] Note that I am again implicitly subscribing to the structuralist philosophy mentioned in Subsection 8.3.3: two Aristotelian diagrams are said to be 'essentially the same', because they have the same configuration of Aristotelian relations. The formulas in both structures are concerned with different topics (public announcements/necessity), and even on a higher level of abstraction they differ significantly: in Figure 8.6 the formulas in the upper corners of the original square 'inside' the octagon are *existential* ($\langle !p\rangle(\neg)Kq$), whereas in Figure 8.7 they are *universal* ($\Box(\neg)p$).

Figure 8.7: Octagon of oppositions for the S5 modal operators



(2003)'s initial results on S5 have been generalized by Moretti (2009a), Smessaert (2009), and others. I will now show that these generalizations can perfectly be transferred from their original context (S5) to the context of public announcement logic.

The first thing to note is that the Aristotelian octagon that was constructed in the previous subsection is not Boolean closed.

Lemma 8.5. *The Aristotelian diagram in Figure 8.6 is not Boolean closed.*

Proof. Consider, for example, the conjunction of $\neg p$ and $K[!p]q$. Just like in the proof of Lemma 8.4, one can show that this conjunction is consistent, but not equivalent to any of the formulas appearing in the octagon in Figure 8.6. □

In Section 8.3, I showed that the Aristotelian square for public announcement logic (Figure 8.2) is not Boolean closed (Lemma 8.4), and then immediately went on to construct its Boolean closure (Theorem 8.2). Now, however, we will proceed in a more indirect way. First, we construct the Boolean closure of the *set of formulas* appearing in the octagon (so not of the octagon itself, *qua* Aristotelian diagram).

Theorem 8.4. *Let F be the set containing the eight formulas appearing in the Aristotelian octagon in Figure 8.6. The Boolean closure of F is the fourteen-element set F', which has the following elements:*[28]

1. p,
2. $\langle !p\rangle Kq$,
3. $K[!p]q$,
4. $[!p]Kq$,
5. $\neg p$,
6. $\langle !p\rangle \neg Kq$,
7. $\neg K[!p]q$,
8. $[!p]\neg Kq$,
9. $\neg p \wedge K[!p]q$,
10. $\neg p \wedge \neg K[!p]q$,
11. $p \vee \neg K[!p]q$,
12. $p \vee K[!p]q$,
13. $[!p]Kq \wedge (p \vee \neg K[!p]q)$,
14. $[!p]\neg Kq \wedge (p \vee K[!p]q)$.

Proof. Note that $F \subseteq F'$. The following tables show that F' is Boolean closed:

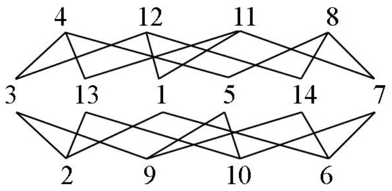
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$\neg$	5	8	7	6	1	4	3	2	11	12	9	10	14	13

$\wedge$	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	2	2	2	$\bot$	6	6	6	$\bot$	$\bot$	1	1	2	3
2		2	2	2	$\bot$	$\bot$	$\bot$	$\bot$	$\bot$	$\bot$	2	2	2	$\bot$
3			3	3	9	$\bot$	$\bot$	9	9	$\bot$	2	3	2	9
4				4	5	$\bot$	10	5	9	10	13	3	13	9
5					5	$\bot$	10	5	9	10	10	9	10	9
6						6	6	6	$\bot$	$\bot$	6	6	$\bot$	6
7							7	7	$\bot$	10	7	6	10	6
8								8	9	10	7	14	10	14
9									9	$\bot$	$\bot$	9	$\bot$	9
10										10	10	$\bot$	10	$\bot$
11											11	1	13	6
12												12	13	14
13													13	$\bot$
14														14

Furthermore, it is a tedious but easy exercise to show that there is no set of formulas X such that $F \subseteq X \subset F'$ and X is Boolean closed. □

[28] In the remainder of this subsection, I will often refer to the formulas in F' using the numbers that are given here. For example, I will refer to the formula $\neg p \wedge \neg K[!p]q$ using the number 10.

Figure 8.8: Hasse diagram for the fourteen formulas in F'



We have constructed the Boolean closure F' of the set F of formulas appearing in the Aristotelian octagon in Figure 8.6. By Definition 8.7, the formulas in F' are exactly the formulas that should appear in the Boolean closure of this octagon (*qua* Aristotelian diagram). The question now arises how to 'organize' these formulas into a helpful and visually elegant Aristotelian diagram.

To get some inspiration, let us first construct a Hasse diagram for F': see Figure 8.8. This Hasse diagram is itself not an Aristotelian diagram, because it does not display all of the Aristotelian relations between the fourteen formulas in F' (it only represents some of the subalternation relations). A tedious but easy exercise leads to the following table, which lists the 79 Aristotelian relations that hold between any of the fourteen formulas in F'.[29]

[29] 'CD' stands for contradiction, 'C' stands for contrariety, 'SC' stands for subcontrariety, and '↙ / ↗' stand for subalternation. In particular, '↙' says that there is a subalternation from the column-formula to the row-formula, and vice versa for '↗'. Finally, 'NO' says that there is no Aristotelian relation at all between the row- and column-formulas; the significance of these cases will be discussed in much more detail in Chapter 9 (in particular, see Theorem 9.8 on p. 281).

	2	3	4	5	6	7	8	9	10	11	12	13	14
1	↙	NO	SC	CD	↙	NO	SC	C	C	↗	↗	NO	NO
2		↗	↗	C	C	C	CD	C	C	↗	↗	↗	C
3			↗	NO	C	CD	SC	↙	C	SC	↗	NO	NO
4				↙	CD	SC	SC	↙	↙	SC	SC	↙	SC
5					C	NO	↗	↙	↙	SC	SC	NO	NO
6						↗	↗	C	C	↗	↗	C	↗
7							↗	C	↙	↗	SC	NO	NO
8								↙	↙	SC	SC	SC	↙
9									C	CD	↗	C	↗
10										↗	CD	↗	C
11											SC	↙	SC
12												SC	↙
13													CD

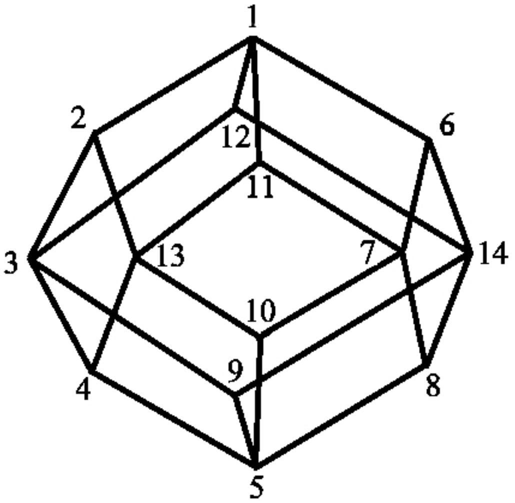
Let us take stock. The aim is to construct the Boolean closure of the Aristotelian octagon in Figure 8.6. We already know the formulas of this Boolean closure: they are the fourteen elements of F'. We also already know the 79 Aristotelian relations which hold between these fourteen formulas: they are listed in the table above.

We are now ready to apply Smessaert's (2009) results, which Moretti (2009a, p. 217) calls "simply *brilliant*". Reconsider the Hasse diagram for F' (Figure 8.8). Smessaert's results say that this Hasse diagram can be transformed into a *rhombic dodecahedron*. This is a three-dimensional structure, which has twelve rhombus-shaped faces, and (of course) fourteen vertices. Figure 8.9 shows this dodecahedron; however, it does *not* show the dodecahedron *qua* Aristotelian diagram, because it does not show the 79 Aristotelian relations holding between the fourteen vertices (for reasons of visual clarity).

Theorem 8.5. *The Boolean closure of the Aristotelian octagon in Figure 8.6 is the Aristotelian rhombic dodecahedron in Figure 8.9.*

Proof. The formulas appearing in the dodecahedron are the elements of F', and so they are certainly the right ones, because of Theorem 8.4. That this dodecahedron completely represents the Aristotelian relations between these fourteen formulas follows from the fact that the six strong Sesmat-Blanché hexagons i – vi in Figure 8.10 can be embedded inside the dodecahedron: each of the 79

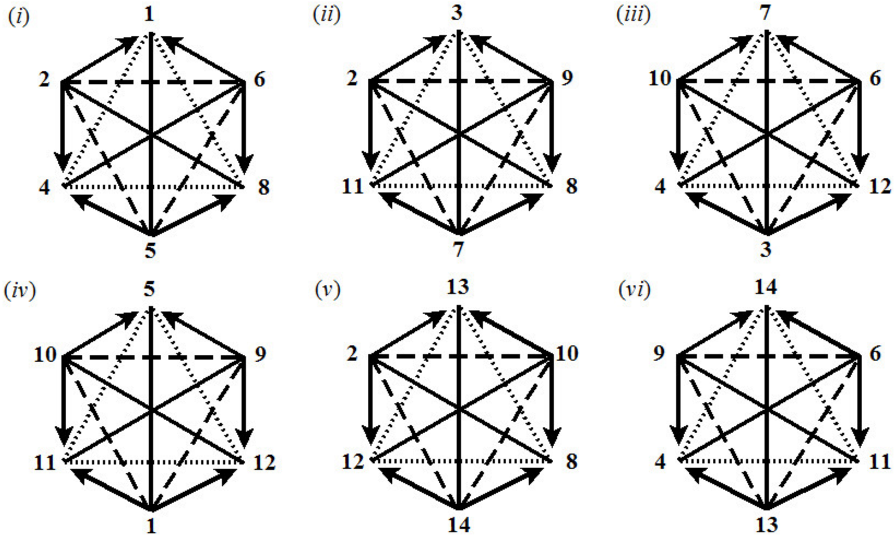
Figure 8.9: Rhombic dodecahedron of oppositions for the epistemic properties of public announcement



Aristotelian relations appears in at least one of these hexagons, and thus in the dodecahedron. □

The six Aristotelian hexagons in Figure 8.10 that are used in the proof are the PAL-analogues of six 'original' Aristotelian hexagons for S5. This analogy illustrates the progress logical geometry has been making over the past few years. Béziau (2003) already constructed hexagons i, ii and iii (for S5). Subsequently, hexagon iv (for S5) was constructed independently (and through different ways of reasoning) by Moretti and Smessaert. Finally, hexagons v and vi (for S5) were constructed by Smessaert (2009), who also proved that no other strong Sesmat-Blanché hexagons besides these six can be embedded inside the rhombic dodecahederon. In this chapter, (the PAL-analogues of) these six hexagons were constructed in one fell swoop. This illustrates how we have been able to exploit the 'initial' analogy between the Aristotelian octagon for the epistemic properties of public announcement (Figure 8.6) and Béziau's Aristotelian octagon for the S5 modal operators (Figure 8.7).

Figure 8.10: Six hexagons that are embeddable in the rhombic dodecahedron



8.4.3 A Comparison with Smessaert's Results

As was emphasized throughout the previous subsection, the construction of the Aristotelian rhombic dodecahedron for the epistemic properties of public announcement (Figure 8.9) is essentially an application of the techniques developed by Smessaert (2009) in the context of S5. It should thus not be surprising that we obtain the same 'end results', viz. the Aristotelian rhombic dodecahedra for S5 and public announcement logic. This also means that new results on the rhombic dodecahedron for S5 also apply to the rhombic dodecahedron for public announcement logic. For example, it turns out that next to the six strong Sesmat-Blanché hexagons mentioned above, many more Aristotelian diagrams (smaller as well as larger ones) can be embedded inside the rhombic dodecahedron. In ongoing work with Smessaert, I am developing an exhaustive typology of these Aristotelian diagrams, defining various 'types' (for example, one type is the strong Sesmat-Blanché hexagon) and exploring how many instances of each type can be embedded inside the rhombic dodecahedron (Smessaert and Demey 2013a). Although many of these results were initially obtained for S5,

Figure 8.11: Modal graph for S5

$$\begin{tikzcd}
\Box p \arrow[d] & \Box\neg p \arrow[d] \\
p \arrow[d] & \neg p \arrow[d] \\
\Diamond p & \Diamond\neg p
\end{tikzcd}$$

because of the analogy discussed above, they automatically 'carry over' to public announcement logic.

Despite the similarities, there are also some differences between Smessaert's original results on S5 and the present results on public announcement logic. The first difference is that the dodecahedron for public announcement logic does *not* arise from a modal graph. Smessaert's dodecahedron for S5 arises from the modal graph for S5 (Figure 8.11), which represents the six non-equivalent modalities expressible in S5 (including the 'naked' modalities). Similarly, the hexagon for the structural properties of public announcement (Figure 8.3) can be seen as arising from the modal graph for the public announcement operators.[30] The rhombic dodecahedron for public announcement logic, however, is based on Lemma 8.2, which talks about the interaction between public announcement and knowledge, rather than about the non-equivalent public announcement operators expressible in public announcement logic. This shows that there are Aristotelian diagrams that do not arise from a modal graph, and thus loosens the connection between logical geometry and modal graphs.

The most important difference between the current results and those of Smessaert, however, is the difference in *complexity*.[31] Smessaert constructed an Aris-

[30] More generally: the Aristotelian hexagon for any partially functional process π (Figure 8.4) can be seen as arising from the modal graph for the dynamic operators representing π.

[31] The term 'complexity' is used in an intuitive sense here (as is explained in the main text). It is well-known that from the perspective of *computational complexity*, PAL is equally complex as S5: the satisfiability problem of both S5 and PAL is NP-complete in the single-agent case,

totelian rhombic dodecahedron for $\mathsf{S5}$, and is thus concerned with *one* modality ($\Box\cdot/\Diamond\cdot$), which is *unary*. The Aristotelian rhombic dodecahedron for the epistemic properties of public announcements, however, is concerned with *two* modalities: the knowledge operator (K) and the public announcement operator ($[!\cdot]\cdot$). The former is also unary, but the latter is *binary*: if one has a public announcement operator $[!\cdot]\cdot$, then one needs to supply two formulas φ and ψ to obtain a well-formed formula $[!\varphi]\psi$. Hence, according to both parameters (number of modalities/arity of the modalities), public announcement logic is more complex than $\mathsf{S5}$.

These differences exhibit the power and wide applicability of Smessaert's techniques. Although Smessaert initially applied them to obtain an Aristotelian rhombic dodecahedron for the 'simple' modal logic $\mathsf{S5}$ (based on the modal graph for that logic), they can also be used to obtain Aristotelian diagrams (which are not necessarily based on modal graphs) for more complex logics, such as public announcement logic.

8.5 Conclusion

In this chapter, I have studied dynamic epistemic logic, and in particular public announcement logic, using the tools of logical geometry. After giving a brief overview of public announcement logic, I constructed an Aristotelian square and hexagon for the structural properties of public announcements, and showed how to generalize them to Aristotelian diagrams for *any* partially functional process. These results support the structuralist philosophy surrounding logical geometry. Finally, I focused on the epistemic properties of public announcements, and constructed an Aristotelian octagon and rhombic dodecahedron.

PSPACE-complete in the multi-agent case without common knowledge, and EXPTIME-complete in the multi-agent case with common knowledge (Halpern and Moses 1992, Lutz 2006).

9 Logical Geometry and Information

9.1 Introduction

The Aristotelian square of oppositions is a diagram that displays four formulas, and certain logical relations holding between them. Although traditionally, it was closely associated with Aristotelian syllogistics, it can be used to study many other logical systems, and nowadays it also has applications in linguistics. In recent years, many, increasingly complex extensions of the square have been discovered and intensively studied. At first sight, there does not seem to be a fundamental difference between the Aristotelian square and its extensions. In practice, however, there is a major difference in 'popularity': while the square is nearly universally known among logicians and formal linguists, many of the larger diagrams are only known by a few specialists.

The main aim of this chapter is to argue that there is indeed a fundamental difference between the square and its extensions, viz. a difference in informativity. To do this, I will develop a formal, well-motivated account of information in (Aristotelian and other) diagrams, and then use it to show that the square is strictly more informative than many of the more complex diagrams.

The argumentation consists of four main steps. The first step is to distinguish between concrete Aristotelian *diagrams* (such as the square and its extensions) and, on a more abstract level, the Aristotelian *geometry* (the set of logical relations visualized in Aristotelian diagrams). This distinction will enable us to provide a more fine-grained analysis later on (in the fourth step).

Second, I will define two new logical geometries, viz. the *opposition* and *implication* geometries (and the corresponding types of diagrams). The Aristotelian geometry can advantageously be seen as *hybrid* between these two new geometries: they solve some problems that have traditionally been associated with the Aristotelian geometry, and they also have several independent motivations.

The third step concerns information in the opposition and implication geometries. I will adopt an account of information that is well-known in logic and semantics, viz. *information as range*, and show that it can be used to compare the informativity not only of statements (as is usually done), but also of logical relations. This yields an informativity ordering on the opposition and implication geometries. I will show that this ordering is highly intuitive, and also fits well with the structural properties of these geometries.

The fourth and final step brings everything together. I will argue that the Aristotelian square is highly informative in two successive steps. First, I will show that the Aristotelian *geometry* is informationally optimal: it is hybrid between the opposition and implication geometries not in some random manner, but exactly so as to maximize informativity. Second, within the Aristotelian geometry, I will make a further distinction between more and less informative *diagrams*, based on whether or not they contain pairs of formulas that are *unconnected* (i.e. that stand in the least informative opposition and implication relations). It turns out that such minimally informative pairs do not occur in the classical square, but do occur in some of its extensions.

This four-step argumentation is reflected in the structure of the chapter. Section 9.2 provides some historical background and examples of the Aristotelian square and its extensions; most importantly, it also introduces the geometry/diagram distinction. Section 9.3 introduces the opposition and implication geometries and discusses their various properties and motivations. Section 9.4 applies the well-known 'information as range'-perspective to the opposition and implication geometries, and discusses some advantages of this application. Section 9.5 shows that the Aristotelian geometry is hybrid between the opposition and implication geometries in an informationally optimal way; it also introduces the notion of unconnectedness and studies in which Aristotelian diagram it occurs. Finally, Section 9.6 wraps things up and suggests some questions for further research.[1]

[1] For the sake of readability, some technical remarks and results have been placed in a separate appendix (see p. 289ff.); they are not essential for the main line of argumentation.

9.2 The Aristotelian Square of Oppositions

This section introduces the Aristotelian square of oppositions. Subsection 9.2.1 defines the Aristotelian geometry and its diagrams, and provides some examples of the square in various logical systems, while Subsection 9.2.2 discusses some extensions to larger diagrams, such as hexagons and octagons. Building on this concise overview, Subsection 9.2.3 raises the main issue that will be addressed in this paper.

9.2.1 A Brief History of the Aristotelian Square

The Aristotelian square of oppositions has a rich tradition, originating—together with the discipline of logic itself—in Aristotle's logical works. It has been studied by some of the most distinguished scholars in the history of logic, such as Avicenna (Chatti 2012), John Buridan (Hughes 1987, Read 2012a), Boole and Frege (Peckhaus 2012).[2] Contemporary logicians too have found it worthwile to show that the logics they are studying give rise to square-like structures. Typical examples include the construction of squares for modal logic (Fitting and Mendelsohn 1998, Carnielli and Pizzi 2008), intuitionistic and linear logic (Mélès 2012), epistemic logic (Lenzen 2012), deontic logic (Moretti 2009b, McNamara 2010) and temporal logic (Rini and Cresswell 2012). Applications of the square to natural language have been explored by linguists such as van der Auwera (1996), Horn (1989, 2012) and Seuren (2010, 2012b,a).

Formally speaking, we will take the Aristotelian square to be a concrete *diagram* that visualizes an underlying abstract *geometry*, i.e. a set of logical relations between formulas (relative to some background logical system S).[3]

Definition 9.1 (Aristotelian geometry)**.** Let S be a logical system, which is assumed to have connectives expressing classical negation ($\neg$), conjunction ($\wedge$) and implication ($\rightarrow$),[4] and a model-theoretic semantics. Let $\mathcal{L}_\mathsf{S}$ be the language of S.

[2] For a more exhaustive historical overview, see Parsons (2012) and Seuren (2010, Chapter 5).

[3] The term 'Aristotelian' is used in a strictly technical sense here, to distinguish the Aristotelian geometry and its diagrams from other kinds of geometries and diagrams that will be introduced later. For a detailed account of the historical origins of the square (and the crucial role of Apuleius), see Londey and Johanson (1984).

[4] It is well-known that in the presence of classical negation, each of $\wedge$ and $\rightarrow$ can be defined in terms of the other: $\varphi \rightarrow \psi = \neg(\varphi \wedge \neg\psi)$, and $\varphi \wedge \psi = \neg(\varphi \rightarrow \neg\psi)$. It does not matter for Definition 9.1 whether both of these connectives are taken as primitive, or one of them is defined

The *Aristotelian relations for* S are defined as follows: the formulas $\varphi, \psi \in \mathcal{L}_\mathsf{S}$ are

$$\begin{array}{llllll}
\mathsf{S}\text{-}\textit{contradictory} & \text{iff} & \mathsf{S} \models \neg(\varphi \wedge \psi) & \text{and} & \mathsf{S} \models \neg(\neg\varphi \wedge \neg\psi), \\
\mathsf{S}\text{-}\textit{contrary} & \text{iff} & \mathsf{S} \models \neg(\varphi \wedge \psi) & \text{and} & \mathsf{S} \not\models \neg(\neg\varphi \wedge \neg\psi), \\
\mathsf{S}\text{-}\textit{subcontrary} & \text{iff} & \mathsf{S} \not\models \neg(\varphi \wedge \psi) & \text{and} & \mathsf{S} \models \neg(\neg\varphi \wedge \neg\psi), \\
\textit{in } \mathsf{S}\text{-}\textit{subalternation} & \text{iff} & \mathsf{S} \models \varphi \rightarrow \psi & \text{and} & \mathsf{S} \not\models \psi \rightarrow \varphi.
\end{array}$$

The *Aristotelian geometry for* S is the set $\mathcal{AG}_\mathsf{S} = \{CD, C, SC, SA\}$ of the four Aristotelian relations for S (the abbreviations stand for contradiction, contrariety, subcontrariety and subalternation, respectively).

When the system S is clear from the context, I will often leave it implicit, and simply talk about 'contrary' instead of 'S-contrary', etc. Intuitively, the first three relations—CD, C and SC—are defined in terms of whether the formulas can be true together (the $\varphi \wedge \psi$ part) and whether they can be false together (the $\neg\varphi \wedge \neg\psi$ part);[5] the fourth relation—SA—is defined in terms of truth propagation.[6]

Definition 9.2 (Aristotelian diagrams)**.** Let S be a logical system as in Definition 9.1. An *Aristotelian diagram for* S is a diagram that visualizes an edge-labeled graph G. The vertices of G are contingent and pairwise non-equivalent formulas $\varphi_1, \dots, \varphi_n \in \mathcal{L}_\mathsf{S}$;[7] the edges of G are labeled by the Aristotelian relations between those formulas, i.e. if φ_i and φ_j stand in any Aristotelian relation, then this is visualized in the diagram, according to the code in Figure 9.1.[8]

Note that Definition 9.2 allows only *contingent* and *pairwise non-equivalent* formulas to appear in Aristotelian diagrams. The first reason for these restrictions is of a historical nature: classically, squares of oppositions only contained non-equivalent contingencies. More importantly, although the Aristotelian *geometry*

in terms of the other. I will return to the interdefinability of $\wedge$ and $\rightarrow$ in Subsection 9.3.3.

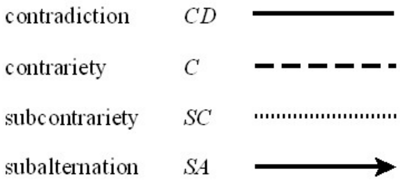
[5] It is well-known that $\neg(\neg\varphi \wedge \neg\psi)$ is equivalent to $\varphi \vee \psi$, but I prefer to stick with the first notation, because it more clearly expresses the idea of φ and ψ being false together.

[6] It should be clear that the Aristotelian relations are *not* defined in terms of properties of the formulas they relate, such as quantity and quality, as is done in many historical studies on Aristotelian logic (Parry and Hacker 1991); also see the discussion in Subsection 8.3.3.

[7] So $\mathsf{S} \not\models \varphi_i$, $\mathsf{S} \not\models \neg\varphi_i$, and $\mathsf{S} \not\models \varphi_i \leftrightarrow \varphi_j$, for $1 \leq i \neq j \leq n$.

[8] It follows immediately from Definition 9.1 that the first three relations are symmetric, and are therefore represented in Figure 9.1 by lines without arrows. We represent φ and ψ being in subalternation by means of an arrow going from φ to ψ, classically referred to as the 'superaltern' and 'subaltern', respectively.

Figure 9.1: Code for visually representing the Aristotelian relations



perfectly allows non-contingencies to enter into multiple Aristotelian relations with other formulas,[9] those relations will be vacuous and visualizing them would needlessly mess up the *diagrams* (Sanford 1968). Furthermore, the restriction to pairwise non-equivalent formulas shows that Aristotelian diagrams are essentially semantic entities: like Hasse diagrams, they represent formulas only up to logical equivalence.[10]

The most prototypical Aristotelian diagrams are those which have exactly four vertices, better known as the Aristotelian squares. Figure 9.2 shows three such Aristotelian squares for fragments of (a) classical propositional logic (CPL), (b) the modal logic S5, and (c) the deontic logic KD.[11] For example, $p \wedge q$ and $p \vee q$ are in CPL-subalternation (CPL $\models (p \wedge q) \rightarrow (p \vee q)$ and CPL $\not\models (p \vee q) \rightarrow (p \wedge q)$), $\Box p$ and $\Box\neg p$ are S5-contrary (S5 $\models \neg(\Box p \wedge \Box\neg p)$ and S5 $\not\models \neg(\neg\Box p \wedge \neg\Box\neg p)$), and $\mathsf{P}p$ and $\mathsf{P}\neg p$ are KD-subcontrary (KD $\not\models \neg(\mathsf{P}p \wedge \mathsf{P}\neg p)$ and KD $\models \neg(\neg\mathsf{P}p \wedge \neg\mathsf{P}\neg p)$). The modal logic S5 will be used as a running example throughout this chapter.[12]

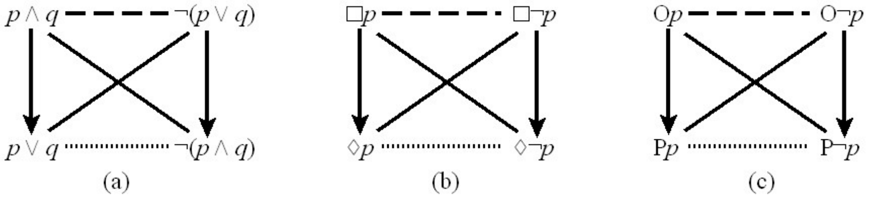
[9] Tautologies are subaltern and subcontrary to any contingent formula. Conversely, contradictions are superaltern and contrary to any contingent formula.

[10] For a more detailed discussion of the connection between Aristotelian diagrams and Hasse diagrams, see Smessaert (2009) and Demey and Smessaert (2013a).

[11] The operators O and P in the deontic square stand for ‘obligatory’ and ‘permitted’, respectively.

[12] Note that I will not consider squares for the quantifiers, and thus sidestep the notoriously difficult issue of *existential import* (Chatti and Schang 2013, Parsons 2012, Read 2012b, Seuren 2012b), since the informativity account to be developed here is entirely independent of it.

Figure 9.2: Aristotelian squares for (a) CPL, (b) the modal logic S5 and (c) the deontic logic KD.



9.2.2 Extensions of the Aristotelian Square

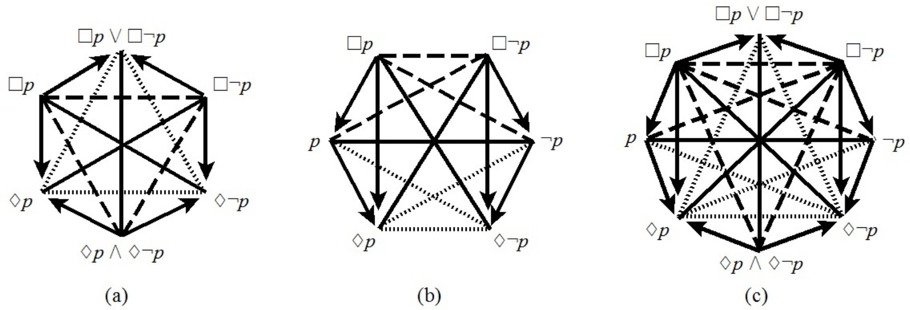
It should be noted that Definition 9.2 does not require an Aristotelian diagram to have only 4 formulas as vertices. Unsurprisingly, then, there have been several proposals throughout history to extend the Aristotelian square to more complex diagrams. The most widely known extension is the Aristotelian hexagon proposed by Jacoby (1950, 1960), Sesmat (1951) and Blanché (1952, 1966). A different Aristotelian hexagon was proposed by Czeżowski (1955), although it was already known by the 13th-century logician William of Sherwood (Khomskii 2012). While the first type of hexagon is Boolean closed, the second one is not; these, and other, differences are studied in Smessaert (2012b). Further two-dimensional generalizations include the octagons described by Béziau (2003) and Seuren (2010). Figure 9.3 shows two hexagons and an octagon for the modal logic S5.[13]

In recent years, even further generalizations have been proposed, moving from the two-dimensional to the three-dimensional realm. For example, Moretti (2009a) and Chatti and Schang (2013) study two types of Aristotelian cubes, while Smessaert (2009) describes an Aristotelian rhombic dodecahedron for the modal logic S5. In Chapter 8 of this thesis, Smessaert's results were generalized, thus obtaining a rhombic dodecahedron for public announcement logic. Other, related Aristotelian diagrams have been studied by Sauriol (1968) and Moretti (2009a, 2012a).

In ongoing work, Smessaert and I are developing an exhaustive typology of Aristotelian diagrams, which allows us to classify all the diagrams mentioned

[13] Note that the octagon in Figure 9.3(c) can be seen as the 'sum' of the hexagons in (a) and (b).

Figure 9.3: (a) Sesmat-Blanché hexagon, (b) Sherwood-Czeżowski hexagon, and (c) Béziau octagon for S5.



above (and many others), and to study their various interrelationships (Smessaert and Demey 2013a). The informational account that is developed in this chapter, however, is conceptually prior to this typology, so I will not go into it any further.

9.2.3 The Success of the Aristotelian Square

From a theoretical perspective, there does not seem to be any fundamental difference between the Aristotelian square and its extensions: both the square and its extensions are just examples of Aristotelian diagrams (cf. Definition 9.2). In practice, however, there is a major difference in popularity: while the square is nearly universally known among logicians and formal linguists, many of the larger diagrams are only known by a few specialists studying them.[14] This might partially be explained by the relative recency of their discovery; however, even the hexagons that were already being investigated in the 1950s have never been able to attract much attention (despite having various interesting properties, as shown by Smessaert 2012b).

Another explanation of the square's success is based on the intuition that this diagram is *highly informative*.[15] Unfortunately, this intuition is quite vague;

[14] This does not mean that these extensions do not have any applications at all. For example, Horn (1990) uses various hexagons to study Gricean maxims and conversational implicatures, while Jaspers (2012) uses the Sesmat-Blanché hexagon to analyze the structure of the color categories from a logical, linguistic and cognitive perspective.

[15] For example, this intuition seems to be implicit in remarks such as the following: "familiarity

for example, what does ‘informative’ mean here?—and are the larger diagrams then supposed to be *less* informative than the square? In the remainder of this chapter, however, I will argue that this intuition is essentially on the right track: I will develop a formal, well-motivated account of information in (Aristotelian and other) geometries and diagrams, and then use it to show that the square is indeed more informative than many of the more complex diagrams.

9.3 The Logical Geometries of Opposition and Implication

This section introduces two new logical geometries in addition to the classical Aristotelian geometry. Subsection 9.3.1 discusses some problems that have traditionally been associated with the Aristotelian geometry. Subsection 9.3.2 defines the opposition geometry and the implication geometry, as well as their associated diagrams. Subsection 9.3.3 shows that these geometries not only solve the problems of the Aristotelian geometry, but also have several independent motivations.

9.3.1 Problems with the Aristotelian Geometry

The Aristotelian geometry, as introduced in Definition 9.1, seems to suffer from a number of problems. For starters, this geometry does not induce a partition on the formula-pairs, and thus fails to provide a full organization of logical space. On the one hand, the Aristotelian relations are *not mutually exclusive*: as was already discussed in Subsection 9.2.1,[16] there exist pairs of formulas that simultaneously stand in two Aristotelian relations. For example, the formulas $p \wedge \neg p$ and p are both contrary and in subalternation, whereas p and $p \vee \neg p$ are both subcontrary and in subalternation.[17] On the other hand, the Aristotelian geometry is *not exhaustive* either: some pairs of formulas—for example, p and $\Diamond p \wedge \Diamond \neg p$—stand in no Aristotelian relation whatsoever. A particular subclass of such pairs

with the square is useful for logicians today as a kind of *lingua franca*, when adapted as a shorthand to express logical relations in specialized applied logics with specialized domains” (Jacquette 2012, p. 81).

[16] In particular, see Footnote 9.

[17] Note that both examples involve a non-contingent formula. This is not a coincidence: if we restrict to contingent formulas, then Lemma 8.3 on p. 223 states that the Aristotelian relations *are* mutually exclusive.

results from the Aristotelian relations' irreflexivity on contingent formulas: no contingent formula stands in any Aristotelian relation whatsoever to itself.

Most importantly, however, the Aristotelian geometry is inherently based on a certain *conceptual confusion*, which is visible in the relations' definitions and which will turn out to have far-reaching consequences. Whereas the first three Aristotelian relations (contradiction, contrariety and subcontrariety) are characterized in terms of the related formulas *possibly being true/false together*, the fourth relation (subalternation) is characterized in terms of *truth propagation*. These two notions are conceptually independent: the former is commutative (φ and ψ can be true together iff ψ and φ can be true together), whereas the latter is directional (truth is propagated *from* φ *to* ψ). The commutativity of 'together' is captured by the conjunctions in the definitions of the first three relations, which are therefore symmetrical: for $R = CD$, C and SC, we have $R(\varphi, \psi)$ iff $R(\psi, \varphi)$. By contrast, the directionality of 'propagation' is captured by the implications in the definition of the fourth relation, which is therefore asymmetrical: if $SA(\varphi, \psi)$, then not $SA(\psi, \varphi)$.

9.3.2 Defining the Opposition and Implication Geometries

I have just argued that the first three Aristotelian relations are conceptually independent from the fourth one. These three relations are all based on the idea of the related formulas possibly being true/false together. Combinatorially speaking, this idea leads to four separate cases:

1. the related formulas cannot be true together, and cannot be false together,
2. the related formulas cannot be true together, but can be false together,
3. the related formulas can be true together, but cannot be false together,
4. the related formulas can be true together, and can be false together.

The first three cases correspond exactly with the Aristotelian relations of contradiction, contrariety and subcontrariety, respectively. The fourth case, however, does not correspond with any Aristotelian relation. The relation corresponding to this case will be called *non-contradiction* (Smessaert 2009, p. 310).[18]

[18] Non-contradiction is clearly different from the Aristotelian relation of subalternation. First of all, there exist pairs of formulas—such as $(p, \neg\neg p)$ and (p, q)—which are in non-contradiction,

In the light of these observations, it is natural to remove the subalternation relation from the Aristotelian geometry, and to replace it with the non-contradiction relation. The new geometry that is thus obtained, will be called the 'opposition geometry'.[19]

Definition 9.3 (opposition geometry)**.** Let S be a logical system as in Definition 9.1. The *opposition relations for* S are defined as follows: the formulas $\varphi, \psi \in \mathcal{L}_\mathsf{S}$ are

S-*contradictory* iff $\mathsf{S} \models \neg(\varphi \wedge \psi)$ and $\mathsf{S} \models \neg(\neg\varphi \wedge \neg\psi)$,
S-*contrary* iff $\mathsf{S} \models \neg(\varphi \wedge \psi)$ and $\mathsf{S} \not\models \neg(\neg\varphi \wedge \neg\psi)$,
S-*subcontrary* iff $\mathsf{S} \not\models \neg(\varphi \wedge \psi)$ and $\mathsf{S} \models \neg(\neg\varphi \wedge \neg\psi)$,
S-*non-contradictory* iff $\mathsf{S} \not\models \neg(\varphi \wedge \psi)$ and $\mathsf{S} \not\models \neg(\neg\varphi \wedge \neg\psi)$.

The *opposition geometry for* S is the set $\mathcal{OG}_\mathsf{S} = \{CD, C, SC, NCD\}$ of the four opposition relations for S (the abbreviation NCD stands for non-contradiction).

Consider again the relation of subalternation, which we have just removed from the Aristotelian geometry to obtain the opposition geometry. This relation is based on the idea of truth propagation (entailment), with truth being propagated from the left formula (φ) to the right one (ψ), i.e. φ entails ψ, and not vice versa. Combinatorially speaking, there are four 'directions' of truth propagation:

1. φ entails ψ, and φ is entailed by ψ,
2. φ entails ψ, but φ is not entailed by ψ,
3. φ does not entail ψ, but φ is entailed by ψ,
4. φ does not entail ψ, and φ is not entailed by ψ.

but not in subalternation. Furthermore, if two contingent formulas φ and ψ are in subalternation, they will also be in non-contradiction, but that characterization would miss the key point that the truth values of φ and ψ are not independent (if φ is true, then ψ has to be true as well).

[19] 'Opposition geometry' (Definition 9.3) is a technical term, on a par with 'Aristotelian geometry' (Definition 9.1) and 'implication geometry' (Definition 9.4), and should thus not be confused with the general framework of oppositional geometry developed by Moretti (2012a); also recall Footnote 2 on p. 217. Finally, note that Definition 9.3 is similar in spirit to Moretti (2009a, 2012b) and Schang (2012b)'s 'question-answer semantics'; however, they propose this as a semantics for the *Aristotelian* geometry, and thus run into trouble when dealing with subalternation (recall Footnote 18 on the distinction between subalternation and non-contradiction).

The second case corresponds to the relation of subalternation, which will also be called left-implication (because truth is propagated from *left* to right). Continuing this naming convention, the relations corresponding to cases 1, 3 and 4 will be called bi-implication, right-implication and non-implication, respectively. Together, these four relations constitute the implication geometry:

Definition 9.4 (implication geometry)**.** Let S be a logical system as in Definition 9.1. The *implication relations for* S are defined as follows: the formulas $\varphi, \psi \in \mathcal{L}_\mathsf{S}$ are in

S-*bi-implication*	iff	$\mathsf{S} \models \varphi \to \psi$	and	$\mathsf{S} \models \psi \to \varphi$,
S-*left-implication*	iff	$\mathsf{S} \models \varphi \to \psi$	and	$\mathsf{S} \not\models \psi \to \varphi$,
S-*right-implication*	iff	$\mathsf{S} \not\models \varphi \to \psi$	and	$\mathsf{S} \models \psi \to \varphi$,
S-*non-implication*	iff	$\mathsf{S} \not\models \varphi \to \psi$	and	$\mathsf{S} \not\models \psi \to \varphi$.

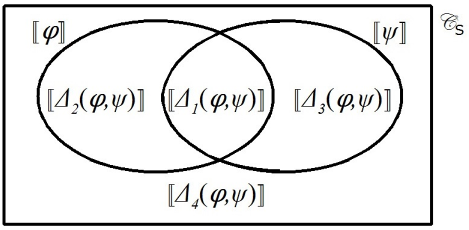
The *implication geometry for* S is the set $\mathcal{IG}_\mathsf{S} = \{BI, LI, RI, NI\}$ of the four implication relations for S (the abbreviations stand for bi-, left-, right- and non-implication, respectively).

Remark 9.1. The opposition and implication relations are all defined by means of the propositional functions $\Delta_1 - \Delta_4$:

- $\Delta_1(\varphi, \psi) := (\varphi \wedge \psi)$,
- $\Delta_2(\varphi, \psi) := (\varphi \wedge \neg\psi)$,
- $\Delta_3(\varphi, \psi) := (\neg\varphi \wedge \psi)$,
- $\Delta_4(\varphi, \psi) := (\neg\varphi \wedge \neg\psi)$.

The opposition relations are defined in terms of (the negations of) Δ_1 and Δ_4. Similarly, recalling that $\alpha \to \beta$ is equivalent to $\neg(\alpha \wedge \neg\beta)$, it should be clear that the implication relations are defined in terms of (the negations of) Δ_2 and Δ_3. Each of these functions provides a complete description of the world with respect to φ and ψ, and is thus related to Carnap (1947)'s notion of *state description*. The indices come from the canonical way of displaying a truth table for a binary, truth-functional connective $\bullet$; for example, the table's *first* row indicates the truth value of $\varphi \bullet \psi$ when φ and ψ are both true, i.e. when $\Delta_1(\varphi, \psi)$ is true. These propositional functions jointly partition logical space, as is illustrated in Figure 9.4. The opposition and implication relations holding between φ and ψ are determined by which of the regions $[\![\Delta_i(\varphi, \psi)]\!]$ are *empty*; this corresponds

Figure 9.4: The partition of logical space ($\mathcal{C}_{\mathsf{S}}$) induced by the propositions $\Delta_i(\varphi, \psi)$

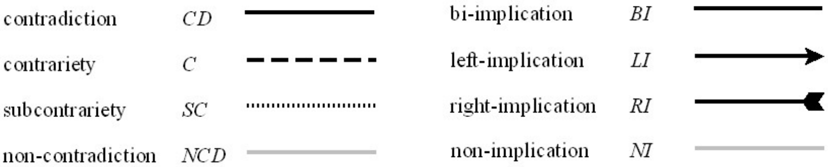


to the fact that the opposition and implication relations are defined in terms of the *negations* of Δ_i.

The opposition and implication geometries are visualized by opposition and implication diagrams, in exactly the same way as the Aristotelian geometry is visualized by Aristotelian diagrams (recall Definition 9.2).

Definition 9.5 (opposition and implication diagrams)**.** Let S be a logical system as in Definition 9.1. An *opposition diagram (resp. implication diagram) for* S is a diagram that visualizes an edge-labeled graph G. The vertices of G are contingent and pairwise non-equivalent formulas $\varphi_1, \ldots, \varphi_n \in \mathcal{L}_{\mathsf{S}}$; the edges of G are labeled by the opposition relations (resp. implication relations) between those formulas, i.e. if φ_i and φ_j stand in any opposition relation (resp. implication relation), then this is visualized in the diagram, according to the code in Figure 9.5.

Figure 9.5: Code for visually representing the opposition and implication relations



Since the two new geometries were obtained by disentangling the Aristotelian geometry, several relations occur in both the Aristotelian geometry and one of the new geometries; obviously, opposition/implication diagrams visualize these relations in the same way as Aristotelian diagrams (compare the codes in Figures 9.1 and 9.5). Furthermore, some opposition and implication relations are visualized in the same way (in particular, solid black lines for CD as well as BI, or solid grey lines for NCD as well as NI).[20] However, this should not cause any confusion, because a diagram for a given geometry visualizes only relations belonging to that geometry (for example, a solid grey line in an opposition diagram can only represent NCD). Finally, the six symmetric opposition and implication relations are represented by lines without arrows; the asymmetric relations of LI and RI are represented by arrows, with the arrow going *from* the relation's first argument *to* its second argument. Thus, $LI(\varphi, \psi)$ and $RI(\varphi, \psi)$ are visualized as $\varphi \mathrel{-\!\!\blacktriangleright} \psi$ and $\varphi \mathrel{-\!\!\blacktriangleleft} \psi$, respectively.[21]

These visual properties are illustrated in Figure 9.6, which shows an Aristotelian diagram, an opposition diagram and an implication diagram for one and the same fragment of $\mathsf{S5}$-formulas: $\{\Box p,\ \Box\neg p,\ \Diamond p\}$.

9.3.3 Motivating the New Geometries

The two geometries introduced in the previous subsection are well-motivated: not only do they solve the problems of the Aristotelian geometry (§ 9.3.3.1), they also shed new light on a number of historical issues (§ 9.3.3.2) and turn out to have various interesting formal properties (§ 9.3.3.3 and § 9.3.3.4).

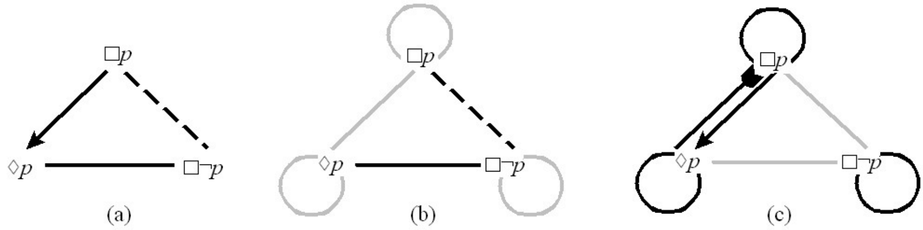
9.3.3.1 Conceptual Clarification

The opposition and implication geometries jointly solve the problems that led to their introduction in the first place, viz. the problems of the Aristotelian geometry that were discussed in Subsection 9.3.1. Both geometries induce a partition on

[20] The contrast within the opposition and implication geometries between (three kinds of) black lines on the one hand and a grey line on the other is motivated by informativity considerations that will be discussed later in the chapter.

[21] The arrow's *head* indicates the direction of truth propagation. In the case of LI, this direction matches the direction of the arrow itself, but in the case of RI, they differ. For example, $LI(\Box p, p)$ is visualized as $\Box p \mathrel{-\!\!\blacktriangleright} p$, because both the LI-relation and truth propagation go from $\Box p$ to p; however, $RI(\Diamond p, p)$ is visualized as $\Diamond p \mathrel{-\!\!\blacktriangleleft} p$, because the RI-relation goes from $\Diamond p$ to p, but truth is propagated from p to $\Diamond p$.

Figure 9.6: (a) Aristotelian diagram, (b) opposition diagram, and (c) implication diagram for an S5-fragment



the formula-pairs: it is easy to show that each pair of formulas stands in one and only one opposition relation and in one and only one implication relation.

More importantly, the conceptual confusion underlying the Aristotelian geometry is dissolved: the opposition geometry is uniformly based on the notion of 'possibly being true/false together' (its relations are defined in terms of Δ_1 and Δ_4), and the implication geometry is uniformly based on the notion of 'truth propagation' (its relations are defined in terms of Δ_2 and Δ_3).

9.3.3.2 Historical Context

The two new geometries are firmly rooted in a long-standing tradition of discussions about the conceptual confusion underlying the Aristotelian geometry. For example, already in the second century AD, Apuleius observed that the relations of contradiction, contrariety and subcontrariety are all based on the notion of 'possibly being true/false together', which he called 'pugna'. Subalternation, however, falls outside the scope of this notion: "[u]nder the truth-functional perspective of *pugnae* we learn quickly that a-i and e-o [i.e. subaltern formulas] are neither in *pugna perfecta* [CD], nor in *pugna dividua* [C/SC], but they are in no *pugna* whatsoever" (Gombocz 1990, p. 126). If Apuleius' notion of 'no *pugna* whatsoever' is viewed as non-contradiction, his *pugna*-perspective clearly anticipates the opposition geometry.

Furthermore, Correia (2012) convincingly argues that there are two complementary perspectives on the square: as a theory of negation and as a theory of logical consequence. Both perspectives have been discussed in separate textual traditions of Aristotle's work: the former is mainly found in commentaries on *De Interpretatione*, while the latter is central in commentaries on *Prior Analytics*. As Correia points out, these perspectives are based on "two kinds of logical relations that commentators distinguished in their comments on the square: relations of opposition [CD, C, SC] and relations of the parts and the whole [SA/LI]" (Correia 2012, p. 47). Hence, the negation- and consequence-perspectives (with their underlying logical relations) clearly anticipate the opposition and implication geometries, respectively.

9.3.3.3 Internal and External Structure

The two new geometries are highly structured, both internally and externally (i.e. with respect to each other). I will first discuss the geometries' internal structure. Since the opposition geometry is based on the commutative notion of 'together', its relations are all symmetric. The implication geometry, however, is based on the directional notion of 'truth propagation'; if the direction of truth propagation is reversed, the roles of left-to-right implication (LI) and right-to-left implication (RI) are changed around (the 'neutral' relations of both-way implication (BI) and neither-way implication (NI) are left untouched). This is summarized in the following lemma.

Lemma 9.1. *For all formulas $\varphi, \psi \in \mathcal{L}_\mathsf{S}$, the following hold:*

$$\begin{array}{lrcrlrcr}
\textit{1a)} & CD(\varphi,\psi) & \textit{iff} & CD(\psi,\varphi), & \textit{1b)} & BI(\varphi,\psi) & \textit{iff} & BI(\psi,\varphi), \\
\textit{2a)} & C(\varphi,\psi) & \textit{iff} & C(\psi,\varphi), & \textit{2b)} & LI(\varphi,\psi) & \textit{iff} & RI(\psi,\varphi), \\
\textit{3a)} & SC(\varphi,\psi) & \textit{iff} & SC(\psi,\varphi), & \textit{3b)} & RI(\varphi,\psi) & \textit{iff} & LI(\psi,\varphi), \\
\textit{4a)} & NCD(\varphi,\psi) & \textit{iff} & NCD(\psi,\varphi), & \textit{4b)} & NI(\varphi,\psi) & \textit{iff} & NI(\psi,\varphi).
\end{array}$$

Proof. All items follow trivially from Definitions 9.3 and 9.4. □

If we use $\mathcal{G}_\mathsf{S}$ to denote the set of all opposition and implication relations ($\mathcal{G}_\mathsf{S} := \mathcal{OG}_\mathsf{S} \cup \mathcal{IG}_\mathsf{S}$),[22] this lemma can be rephrased in a slightly more compact way. The advantages of this rephrasing will become clear later on.

Corollary 9.1. *There exists a mapping $F\colon \mathcal{G} \to \mathcal{G}$ such that for all relations $R \in \mathcal{G}$, it holds for all $\varphi, \psi \in \mathcal{L}_\mathsf{S}$ that $R(\varphi,\psi)$ iff $F(R)(\psi,\varphi)$.*

[22] Note that this set includes the original Aristotelian relations, i.e. $\mathcal{AG} \subseteq \mathcal{G}$.

Proof. The definition of F can be straightforwardly 'read off' from Lemma 9.1, i.e. put $F(CD) := CD$, $F(C) := C$, $F(SC) := SC$, $F(NCD) := NCD$, $F(BI) := BI$, $F(LI) := RI$, $F(RI) := LI$, and $F(NI) := NI$. □

Another, independent way in which the new geometries are internally structured, is that if two formulas stand in some opposition (resp. implication) relation, their negations stand in some opposition (resp. implication) relation as well. Details can be found in the following lemma and corollary.

Lemma 9.2. *For all formulas $\varphi, \psi \in \mathcal{L}_{\mathsf{S}}$, the following hold:*

1a)	$CD(\varphi,\psi)$	*iff*	$CD(\neg\varphi,\neg\psi)$,	*1b)*	$BI(\varphi,\psi)$	*iff*	$BI(\neg\varphi,\neg\psi)$,
2a)	$C(\varphi,\psi)$	*iff*	$SC(\neg\varphi,\neg\psi)$,	*2b)*	$LI(\varphi,\psi)$	*iff*	$RI(\neg\varphi,\neg\psi)$,
3a)	$SC(\varphi,\psi)$	*iff*	$C(\neg\varphi,\neg\psi)$,	*3b)*	$RI(\varphi,\psi)$	*iff*	$LI(\neg\varphi,\neg\psi)$,
4a)	$NCD(\varphi,\psi)$	*iff*	$NCD(\neg\varphi,\neg\psi)$,	*4b)*	$NI(\varphi,\psi)$	*iff*	$NI(\neg\varphi,\neg\psi)$.

Proof. All items follow trivially from Definitions 9.3 and 9.4. □

Corollary 9.2. *There exists a mapping $N12\colon \mathcal{G} \to \mathcal{G}$ such that for all relations $R \in \mathcal{G}$, it holds for all $\varphi, \psi \in \mathcal{L}_{\mathsf{S}}$ that $R(\varphi,\psi)$ iff $N12(R)(\neg\varphi,\neg\psi)$.*

Proof. As before, the definition of $N12$ can be 'read off' from Lemma 9.2. □

It should be noted that for all $R \in \mathcal{G}$, it holds that $F(N12(R)) = N12(F(R))$; we can thus define the mapping $FN12\colon \mathcal{G} \to \mathcal{G}$ by putting $FN12 := F \circ N12 = N12 \circ F$. If we use Id to note the identity mapping on $\mathcal{G}$, the internal structure of the opposition and implication geometries can be summarized as follows:

Remark 9.2. The set $\{Id, F, N12, FN12\}$ is closed under composition ($\circ$), and forms a group that acts faithfully on $\mathcal{G}$. This group is isomorphic to the Klein four-group. The separate geometries $\mathcal{OG}$ and $\mathcal{IG}$ are invariant under this group action. More details can be found in Remark 9.6 in the appendix.[23]

I have argued above that the opposition and implication geometries are conceptually independent: the former is based on the notion of 'possibly being true/false together', while the latter is based on the notion of 'truth propagation'. This does not imply, however, that there are absolutely no connections

[23] For more background on group theory, see Rotman (1995), in particular p. 55ff. and p. 345ff.

between both geometries. Consider, for example, the opposition relation of contrariety. If $C(\varphi,\psi)$, then φ and ψ cannot be true together, which is, by itself, a ‘directionless’ situation. However, we can *impose* a direction upon it, in two complementary ways:

- If the *first* formula (φ) is true, then the *second* one (ψ) has to be false (because otherwise both formulas would be true together after all).
- If the *second* formula (ψ) is true, then the *first* one (φ) has to be false (because otherwise both formulas would be true together after all).[24]

It is easy to see that these two ways of symmetry breaking correspond exactly to $LI(\varphi,\neg\psi)$ and $RI(\neg\varphi,\psi)$, and have thus taken us to the implication geometry. The following lemma lists similar ways in which oppositional facts can be expressed using implication relations, and vice versa.[25]

Lemma 9.3. *For all formulas $\varphi,\psi\in\mathcal{L}_\mathsf{S}$, the following hold:*

$$\begin{array}{lrcrlrcr}
\textit{1a)} & CD(\varphi,\psi) & \textit{iff} & BI(\neg\varphi,\psi), & \textit{1b)} & CD(\varphi,\psi) & \textit{iff} & BI(\varphi,\neg\psi),\\
\textit{2a)} & C(\varphi,\psi) & \textit{iff} & RI(\neg\varphi,\psi), & \textit{2b)} & C(\varphi,\psi) & \textit{iff} & LI(\varphi,\neg\psi),\\
\textit{3a)} & SC(\varphi,\psi) & \textit{iff} & LI(\neg\varphi,\psi), & \textit{3b)} & SC(\varphi,\psi) & \textit{iff} & RI(\varphi,\neg\psi),\\
\textit{4a)} & NCD(\varphi,\psi) & \textit{iff} & NI(\neg\varphi,\psi), & \textit{4b)} & NCD(\varphi,\psi) & \textit{iff} & NI(\varphi,\neg\psi),\\
\textit{5a)} & BI(\varphi,\psi) & \textit{iff} & CD(\neg\varphi,\psi), & \textit{5b)} & BI(\varphi,\psi) & \textit{iff} & CD(\varphi,\neg\psi),\\
\textit{6a)} & LI(\varphi,\psi) & \textit{iff} & SC(\neg\varphi,\psi), & \textit{6b)} & LI(\varphi,\psi) & \textit{iff} & C(\varphi,\neg\psi),\\
\textit{7a)} & RI(\varphi,\psi) & \textit{iff} & C(\neg\varphi,\psi), & \textit{7b)} & RI(\varphi,\psi) & \textit{iff} & SC(\varphi,\neg\psi),\\
\textit{8a)} & NI(\varphi,\psi) & \textit{iff} & NCD(\neg\varphi,\psi), & \textit{8b)} & NI(\varphi,\psi) & \textit{iff} & NCD(\varphi,\neg\psi).
\end{array}$$

Proof. All items follow trivially from Definitions 9.3 and 9.4. □

Corollary 9.3. *There exist mappings $N1, N2\colon \mathcal{G}\to\mathcal{G}$ such that for all relations $R\in\mathcal{G}$, the following holds for all $\varphi,\psi\in\mathcal{L}_\mathsf{S}$:*
a) $R(\varphi,\psi)$ *iff* $N1(R)(\neg\varphi,\psi)$,
b) $R(\varphi,\psi)$ *iff* $N2(R)(\varphi,\neg\psi)$.

[24] These facts were already known by the 12th-century logician Peter of Spain, who called them the ‘law of contraries’ (Horn 2010).

[25] Lemma 9.3 consists of an *a*- and a *b*-series, which describe the effects of negating the *first*, resp. the *second* argument of a given relation. The symmetry breaking/creating required to connect the opposition and implication geometries is manifested in the fact that *exactly one* argument is negated. This is to be contrasted with Lemma 9.2, in which *both* arguments are negated, and the geometries are kept apart (opposition relations are connected with opposition relations, implication relations with implication relations).

Proof. The definitions of $N1$ and $N2$ can be 'read off' from the a- and b-series of items, respectively, in Lemma 9.3. □

These mappings $N1$ and $N2$ are obviously related to the mapping $N12$ defined above: $N12 = N1 \circ N2 = N2 \circ N1$. If we define two additional mappings $FN1, FN2\colon \mathcal{G} \to \mathcal{G}$ by $FN1 = F \circ N1$ and $FN2 = F \circ N2$, the close relationship between the opposition and implication geometries can be summarized as follows:

Remark 9.3. The set $\{Id, N1, N2, N12, F, FN1, FN2, FN12\}$ is closed under composition, and forms a group that acts faithfully on $\mathcal{G}$. This group is isomorphic to the dihedral group of order 8. More details can be found in Remark 9.7 in the appendix.

It should be emphasized that the rich structure of the opposition and implication geometries does not primarily consist in the *individual* items of Lemmas 9.1–9.3, but rather in the fact that they *interact* with each other in interesting ways. These interactions can concisely be described using the language of group theory, as illustrated in Remarks 9.2–9.3 and Remarks 9.6–9.7.

9.3.3.4 Geometries and Connectives

There are 4 opposition relations and 4 implication relations, and thus $4 \times 4 = 16$ possible combinations of an opposition and an implication relation. On the other hand, it is well-known that there are $2^4 = 16$ binary, truth-functional connectives (Enderton 2001, pp. 50–51). I will now show that this numerical equality is not a coincidence, because there exists a canonical correspondence between pairs of opposition and implication relations and binary connectives.

Each binary, truth-functional connective $\bullet$ can be identified with its truth table, i.e. with the 4-tuple $(\bullet_1, \bullet_2, \bullet_3, \bullet_4) \in \{0,1\}^4$, where $\bullet_i$ is the truth value of the formula $\varphi \bullet \psi$ on row i, i.e. in case $\Delta_i(\varphi, \psi)$ is true. Formally, this means that

$$\text{if } \bullet_i = 1, \text{ then } \mathsf{S} \models \Delta_i(\varphi,\psi) \rightarrow (\varphi \bullet \psi), \tag{9.1}$$

$$\text{if } \bullet_i = 0, \text{ then } \mathsf{S} \models \Delta_i(\varphi,\psi) \rightarrow \neg(\varphi \bullet \psi). \tag{9.2}$$

For example, conjunction is $\wedge = (1,0,0,0)$, while (inclusive) disjunction is $\vee = (1,1,1,0)$. This identification between connectives and their truth tables is used in the following definition:[26]

Definition 9.6. Given an opposition relation $R \in \mathcal{OG}$ and an implication relation $S \in \mathcal{IG}$, we define the binary, truth-functional connective $\bullet^{(R,S)}$ by putting, for $1 \leq i \leq 4$:

$$\bullet_i^{(R,S)} := \begin{cases} 0 & \text{if for all } \varphi, \psi \in \mathcal{L}_{\mathsf{S}} \text{ such that } R(\varphi,\psi) \text{ and } S(\varphi,\psi)\colon \models \neg\Delta_i(\varphi,\psi), \\ 1 & \text{if there exist } \varphi, \psi \in \mathcal{L}_{\mathsf{S}} \text{ such that } R(\varphi,\psi) \text{ and } S(\varphi,\psi)\colon \not\models \neg\Delta_i(\varphi,\psi). \end{cases}$$

As noted in Remark 9.1, the opposition relation R is defined in terms of $\neg\Delta_1$ and $\neg\Delta_4$, and thus determines the values of $\bullet_1^{(R,S)}$ and $\bullet_4^{(R,S)}$; similarly, the implication relation S is defined in terms of $\neg\Delta_2$ and $\neg\Delta_3$, and thus determines the values of $\bullet_2^{(R,S)}$ and $\bullet_3^{(R,S)}$. In total, the pair $(R,S) \in \mathcal{OG} \times \mathcal{IG}$ yields the connective $\bullet^{(R,S)}$. For example, (SC, NI) yields the connective $\bullet^{(SC,NI)} = (1,1,1,0) = \vee$, and (SC, LI) yields the connective $\bullet^{(SC,LI)} = (1,0,1,0)$.

Definition 9.6 thus associates each pair of an opposition relation R and an implication relation S with a truth-functional, binary connective $\bullet^{(R,S)}$. It is easy to see that this mapping $(R,S) \longmapsto \bullet^{(R,S)}$ is a bijection:

- it is *injective*: for all opposition relations R, R' and implication relations S, S', $\bullet^{(R,S)} = \bullet^{(R',S')}$ implies that $R = R'$ and $S = S'$,
- it is *surjective*: for every binary, truth-functional connective $\bullet$, there exist an opposition relation R and an implication relation S such that $\bullet = \bullet^{(R,S)}$.

The mere existence of a bijection between $\mathcal{OG} \times \mathcal{IG}$ and the set of all truth-functional, binary connectives should come as no surprise, since we already knew that both sets have the same cardinality (viz. 16). Theorem 9.1 below states that the bijection described in Definition 9.6 is canonical, and thus provides a positive answer to the question whether "each [binary] logical connective corresponds to a relation of opposition" (Schang 2012a, p. 152)—at least, if Schang's 'relation of opposition' is re-interpreted as a 'pair of an opposition relation and an implication relation'.

[26] Definition 9.6 might look cumbersome, because it involves quantifying over formulas. However, it follows immediately from Definitions 9.3–9.4 that if $R(\varphi,\psi)$, $S(\varphi,\psi)$, $R(\varphi',\psi')$ and $S(\varphi',\psi')$, then for $1 \leq i \leq 4$: $\models \neg\Delta_i(\varphi,\psi) \Leftrightarrow \models \neg\Delta_i(\varphi',\psi')$. This shows that the quantification over formulas in Definition 9.6 is 'innocent': if there exists *at least one* pair of formulas (φ,ψ) standing in the relations R and S for which it holds that $\models \neg\Delta_i(\varphi,\psi)$, then this holds for *all* such pairs of formulas.

Theorem 9.1. *Consider an opposition relation $R \in \mathcal{OG}$ and an implication relation $S \in \mathcal{IG}$. Then for all formulas $\varphi, \psi \in \mathcal{L}_{\mathsf{S}}$, the following holds:*

$$\textit{if } R(\varphi,\psi) \textit{ and } S(\varphi,\psi), \textit{ then } \mathsf{S} \models \varphi \bullet^{(R,S)} \psi.$$

Proof. Let $\varphi, \psi \in \mathcal{L}_{\mathsf{S}}$ be arbitrary formulas and suppose that $R(\varphi,\psi)$ and $S(\varphi,\psi)$. Let $\mathbb{M}$ be an arbitrary model (of the semantics of the system S); we will show that $\mathbb{M} \models \varphi \bullet^{(R,S)} \psi$. By definition of the propositional functions Δ_i, there exists exactly one $i \in \{1,2,3,4\}$ such that $\mathbb{M} \models \Delta_i(\varphi,\psi)$. Hence $\not\models \neg\Delta_i(\varphi,\psi)$, and thus it follows by Definition 9.6 that $\bullet_i^{(R,S)} = 1$. Given the connection between a connective and its truth table—as formally expressed by (9.1) and (9.2)—, it thus follows that $\models \Delta_i(\varphi,\psi) \to (\varphi \bullet^{(R,S)} \psi)$. Hence, $\mathbb{M} \models \Delta_i(\varphi,\psi)$ entails that $\mathbb{M} \models \varphi \bullet^{(R,S)} \psi$. □

Consider, for example, the relations SC and NI, and recall that $\bullet^{(SC,NI)} = (1,1,1,0) = \vee$. Theorem 9.1 now states that for any formulas φ, ψ standing in these relations, it holds that $\models \varphi \vee \psi$. For another example, consider the relations SC and LI, and recall that $\bullet^{(SC,LI)} = (1,0,1,0)$; Theorem 9.1 now states that for any formulas φ, ψ standing in these relations, it holds that $\models \psi$.[27]

The correspondence established above is certainly not the only connection between the binary, truth-functional connectives and logical geometry. For example, several authors have noted that these connectives can be used to decorate a rhombic dodecahedron (Zellweger 1997, Kauffman 2001) and related diagrams (Sauriol 1968, Luzeaux et al. 2008, Moretti 2009a, Dubois and Prade 2012). Such diagrams visualize the Aristotelian relations that hold among the 16 propositions of the form $p \bullet q$, where $\bullet$ is a binary, truth-functional connective. Hence, the connectives appear at the object level: they are (inside) the *relata*, i.e. the concrete formulas standing in the Aristotelian relations. Theorem 9.1, however, is of a fundamentally different nature, because it operates on the metalevel: it does not link the connectives with the relata of the opposition and implication relations, but rather with these *relations themselves*.

Additionally, Theorem 9.1 immediately leads to Theorem 9.2 below, which states that *contingent* formulas can stand in only 7 out of the 16 combinatorially possible pairs of opposition and implication relations. This restriction will turn out to have a number of applications in the remainder of the chapter.

[27] Theorem 9.1 also has a partial converse, which is of less importance for the sake of our argument; more information about this converse can be found in Lemma 9.7 and Remark 9.8 in the appendix.

Theorem 9.2. *Consider arbitrary formulas $\varphi, \psi \in \mathcal{L}_{\mathsf{S}}$, and suppose that φ and ψ are contingent. Then φ and ψ stand in one of the following 7 pairs of relations:*

$$\begin{array}{lcr} (NCD, BI) & & (CD, NI) \\ (NCD, LI) & & (C, NI) \\ (NCD, RI) & & (SC, NI) \\ & (NCD, NI) & \end{array}$$

Proof. It suffices to show that φ and ψ do not stand in any of the 9 other pairs:

- φ and ψ do not stand in (CD, BI):

  For a reductio, suppose they *do* stand in those relations; since $\bullet^{(CD,BI)} = (0,0,0,0) = \bot$, it follows by Theorem 9.1 that $\models \bot$, which contradicts the consistency of S. Note that this case does not even rely on the contingency of φ and ψ.

- φ and ψ do not stand in (SC, LI):

  For a reductio, suppose they *do* stand in those relations; since $\bullet^{(SC,LI)} = (1,0,1,0)$, it follows by Theorem 9.1 that $\models \psi$, which contradicts the contingency of ψ. The cases (SC, RI), (C, LI) and (C, RI) yield the connectives $(1,1,0,0)$, $(0,0,1,1)$ and $(0,1,0,1)$, respectively, and can thus be treated analogously.

- φ and ψ do not stand in (CD, LI):

  For a reductio, suppose they *do* stand in those relations; since $\bullet^{(CD,LI)} = (0,0,1,0)$, it follows by Theorem 9.1 that $\models \neg\varphi \wedge \psi$, and hence also $\models \neg\varphi$ and $\models \psi$, which contradict the contingency of both φ and ψ. The cases (SC, BI), (CD, LI) and (C, BI) yield the connectives $(1,0,0,0)$, $(0,0,1,0)$ and $(0,0,0,1)$, respectively, and can thus be treated analogously. □

Theorems 9.1 and 9.2 connect the binary, truth-functional connectives on the one hand with pairs consisting of an opposition and an implication relation ($\mathcal{OG} \times \mathcal{IG}$) on the other. I will finish this subsection by showing that this connection generalizes the connection between the original Aristotelian relations ($\mathcal{AG}$) and their defining connectives, which was already hinted at by Bocheński (1959) and Williamson (1972).

Remark 9.4. Consider arbitrary contingent formulas $\varphi, \psi \in \mathcal{L}_{\mathsf{S}}$. If $CD(\varphi, \psi)$ or $C(\varphi, \psi)$ or $SC(\varphi, \psi)$, then it follows by Theorem 9.2 that $NI(\varphi, \psi)$. Similarly, if $SA(\varphi, \psi)$, i.e. $LI(\varphi, \psi)$, then $NCD(\varphi, \psi)$. By Definition 9.6, it holds that $\bullet^{(CD,NI)} = (0,1,1,0) = \veebar$ (exclusive disjunction), $\bullet^{(C,NI)} = (0,1,1,1) = |$ (Sheffer's stroke), $\bullet^{(SC,NI)} = (1,1,1,0) = \vee$ and $\bullet^{(NCD,LI)} = (1,0,1,1) = \rightarrow$. Then by Theorem 9.1 it follows that:

- if $CD(\varphi, \psi)$, then $\mathsf{S} \models \varphi \veebar \psi$, i.e. $\mathsf{S} \models \neg\Delta_1(\varphi, \psi)$ and $\mathsf{S} \models \neg\Delta_4(\varphi, \psi)$,
- if $C(\varphi, \psi)$, then $\mathsf{S} \models \varphi \,|\, \psi$, i.e. $\mathsf{S} \models \neg\Delta_1(\varphi, \psi)$,
- if $SC(\varphi, \psi)$, then $\mathsf{S} \models \varphi \vee \psi$, i.e. $\mathsf{S} \models \neg\Delta_4(\varphi, \psi)$,
- if $SA(\varphi, \psi)$, then $\mathsf{S} \models \varphi \rightarrow \psi$, i.e. $\mathsf{S} \models \neg\Delta_2(\varphi, \psi)$.

These entailments are entirely natural; after all, they merely express that the $\models$-parts of Definition 9.1 are necessary (but not sufficient)[28] conditions for the Aristotelian relations. Bocheński (1959, p. 14) uses these entailments to *define* the Aristotelian relations, i.e. he views them as expressing necessary *and* sufficient conditions (e.g. $SC(\varphi, \psi) :\Leftrightarrow \; \mathsf{S} \models \varphi \vee \psi$). In comparison to Definition 9.1, Bocheński's definition can thus be seen as keeping the $\models$-conditions, while leaving out the $\not\models$-conditions.[29] Obviously, both definitions are not equivalent; for example, although $SC(\varphi, \psi)$ entails that $\mathsf{S} \models \varphi \vee \psi$ according to both definitions, the converse is valid according to Bocheński's definition, but not according to our Definition 9.1 (recall Remark 9.8 about the converse of Theorem 9.1 being only partial).

9.4 Information in the Opposition and Implication Geometries

This section is an investigation into the informativity of the opposition and implication geometries. Subsection 9.4.1 introduces a well-known perspective on

[28] Of course, the $\models$- and $\not\models$-parts together *are* sufficient.

[29] Seuren (2010, p. 49) defines the Aristotelian relations in a similar way. Sanford (1968) compares the usual definition (see Definition 9.1) with that of Bocheński, and judges the former to be preferable.

informativity, called ‘information as range’. Subsection 9.4.2 applies this perspective to the opposition and implication geometries, and Subsection 9.4.3 discusses some advantages of this application.

9.4.1 Information as Range

The ‘information as range’-perspective on information is well-known in logic and formal semantics.[30] We start by associating with each statement σ a set $\mathbb{I}(\sigma)$, which is called the *information range* of σ, and whose elements are often referred to as ‘states’ or ‘possible worlds’. The ‘information as range’-perspective states that the informativity of a statement is inversely correlated with the size of its information range: “the more worlds there still are in the information range, the less information it contains” (Gamut 1991, p. 54). This inverse correlation is formally expressed by the definition of the *informativity ordering* $\leq_i$, which looks as follows:

$$\sigma \leq_i \tau \;:\Leftrightarrow\; \mathbb{I}(\sigma) \supseteq \mathbb{I}(\tau).$$

Informally, this definition states that the statement τ is at least as informative as the statement σ iff τ’s information range is a subset of σ’s information range.

Since $\supseteq$ is a partial ordering (reflexive, transitive and antisymmetric), the informativity ordering $\leq_i$ is a preordering (reflexive and transitive).[31] A strict informativity ordering $<_i$ can be defined by putting $\sigma <_i \tau :\Leftrightarrow (\sigma \leq_i \tau$ and $\tau \not\leq_i \sigma)$; this is a strict partial ordering (irreflexive and transitive) (Harel et al. 2000, pp. 6–11).

In formal semantics, the information range of a statement σ is usually identified with its *truth set*, i.e. the set of all possible worlds w (in a given model $\mathbb{M}$) that make σ true: $\mathbb{I}(\sigma) = [\![\sigma]\!]^{\mathbb{M}} = \{w \in \mathbb{M} \mid w \models \sigma\}$. Consider the following example from Löbner (2002, pp. 64–66): let σ be ‘Donald Duck is a bird’ and τ ‘Donald Duck is a duck’. Since every possible world in which Donald Duck is a duck is also a world in which he is a bird, but not vice versa, it holds that $\mathbb{I}(\tau) = [\![\tau]\!] \subseteq [\![\sigma]\!] = \mathbb{I}(\sigma)$ and $\mathbb{I}(\sigma) \not\subseteq \mathbb{I}(\tau)$, and hence the ‘information as

[30] A contemporary overview of this and other logical perspectives on information, which emphasizes their dynamic aspects, can be found in van Benthem and Martinez (2008). Demey (2012b) uses a version of the ‘information as range’-perspective to obtain a logical account of the informativity of narratives.

[31] The information ordering $\leq_i$ is not antisymmetric, because from $\sigma \leq_i \tau$ and $\tau \leq_i \sigma$ it follows that the statements σ and τ are equally informative (i.e. $\mathbb{I}(\sigma) = \mathbb{I}(\tau)$), but *not* that they are identical (i.e. not $\sigma = \tau$).

range'-perspective states that $\sigma \leq_i \tau$ and $\tau \nleq_i \sigma$, respectively, and thus $\sigma <_i \tau$. This matches the semantic judgment that the nominal predicate 'is a duck' is strictly more informative than the nominal predicate 'is a bird'.

It should be emphasized that unlike other, more quantitatively oriented theories of information (Harremoës and Topsœ 2008), the 'information as range'-perspective does not yield any *absolute* informativity judgments (of the form 'σ has informativity k', where $k \in [0, 1]$), but only *comparative* informativity judgments (of the form 'σ is at least as informative as τ' and 'σ is strictly more informative than τ'). For our current purposes, however, such comparative judgments will suffice.

9.4.2 Information in the Opposition and Implication Geometries

I will now show how the 'information as range'-perspective introduced above can be used to compare the informativity of opposition and implication relations. However, this perspective concerns the informativity of *statements* rather than *relations*. Therefore, it is first applied to statements of the form $R(\varphi, \psi)$, and subsequently, this analysis is lifted from statements about relations to the relations themselves.

For any opposition or implication relation R and formulas $\varphi, \psi \in \mathcal{L}_\mathsf{S}$, we consider the statement $R(\varphi, \psi)$, which says that φ and ψ stand in the relation R. This statement does not belong to the logic's object language ($\mathcal{L}_\mathsf{S}$), but rather to its metalanguage. Hence, it does not make sense to talk about $R(\varphi, \psi)$ being *true* in a given S-model $\mathbb{M}$; however, it does make sense to talk about $R(\varphi, \psi)$ being *compatible* with $\mathbb{M}$. Consequently, the information range of the statement $R(\varphi, \psi)$ does not consist of the models in which it is true, but rather of the models with which it is compatible.

Definition 9.7. Consider a relation $R \in \mathcal{OG} \cup \mathcal{IG}$ and formulas $\varphi, \psi \in \mathcal{L}_\mathsf{S}$. Let $\mathcal{C}_\mathsf{S}$ be the class of all models of S. Then we define:

1. a model $\mathbb{M} \in \mathcal{C}_\mathsf{S}$ is *compatible* with the statement $R(\varphi, \psi)$ iff

$$\text{for } 1 \leq i \leq 4\text{: } \Big(R(\varphi, \psi) \Rightarrow \mathsf{S} \models \neg\Delta_i(\varphi, \psi)\Big) \Longrightarrow \mathbb{M} \models \neg\Delta_i(\varphi, \psi),$$

2. the *information range* of the statement $R(\varphi, \psi)$ is

$$\mathbb{I}(R(\varphi, \psi)) := \big\{\mathbb{M} \in \mathcal{C}_\mathsf{S} \mid \mathbb{M} \text{ is compatible with } R(\varphi, \psi)\big\}.$$

An S-model $\mathbb{M}$ is thus compatible with a statement $R(\varphi,\psi)$ iff it is not a countermodel to any of the universal claims in terms of which the truth of the statement $R(\varphi,\psi)$ is defined; in other words, iff all formulas $\neg\Delta_i(\varphi,\psi)$ that $R(\varphi,\psi)$ entails to be tautological (cf. Definitions 9.3 and 9.4) are satisfied by $\mathbb{M}$. Note that for most logical systems S, the class $\mathcal{C}_\mathsf{S}$ of all S-models is a proper class, and thus the information ranges of statements $R(\varphi,\psi)$ will be proper classes too. This is not a problem for the 'information as range'-perspective, however, because this perspective only makes use of *comparative* statements, and it makes perfect sense to say that $X \subseteq Y$ for proper classes X and Y (Jech 2002, p. 6).

Definition 9.7 provides a 'top-down' perspective on the information range of a statement $R(\varphi,\psi)$: we start from the class of all models, and remove those that are not compatible with $R(\varphi,\psi)$. Lemma 9.4 provides an alternative, 'bottom-up' perspective, by characterizing the information range of $R(\varphi,\psi)$ as a union of truth classes, i.e. classes of models of the form $[\![\, \Delta_i(\varphi,\psi) \,]\!] = \{\mathbb{M} \in \mathcal{C}_\mathsf{S} \mid \mathbb{M} \models \Delta_i(\varphi,\psi)\}$.

Lemma 9.4. *Consider arbitrary formulas $\varphi,\psi \in \mathcal{L}_\mathsf{S}$. Then the following hold:*

$$\begin{array}{lcl}
\mathbb{I}(CD(\varphi,\psi)) & = & [\![\, \Delta_2(\varphi,\psi) \,]\!] \cup [\![\, \Delta_3(\varphi,\psi) \,]\!], \\
\mathbb{I}(C(\varphi,\psi)) & = & [\![\, \Delta_2(\varphi,\psi) \,]\!] \cup [\![\, \Delta_3(\varphi,\psi) \,]\!] \cup [\![\, \Delta_4(\varphi,\psi) \,]\!], \\
\mathbb{I}(SC(\varphi,\psi)) & = & [\![\, \Delta_1(\varphi,\psi) \,]\!] \cup [\![\, \Delta_2(\varphi,\psi) \,]\!] \cup [\![\, \Delta_3(\varphi,\psi) \,]\!], \\
\mathbb{I}(NCD(\varphi,\psi)) & = & [\![\, \Delta_1(\varphi,\psi) \,]\!] \cup [\![\, \Delta_2(\varphi,\psi) \,]\!] \cup [\![\, \Delta_3(\varphi,\psi) \,]\!] \cup [\![\, \Delta_4(\varphi,\psi) \,]\!] = \mathcal{C}_\mathsf{S}, \\
\mathbb{I}(BI(\varphi,\psi)) & = & [\![\, \Delta_1(\varphi,\psi) \,]\!] \cup [\![\, \Delta_4(\varphi,\psi) \,]\!], \\
\mathbb{I}(LI(\varphi,\psi)) & = & [\![\, \Delta_1(\varphi,\psi) \,]\!] \cup [\![\, \Delta_3(\varphi,\psi) \,]\!] \cup [\![\, \Delta_4(\varphi,\psi) \,]\!], \\
\mathbb{I}(RI(\varphi,\psi)) & = & [\![\, \Delta_1(\varphi,\psi) \,]\!] \cup [\![\, \Delta_2(\varphi,\psi) \,]\!] \cup [\![\, \Delta_4(\varphi,\psi) \,]\!], \\
\mathbb{I}(NI(\varphi,\psi)) & = & [\![\, \Delta_1(\varphi,\psi) \,]\!] \cup [\![\, \Delta_2(\varphi,\psi) \,]\!] \cup [\![\, \Delta_3(\varphi,\psi) \,]\!] \cup [\![\, \Delta_4(\varphi,\psi) \,]\!] = \mathcal{C}_\mathsf{S}.
\end{array}$$

Proof. We prove the first item. Recall that by Definition 9.3, $CD(\varphi,\psi)$ entails that $\mathsf{S} \models \neg\Delta_i(\varphi,\psi)$ for $i = 1,4$, and hence, by Definition 9.7, an S-model $\mathbb{M}$ is compatible with $CD(\varphi,\psi)$ iff $\mathbb{M} \models \neg\Delta_1(\varphi,\psi)$ and $\mathbb{M} \models \neg\Delta_4(\varphi,\psi)$ (†). Furthermore, note that it follows from the definitions of Δ_i that $\mathsf{S} \models \bigvee_{i=1}^{i=4} \Delta_i(\varphi,\psi)$ (‡). We thus get the following chain of identities:

$$\begin{array}{rcll}\mathbb{I}(CD(\varphi,\psi)) & = & \{\mathbb{M}\in\mathcal{C}_{\mathsf{S}} \mid \mathbb{M} \text{ is compatible with } CD(\varphi,\psi)\} & \\ & = & \{\mathbb{M}\in\mathcal{C}_{\mathsf{S}} \mid \mathbb{M}\models\neg\Delta_1(\varphi,\psi) \text{ and } \mathbb{M}\models\neg\Delta_4(\varphi,\psi)\} & (\dagger)\\ & = & \{\mathbb{M}\in\mathcal{C}_{\mathsf{S}} \mid \mathbb{M}\models\Delta_2(\varphi,\psi) \text{ or } \mathbb{M}\models\Delta_3(\varphi,\psi)\} & (\ddagger)\\ & = & [\![\,\Delta_2(\varphi,\psi)\,]\!]\cup[\![\,\Delta_3(\varphi,\psi)\,]\!]. & \end{array}$$

The other items are proved completely analogously. □

We are now ready to move from the informativity of statements to that of relations. By universally quantifying over $\mathcal{L}_{\mathsf{S}}$, we lift the informativity ordering $\leq_i$ of statements of the form $R(\varphi,\psi)$ to an informativity ordering $\leq_i^{\forall}$ of the relations R themselves.

Definition 9.8. Consider relations $R,S\in\mathcal{OG}\cup\mathcal{IG}$. Then we define:

$$R\leq_i^{\forall} S \;:\Leftrightarrow\; \forall\varphi,\psi\in\mathcal{L}_{\mathsf{S}}\colon R(\varphi,\psi)\leq_i S(\varphi,\psi).$$

Since $\leq_i$ is a preordering, its lifted version $\leq_i^{\forall}$ is a preordering as well. The strict version of this ordering is defined as follows:

$$R<_i^{\forall} S :\Leftrightarrow (R\leq_i^{\forall} S \text{ and } S\not\leq_i^{\forall} R).$$

This is a strict partial ordering (Harel et al. 2000, p. 11).

Definition 9.8 defines the informativity ordering $\leq_i^{\forall}$ for $\mathcal{OG}\cup\mathcal{IG}$, i.e. for opposition and implication relations collectively. Theorems 9.3 and 9.4 describe how $\leq_i^{\forall}$ orders the opposition and implication geometries separately.[32]

Theorem 9.3. *The opposition geometry $\mathcal{OG}$ is ordered by $\leq_i^{\forall}$ as follows:*

- $NCD\leq_i^{\forall} C, NCD\leq_i^{\forall} SC, NCD\leq_i^{\forall} CD, C\leq_i^{\forall} CD$ *and* $SC\leq_i^{\forall} CD$*,*
- *for all* $R\in\mathcal{OG}\colon R\leq_i^{\forall} R$*,*
- *for all other pairs* $(R,S)\in\mathcal{OG}^2\colon R\not\leq_i^{\forall} S$*.*

Proof. We prove that $C\leq_i^{\forall} CD$ (the other items of the form $R\leq_i^{\forall} S$ are proved analogously). Consider arbitrary formulas $\varphi,\psi\in\mathcal{L}_{\mathsf{S}}$; it suffices to show that $C(\varphi,\psi)\leq_i CD(\varphi,\psi)$. It follows from Lemma 9.4 that

$$\begin{aligned}\mathbb{I}(CD(\varphi,\psi)) &= [\![\,\Delta_2(\varphi,\psi)\,]\!]\cup[\![\,\Delta_3(\varphi,\psi)\,]\!]\\ &\subseteq [\![\,\Delta_2(\varphi,\psi)\,]\!]\cup[\![\,\Delta_3(\varphi,\psi)\,]\!]\cup[\![\,\Delta_4(\varphi,\psi)\,]\!]=\mathbb{I}(C(\varphi,\psi)).\end{aligned}$$

[32] The only cross-geometry informativity statements that hold are $NCD\leq_i^{\forall} R$ and $NI\leq_i^{\forall} R$, for all relations $R\in\mathcal{OG}\cup\mathcal{IG}$. I will return to such cross-geometry statements in Subsection 9.5.1; in particular, see Definition 9.9.

By the definition of the $\leq_i$, this means that $C(\varphi,\psi) \leq_i CD(\varphi,\psi)$.

We now prove that $CD \not\leq_i^\forall C$ (the other items of the form $R \not\leq_i^\forall S$ are proved analogously). It suffices to show that $\exists\varphi,\psi \in \mathcal{L}_\mathsf{C}\colon CD(\varphi,\psi) \not\leq_i C(\varphi,\psi)$. Let $\varphi := p$ and $\psi := q$. Note that $[\![\, \Delta_4(p,q) \,]\!]$ is non-empty (there certainly exists a model $\mathbb{M}$ such that $\mathbb{M} \models \Delta_4(p,q)$); hence

$$\begin{aligned}\mathbb{I}(C(p,q)) &= [\![\, \Delta_2(p,q) \,]\!] \cup [\![\, \Delta_3(p,q) \,]\!] \cup [\![\, \Delta_4(p,q) \,]\!] \\ &\not\subseteq [\![\, \Delta_2(p,q) \,]\!] \cup [\![\, \Delta_3(p,q) \,]\!] = \mathbb{I}(CD(\varphi,\psi)).\end{aligned}$$

Again, by the definition of $\leq_i$, this means that $CD(p,q) \not\leq_i C(p,q)$. □

Theorem 9.4. *The implication geometry $\mathcal{IG}$ is ordered by $\leq_i^\forall$ as follows:*

- $NI \leq_i^\forall LI, NI \leq_i^\forall RI, NI \leq_i^\forall BI, LI \leq_i^\forall BI$ *and* $RI \leq_i^\forall BI$*,*
- *for all* $R \in \mathcal{IG}\colon R \leq_i^\forall R$*,*
- *for all other pairs of relations* $(R,S) \in \mathcal{IG}^2\colon R \not\leq_i^\forall S$*.*

Proof. Completely analogous to the proof of Theorem 9.3. □

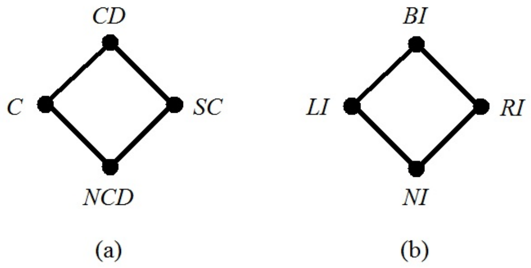
It follows from Theorem 9.3 that $C \leq_i^\forall CD$ and $CD \not\leq_i^\forall C$, and hence $C <_i^\forall CD$. Similarly, we get that $NCD <_i^\forall C$, $NCD <_i^\forall SC$, $NCD <_i^\forall CD$, and $SC <_i^\forall CD$. Furthermore, C and SC are $\leq_i$-incomparable. The opposition relations are thus ordered by informativity as in Figure 9.7(a). Completely analogously, it follows from Theorem 9.4 that the implication relations are ordered as in Figure 9.7(b). The relations NCD and NI are thus the *least informative* in their respective geometries,[33] a property that will become crucial in Subsection 9.5.2.

9.4.3 Motivating the Information Account

The application of the ‘information as range’-perspective to the opposition and implication geometries is well-motivated: it matches well with our intuitive informativity judgments (§ 9.4.3.1), and it intertwines seamlessly with the structure of both geometries (§ 9.4.3.2).

[33] This is reflected in the code in Figure 9.5, which visualizes these two relations in grey instead of black (recall Footnote 20).

Figure 9.7: Informativity ordering of (a) the opposition geometry and (b) the implication geometry.



9.4.3.1 Intuitive Informativity Judgments

In the previous subsection, the ‘information as range’-perspective was used to order the opposition and implication relations according to informativity. For example, it entails that NCD is the least informative opposition relation, that CD is the most informative opposition relation, and that C and SC are in between (cf. Theorem 9.3 and Figure 9.7).[34] These theoretical claims seem to match our intuitive judgments about the relative informativity of the opposition relations.

I will describe a game to explain this. Recall that there are 16 binary, truth-functional connectives. For each connective $\bullet$, we consider the formula $p \bullet q$; these 16 formulas form a Boolean algebra $\mathbb{B}_4$. We randomly choose a formula from $\mathbb{B}_4$, let’s say q, and show it to the opponent *Op*. We then randomly select another formula φ from $\mathbb{B}_4$ (possibly the one that was chosen before), but do *not* show it to *Op*. Instead, we determine the opposition relation that holds between q and φ, and communicate this to *Op*. Based on this information, *Op* has to guess the value of φ. There are four cases:

1. $CD(q, \varphi)$: then *Op* knows that $\varphi = \neg q$,
2. $C(q, \varphi)$: then *Op* knows that $\varphi \in \{p \wedge \neg q,\ \neg p \wedge \neg q,\ \bot\}$,
3. $SC(q, \varphi)$: then *Op* knows that $\varphi \in \{p \vee \neg q,\ \neg p \vee \neg q,\ \top\}$,

[34] We will focus exclusively on the opposition geometry in this paragraph. However, all claims straightforwardly carry over to the implication geometry.

4. $NCD(q, \varphi)$: then *Op* knows that

$$\varphi \in \{p,\ q,\ p \wedge q,\ \neg p,\ p \vee q,\ p \to q,\ p \leftrightarrow q,\ p \veebar q,\ \neg p \wedge q\}.$$

In the first case, *Op* comes to know the exact value of φ; CD is thus the most informative opposition relation. In the second and third cases, *Op* comes to know that φ has one out of 3 values, but she remains uncertain as to which of these 3; hence, C and SC are strictly less informative than CD. In the fourth case, *Op* merely comes to know that φ has one out of 9 values; hence, NCD is strictly less informative than C and SC, and thus the least informative opposition relation.

One might object that this intuitive scenario only works because the formula q sits in the *middle* level of $\mathbb{B}_4$, and that formulas in other levels will yield results that match less well with the informativity claims made by the account developed here.[35] For example, q has the same number of contraries and subcontraries (viz. 3), while for formulas in other levels, this might not be the case. However, it should be noted that in general, the Boolean algebra $\mathbb{B}_n$ (with $n > 1$) has levels $L_0, L_1, L_2, \ldots L_{n-2}, L_{n-1}, L_n$, and that every formula not belonging to L_0, L_1, L_{n-1} or L_n yields the right *comparative* results, i.e. the numbers of its contraries and of its subcontraries will be strictly between the number of its contradictories and the number of its noncontradictories.[36] Furthermore, since $|L_k| = \binom{n}{k} = \frac{n!}{k!(n-k)!}$, it holds that

$$\lim_{n\to\infty} \frac{|L_0| + |L_1| + |L_{n-1}| + |L_n|}{|\mathbb{B}_n|} = \lim_{n\to\infty} \frac{1+n+n+1}{2^n} = 0.$$

Hence, the chance that a randomly chosen formula belongs to L_0, L_1, L_{n-1} or L_n (and thus yields results that do not entirely match with the informativity claims made by our account) vanishes for sufficiently large Boolean algebras.

The remarks made above apply not only to Boolean algebras of formulas, but to finite Boolean algebras *in general*, since the opposition and implication relations can be defined for any Boolean algebra $\mathbb{B} = \langle B, \wedge_\mathbb{B}, \vee_\mathbb{B}, \neg_\mathbb{B}, \bot_\mathbb{B}, \top_\mathbb{B}, \leq_\mathbb{B}\rangle$;

[35] For a formal definition of the notion of 'level' in a Boolean algebra, or, more generally, in a poset, see Engel (1997, p. 7).

[36] More precisely, a formula in level L_i has 1 contradictory, $2^{n-i} - 1$ contraries, $2^i - 1$ subcontraries, and $(2^{n-i} - 1) \cdot (2^i - 1)$ noncontradictories. Note that $1 < \{2^{n-i} - 1, 2^i - 1\} < (2^{n-i} - 1) \cdot (2^i - 1)$ iff $1 < i < n - 1$, i.e. a formula yields the right comparative results iff it does not belong to L_0, L_1, L_{n-1} or L_n. Finally, note that if $i \approx \frac{n}{2}$, then $2^{n-i} - 1 \approx 2^i - 1$, i.e. formulas sitting (approximately) in the middle level of $\mathbb{B}_n$ indeed have (approximately) the same number of contraries and subcontraries.

for example, contrariety and left-implication in $\mathbb{B}$ are typically defined as follows (for any $x, y \in B$):[37]

$$\begin{array}{llllll} C_{\mathbb{B}}(x,y) & \text{iff} & x \wedge_{\mathbb{B}} y = \bot_{\mathbb{B}} & \text{and} & x \vee_{\mathbb{B}} y \neq \top_{\mathbb{B}}, \\ LI_{\mathbb{B}}(x,y) & \text{iff} & x \leq_{\mathbb{B}} y & \text{and} & y \not\leq_{\mathbb{B}} x. \end{array}$$

9.4.3.2 Coherence with the Structure of the Geometries

A major theoretical advantage of the informativity perspective on the opposition and implication geometries is that it intertwines seamlessly with the internal and external structure of these geometries, which was described in Subsection 9.3.3 (§ 9.3.3.3).

The informativity ordering of the opposition geometry is fully described by Theorem 9.3, and visualized by Figure 9.7(a). It is immediately clear from this figure that from the informativity perspective, C and SC play symmetrical roles: both are strictly in between NCD and CD. Formally, this can be expressed as follows:

$$NCD <_i^\forall C <_i^\forall CD \quad \text{and} \quad NCD <_i^\forall SC <_i^\forall CD.$$

However, there is a theoretical redundancy here, since each of these two series of inequalities actually *follows from* the other one. Using the mapping $N12$ that was defined in Corollary 9.2, this can be reformulated as follows:

Lemma 9.5. *For all relations $R, S \in \mathcal{OG}$: $R \leq_i^\forall S$ iff $N12(R) \leq_i^\forall N12(S)$.*

Proof. For all $R, S \in \mathcal{OG}$, it holds that

$$\begin{array}{llll} R \leq_i^\forall S & \Leftrightarrow & \forall \varphi, \psi \in \mathcal{L}_{\mathsf{S}}\colon R(\varphi,\psi) \leq_i S(\varphi,\psi) & \\ & \Leftrightarrow & \forall \varphi, \psi \in \mathcal{L}_{\mathsf{S}}\colon N12(R)(\neg\varphi,\neg\psi) \leq_i N12(S)(\neg\varphi,\neg\psi) & (\dagger) \\ & \Leftrightarrow & \forall \varphi, \psi \in \mathcal{L}_{\mathsf{S}}\colon N12(R)(\varphi,\psi) \leq_i N12(S)(\varphi,\psi) & (\ddagger) \\ & \Leftrightarrow & N12(R) \leq_i^\forall N12(S). & \end{array}$$

[37] The opposition relations are thus typically defined in terms of $\wedge_{\mathbb{B}}$ and $\vee_{\mathbb{B}}$, while the implication relations are typically defined in terms of $\leq_{\mathbb{B}}$. This suggests that the distinction between the opposition and implication geometries is analogous to the distinction between the algebraic and order-theoretic perspectives on Boolean algebras (Davey and Priestley 2002, pp. 33–41). Furthermore, it is well-known that both perspectives are equivalent to each other (via $x \leq_{\mathbb{B}} y \Leftrightarrow x \wedge_{\mathbb{B}} y = x$); this is analogous to the connection between the opposition and implication geometries described in Lemma 9.3.

The †-labeled equivalence holds because of Corollary 9.2. The ‡-labeled equivalence holds because of the universal quantification over $\mathcal{L}_\mathsf{S}$ and the fact that $R(\varphi, \psi)$ iff $R(\neg\neg\varphi, \neg\neg\psi)$ for any opposition relation R and formulas φ, ψ. □

Similar remarks can be made about the connection between the informativity ordering of the implication geometry—as described by Theorem 9.4 and visualized by Figure 9.7(b)—and the internal structure of this geometry (if $\mathcal{OG}$ is replaced with $\mathcal{IG}$ in Lemma 9.5, the proof remains valid).

I now turn to the connection between the informativity perspective and the geometries' external structure (i.e. how they are related to each other). It is immediately clear from Figure 9.7 that $\mathcal{OG}$ and $\mathcal{IG}$ are ordered in exactly the same way. Formally, this can be expressed as follows:

$$NCD <_i^\forall \{C, SC\} <_i^\forall CD \quad \text{and} \quad NI <_i^\forall \{LI, RI\} <_i^\forall BI.$$

However, there is a theoretical redundancy here, since each of these two series of inequalities actually *follows from* the other one. Using the mapping $N2$ that was defined in Corollary 9.3, this can be reformulated as follows:

Lemma 9.6. *For all relations $R, S \in \mathcal{OG} \cup \mathcal{IG}$: $R \leq_i^\forall S$ iff $N2(R) \leq_i^\forall N2(S)$.*

Proof. Completely analogous to the proof of Lemma 9.5, but based on Corollary 9.3 instead of Corollary 9.2. □

In sum, there are certain facts about the informativity ordering of the opposition and implication geometries that can be obtained in two distinct ways:

1. by deriving them directly from the 'information as range'-perspective on these geometries (Definitions 9.7 and 9.8); this was done in Theorems 9.3 and 9.4;

2. by combining that perspective with the geometries' internal and external structure (Corollaries 9.2 and 9.3); this was done in Lemmas 9.5 and 9.6.

These considerations can be seen as evidence for the theoretical robustness of the informativity account that was described in this section.

9.5 Information in the Aristotelian Geometry and its Diagrams

In the previous two sections, I have introduced the opposition and implication geometries, and shown how the 'information as range'-perspective can be applied to them. This conceptual machinery will now be used to explicate the intuition that the Aristotelian square is highly informative. This will be done in two successive steps: in Subsection 9.5.1 I will show that the Aristotelian geometry is an informative *geometry*, and in Subsection 9.5.2 I will show that within this geometry, the well-known square is a highly informative *diagram*. Finally, in Subsection 9.5.3 I will reassess the purported problems of the Aristotelian geometry in the light of these informativity considerations.

9.5.1 Information in the Aristotelian Geometry

The Aristotelian geometry (Definition 9.1) can be characterized as being *hybrid* between the opposition geometry (Definition 9.3) and the implication geometry (Definition 9.4): it consists of three opposition relations (CD, C and SC) and one implication relation (LI, i.e. SA).[38] From an informativity perspective, the former three are the most informative relations in the opposition geometry, while the latter is second most informative in the implication geometry (Figure 9.7). Hence, the Aristotelian geometry is hybrid in an informationally optimal way.

One might object that for the Aristotelian geometry to be truly informationally optimal, it would have to include BI, since that implication relation is strictly more informative than LI. Additionally, since RI is second most informative too, it seems arbitrary to include LI and not RI. I will now provide a more formal account of the hybrid nature of the Aristotelian geometry, which adequately addresses both these objections, and thus supports the conclusion regarding its informational optimality.

For our purposes, it will be necessary to compare informativity 'across geometries'. For example, considering Figure 9.7, there is a clear sense in which CD is strictly more informative in $\mathcal{OG}$ than RI is in $\mathcal{IG}$: CD is the most informative relation in $\mathcal{OG}$, while RI is only amongst the second most informative

[38] This should come as no surprise, since the opposition and implication geometries were obtained in Subsection 9.3.2 precisely by conceptually disentangling the Aristotelian geometry.

relations in $\mathcal{IG}$. Still, one can check that it does not hold that $RI <_i^\forall CD$.[39] However, the mapping $N2$ defined in Corollary 9.3 *does* enable us to make such cross-geometrical informativity comparisons (this is justified because informativity is invariant under this mapping; recall Lemma 9.6). For example, although it does not hold that $RI <_i^\forall CD$, it *does* hold that $N2(RI) = SC <_i^\forall CD$, and therefore CD will be called the 'winner' of $\{CD, RI\}$.

Definition 9.9. Consider arbitrary relations $R \in \mathcal{OG}$ and $S \in \mathcal{IG}$. Then the *winner* of $\{R, S\}$ is defined as follows:

- S is the winner iff $R <_i^\forall N2(S)$,
- R is the winner iff $N2(S) <_i^\forall R$.

For example, BI is the winner of $\{C, BI\}$, since $C <_i^\forall CD = N2(BI)$. Furthermore, $\{C, RI\}$ does not yield a winner at all: $C \not<_i^\forall SC = N2(RI)$, so RI is not the winner; and $N2(RI) = SC \not<_i^\forall C$, so C is not the winner either.

We are now ready to discuss the informational optimality of the Aristotelian geometry in a formally precise sense. Theorem 9.5 below states that all Aristotelian relations (between contingent formulas) are winners, i.e. informationally optimal.

Theorem 9.5. *Consider arbitrary contingent formulas $\varphi, \psi \in \mathcal{L}_\mathsf{S}$. Let $R \in \mathcal{OG}$ and $S \in \mathcal{IG}$ be the unique relations such that $R(\varphi, \psi)$ and $S(\varphi, \psi)$, respectively.*

1. *If $R \in \mathcal{AG}$, then R is the winner of $\{R, S\}$.*
2. *If $S \in \mathcal{AG}$, then S is the winner of $\{R, S\}$.*

Proof. First, suppose that $R \in \mathcal{AG}$. Since φ and ψ are contingent, it follows by Theorem 9.2 that $R \in \{CD, C, SC\}$ and $S = NI$, and hence R is the winner of $\{R, S\}$. Second, suppose that $S \in \mathcal{AG}$. Since φ and ψ are contingent, it follows by Theorem 9.2 that $R = NCD$ and $S = LI$, and hence S is the winner of $\{R, S\}$. □

Since Theorem 9.5 states that all Aristotelian relations are winners, it is natural to ask the converse question: are all winners Aristotelian? Theorem 9.6 states that this is indeed by and large the case.

[39] Theorems 9.3 and 9.4 apply only 'locally' to $\mathcal{OG}$ and $\mathcal{IG}$, respectively; recall Footnote 32.

Theorem 9.6. *Let $\varphi, \psi \in \mathcal{L}_\mathsf{S}$ and $R \in \mathcal{OG}$, $S \in \mathcal{IG}$ be as in Theorem 9.5.*

1. *If R is the winner of $\{R, S\}$, then $R \in \mathcal{AG}$.*
2. *If S is the winner of $\{R, S\}$, then $S \in \mathcal{AG} \cup \{BI, RI\}$.*

Proof. First, suppose that R is the winner of $\{R, S\}$. Since φ and ψ are contingent, it follows by Theorem 9.2 that $R \in \{CD, C, SC\} \subseteq \mathcal{AG}$. Second, suppose that S is the winner of $\{R, S\}$. Since φ and ψ are contingent, it follows by Theorem 9.2 that $S \in \{BI, LI, RI\} \subseteq \mathcal{AG} \cup \{BI, RI\}$. □

The two cases where the winner is BI or RI (and thus does not belong to $\mathcal{AG}$) correspond exactly to the two objections against the informational optimality of $\mathcal{AG}$ that were raised at the beginning of this subsection. I will now discuss how these cases are resolved in the class of Aristotelian diagrams.[40]

First, note that by definition, Aristotelian diagrams are semantic entities, i.e. they do not contain any distinct equivalent formulas (Definition 9.2). Logical equivalence coincides with BI, and therefore, in any diagram, BI holds exactly between each formula and itself.[41] The BI-relations thus need not be visualized explicitly in the Aristotelian diagrams: their place is predetermined by their definition and does not vary from diagram to diagram (they occur exactly as the ‘loops’ between each formula and itself).

Second, note that if $RI(\varphi, \psi)$, then Theorem 9.2 yields $NCD(\varphi, \psi)$. By Lemma 9.1 it follows that $NCD(\psi, \varphi)$ and $LI(\psi, \varphi)$. The winner of $\{NCD, LI\}$ is LI, which *does* belong to $\mathcal{AG}$. Therefore, the RI-relations need not be visualized explicitly in the Aristotelian diagrams: they correlate exactly with the LI-relations (because LI and RI offer two complementary perspectives on truth propagation; recall Footnote 21).[42]

[40] Since this discussion applies to *all* Aristotelian diagrams, it rightly belongs in this subsection. The next subsection, in contrast, will distinguish between various *particular* Aristotelian diagrams, e.g. the concrete square, the concrete Sesmat-Blanché hexagon, etc.

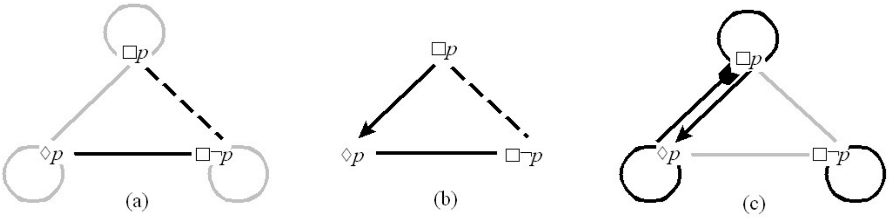
[41] This argument is made fully precise in Definitions 9.11 and 9.12 and Lemma 9.8 in the appendix.

[42] From a theoretical perspective, the case of LI/RI described above (which is based on the equivalence $LI(\varphi, \psi) \Leftrightarrow RI(\psi, \varphi)$, cf. Lemma 9.1) seems to be exactly similar to the case of C/SC (which is based on the equivalence $C(\varphi, \psi) \Leftrightarrow SC(\neg\varphi, \neg\psi)$, cf. Lemma 9.2) and to that of C/LI (which is based on the equivalence $C(\varphi, \psi) \Leftrightarrow LI(\varphi, \neg\psi)$, cf. Lemma 9.3). This might suggest that SC and LI can unproblematically be left out of the Aristotelian diagrams as well. From a visual perspective, however, the latter two cases are entirely different from the first. The

The remarks above suggest that from an *information visualization* viewpoint, BI and RI are redundant (even though they are among the more informative implication relations). After all, their place in the Aristotelian diagrams is predictable, either absolutely (the BI-relations occur exactly as the 'loops'), or on the basis of the other Aristotelian relations (the RI-relations occur exactly wherever there are LI-relations in the converse direction). It is important to emphasize, however, that this visual redundancy of BI and RI is perfectly compatible with their importance in the internal and external structure of the implication geometry (as discussed in § 9.3.3.3).

Example 9.1. Consider the fragment $\{\Box p, \Diamond p, \Box\neg p\}$. Parts (a), (b) and (c) of Figure 9.8 show the respective opposition, Aristotelian and implication diagrams for this fragment (using the code of Figure 9.5).[43]

Figure 9.8: The (a) opposition, (b) Aristotelian and (c) implication diagram for $\{\Box p, \Diamond p, \Box\neg p\}$



For each pair of distinct formulas, the Aristotelian diagram contains the winner of the corresponding opposition and implication relations:

- $C(\Box p, \Box\neg p)$ and $NI(\Box p, \Box\neg p)$: the winner of $\{C, NI\}$ is $C \in \mathcal{AG}$;

LI/RI case does not require considering formulas other than φ and ψ; hence, in the diagrams, the RI relations occur in exactly the same place as the original LI relations (but in the reverse direction). On the other hand, the C/SC and C/LI cases require considering formulas other than φ and ψ, viz. $\neg\varphi$ and/or $\neg\psi$; hence, in the diagrams, the SC and LI relations occur in other places than the original C relations.

[43] Note that Figure 9.8 shows the same three diagrams as Figure 9.6, but in a different order: I will henceforth put the Aristotelian diagram in between the opposition and implication diagrams, to reflect the fact that $\mathcal{AG}$ is hybrid between $\mathcal{OG}$ and $\mathcal{IG}$.

- $CD(\Diamond p, \Box\neg p)$ and $NI(\Diamond p, \Box\neg p)$: the winner of $\{CD, NI\}$ is $CD \in \mathcal{AG}$;
- $NCD(\Box p, \Diamond p)$ and $LI(\Box p, \Diamond p)$: the winner of $\{NCD, LI\}$ is $LI \in \mathcal{AG}$.

Furthermore, note that for each formula φ in this fragment, it holds that $NCD(\varphi, \varphi)$ and $BI(\varphi, \varphi)$. Although BI is the winner of $\{NCD, BI\}$, it is not visualized in the Aristotelian diagram, because of the reasons stated above. Finally, note that $NCD(\Diamond p, \Box p)$ and $RI(\Diamond p, \Box p)$; although RI is the winner of $\{NCD, RI\}$, it is not visualized in the Aristotelian diagram (but its converse LI *is* visualized; cf. the third item above).

In sum, Theorems 9.5 and 9.6 together state that a relation between contingent formulas is a winner if and only if it is Aristotelian (modulo BI and RI). Hence, each Aristotelian diagram offers an informationally optimal picture of its vertices: all winners are represented in the Aristotelian diagram (modulo BI and RI), and all Aristotelian relations are winners, i.e., contrapositively, all non-winners are not represented in the Aristotelian diagram.

9.5.2 Information in the Aristotelian Diagrams

In the previous subsection I have argued that the Aristotelian geometry is informationally optimal in a *positive* sense: a relation between contingent formulas is Aristotelian if and only if it is a winner (i.e. cross-geometrically *most* informative). In this subsection I will focus on a *negative* aspect of this informational optimality, by showing that Aristotelian diagrams avoid (the combination of) the *least* informative relations as much as possible. It turns out that this minimally informative combination does not occur in certain diagrams, but is unavoidable in others.

One question that was left unanswered in the previous subsection is: what if $\{R, S\}$ does not yield a winner at all? Such cases certainly exist; cf. the $\{C, RI\}$ example below Definition 9.9. However, in the case of contingent formulas, if $\{R, S\}$ does not yield a winner, it follows by contraposition on Theorem 9.5 that neither R nor S is an Aristotelian relation. Moreover, the following theorem states that in this case, R and S are uniquely identified.

Theorem 9.7. *Let $\varphi, \psi \in \mathcal{L}_{\mathsf{S}}$ and $R \in \mathcal{OG}, S \in \mathcal{IG}$ be as in Theorem 9.5. If neither R nor S is the winner of $\{R, S\}$, then $R = NCD$ and $S = NI$.*

Proof. Since R is not the winner of $\{R, S\}$, it follows by Theorem 9.2 that $R \notin \{CD, C, SC\}$, and thus $R = NCD$. Similarly, since S is not the winner of $\{R, S\}$, it follows that $S \notin \{BI, LI, RI\}$, and thus $S = NI$. □

The combination of NCD and NI effectively occurs; for example, it is easy to check that p and $\Diamond p \wedge \Diamond\neg p$ are both non-contradictory and in non-implication. This combination of relations will be crucial to the remainder of this subsection, and is therefore given a separate name, viz. 'unconnectedness'.[44]

Definition 9.10. Let $\varphi, \psi \in \mathcal{L}_\mathsf{S}$ be arbitrary formulas. Then φ and ψ are said to be *unconnected*, written $U(\varphi, \psi)$, iff they are in non-contradiction and non-implication. Formally: $U(\varphi, \psi) :\Leftrightarrow NCD(\varphi, \psi)$ and $NI(\varphi, \psi)$.

The term 'unconnectedness' suggests that unconnected formulas stand in no Aristotelian relation at all.[45] The following theorem shows that this is essentially correct (and thus justifies the term 'unconnectedness').[46]

Theorem 9.8. *Consider formulas φ, ψ in an arbitrary Aristotelian diagram. Then:*

φ and ψ do not stand in any Aristotelian relation $\Leftrightarrow$ φ and ψ are unconnected.

Proof. Let $R \in \mathcal{OG}$ and $S \in \mathcal{IG}$ be the unique relations such that $R(\varphi, \psi)$ and $S(\varphi, \psi)$, respectively. For the right-to-left direction, note that if φ and ψ are unconnected, then $R = NCD \notin \mathcal{AG}$ and $S = NI \notin \mathcal{AG}$. We now prove the left-to-right direction. Assume that $R \notin \mathcal{AG}$ and $S \notin \mathcal{AG}$. Since φ and ψ belong to an Aristotelian diagram, they are (by Definition 9.2) contingent and non-equivalent. Hence $S \neq BI$. Furthermore, note that if $S = RI$, then (by Lemma 9.1) $LI(\psi, \varphi)$, and since LI *is* an Aristotelian relation, this contradicts the assumption that φ and ψ do not stand in any Aristotelian relation; therefore,

[44] Although its combinatorial and informational properties have not been systematically explored so far, the notion of unconnectedness as such has surfaced at various places in the literature, usually under the label 'logical independence'; for example, see Hughes (1987, p. 99), Béziau (2003, p. 226), Karger (2003, p. 435), Seuren (2010, p. 50), Campos-Benítez (2012, pp. 101ff.), Jacquette (2012, p. 86) and Read (2012a, p. 104).

[45] For example, Campos-Benítez states that "independent sentences [...] are not contrary neither subcontrary nor contradictory or subaltern: they have no relationship at all" (Campos-Benítez 2012, p. 103).

[46] In other words, Theorem 9.8 characterizes the *absence* of any Aristotelian relation in a positive way, viz. as the joint *presence* of an opposition and an implication relation.

$S \neq RI$. Hence, by contraposition on Theorem 9.6 we get that neither R nor S is the winner of $\{R, S\}$. By Theorem 9.7 it follows that $R = NCD$ and $S = NI$, i.e. φ and ψ are unconnected. □

Remark 9.5. The proofs of Theorems 9.5–9.8 all make essential use of Theorem 9.2. The 7 pairs of opposition and implication relations singled out by that theorem are structured according to informativity: they are exactly the pairs that consist of (i) the least informative relation in either the opposition or the implication geometry (NCD or NI, respectively) and (ii) any of the four relations in the other geometry. Formally, this means that the set of those 7 pairs can be written as

$$(\{NCD\} \times \mathcal{IG}) \cup (\mathcal{OG} \times \{NI\}).$$

Of course, in this way the unconnectedness combination gets 'counted twice', since $(NCD, NI) \in (\{NCD\} \times \mathcal{IG}) \cap (\mathcal{OG} \times \{NI\})$. The importance of these 7 pairs was already noted by the medieval logician John Buridan (Karger 2003) and later rediscovered by Doyle (1952), although these authors did not view them as pairs of more primitive notions, nor in the light of informativity considerations as we have done here.

Unconnectedness is thus the combination of the two relations that are *least informative* in their respective geometries.[47] The Aristotelian diagrams avoid this minimally informative combination as much as possible. This can be seen as the negative counterpart of the informativity claim argued for in the previous subsection, viz. that the Aristotelian diagrams consist entirely and exclusively of winners. These negative and positive theses jointly constitute the informational optimality of the Aristotelian diagrams.

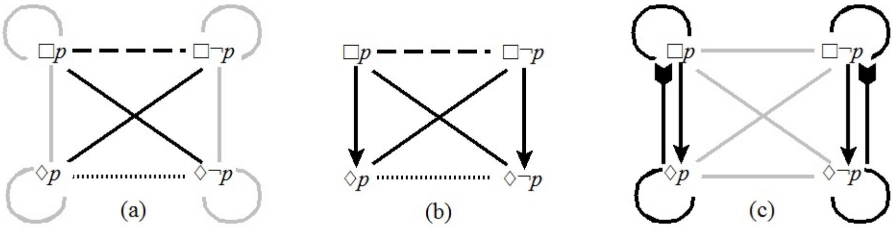
Obviously, whether or not unconnectedness occurs in a given Aristotelian diagram is fully determined by whether there are unconnected formulas amongst its vertices. Considering the diagrams mentioned in Section 9.2, it turns out that some of them have unconnectedness, while others do not. We consider them one by one.[48]

[47] This might explain why some authors, while acknowledging the existence of this relation, deny its logical relevance. For example, according to Seuren, unconnectedness is "a legitimate relation between L-propositions producing truth under certain conditions, yet [...] plays no role [...] in any logic" (Seuren 2010, p. 50). Similarly, Béziau thinks that by treating unconnectedness as a 'real' logical relation, "we are going too far and confusing here negation with distinction" (Béziau 2003, p. 226).

[48] These diagrams fit into an exhaustive typology that is currently being developed (Smessaert

We begin by considering the usual Aristotelian square for the fragment $\mathcal{F}_4 = \{\Box p, \Box\neg p, \Diamond p, \Diamond\neg p\}$, which was already displayed in Figure 9.2(b). Note that there is no unconnectedness in this fragment: for all φ, ψ, it holds that if $NCD(\varphi, \psi)$ then not $NI(\varphi, \psi)$ (or equivalently, if $NI(\varphi, \psi)$ then not $NCD(\varphi, \psi)$), and hence, there are no φ, ψ such that $NCD(\varphi, \psi)$ and $NI(\varphi, \psi)$ simultaneously, i.e. such that $U(\varphi, \psi)$. Visually speaking, each grey NCD relation in the opposition square in Figure 9.9(a) corresponds to a black ($LI/RI/BI$) relation in the implication square in Figure 9.9(c), and vice versa, each grey NI relation on the right corresponds to a black ($CD/C/SC$) relation on the left. Therefore, the Aristotelian square in Figure 9.9(b) contains no unconnectedness: each pair of distinct formulas stands in some Aristotelian relation.

Figure 9.9: The (a) opposition, (b) Aristotelian and (c) implication square for $\{\Box p, \Box\neg p, \Diamond p, \Diamond\neg p\}$



Next, we consider two types of Aristotelian hexagons: the Sesmat-Blanché hexagon for the fragment $\mathcal{F}_{6a} = \mathcal{F}_4 \cup \{\Box p \vee \Box\neg p, \Diamond p \wedge \Diamond\neg p\}$ and the Sherwood-Czeżowski hexagon for the fragment $\mathcal{F}_{6b} = \mathcal{F}_4 \cup \{p, \neg p\}$, which were already displayed in Figure 9.3(a) and (b), respectively. Neither of these fragments contains any unconnectedness, so each pair of distinct formulas in the Aristotelian hexagons in Figures 9.10(b) and 9.11(b) stands in some Aristotelian relation.

None of the Aristotelian diagrams considered thus far contains unconnectedness. This changes, however, when we turn to the Béziau octagon for the fragment $\mathcal{F}_8 = \mathcal{F}_{6a} \cup \mathcal{F}_{6b}$, cf. Figure 9.3(c) above (despite the fact that it is the 'sum' of the hexagons for $\mathcal{F}_{6a}$ and $\mathcal{F}_{6b}$, which themselves do not contain any un-

and Demey 2013a). This typology includes several other types of Aristotelian diagrams, which have various proportions of unconnectedness among their relations.

Figure 9.10: The (a) opposition, (b) Aristotelian and (c) implication Sesmat-Blanché hexagon for $\{\Box p, \Box\neg p, \Diamond p, \Diamond\neg p, \Box p \vee \Box\neg p, \Diamond p \wedge \Diamond\neg p\}$

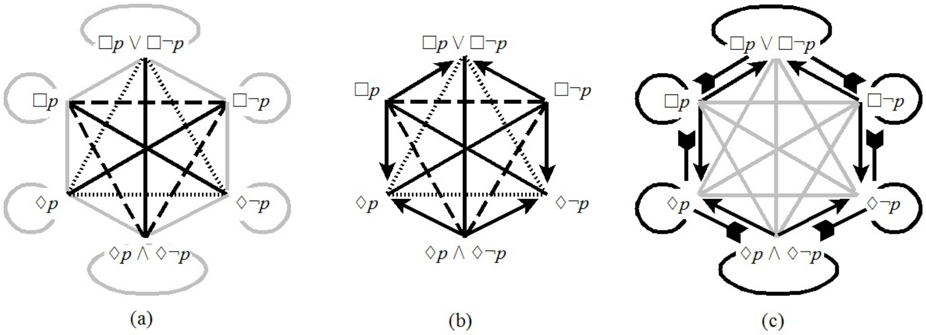


Figure 9.11: The (a) opposition, (b) Aristotelian and (c) implication Sherwood-Czeżowski hexagon for $\{\Box p, \Box\neg p, \Diamond p, \Diamond\neg p, p, \neg p\}$

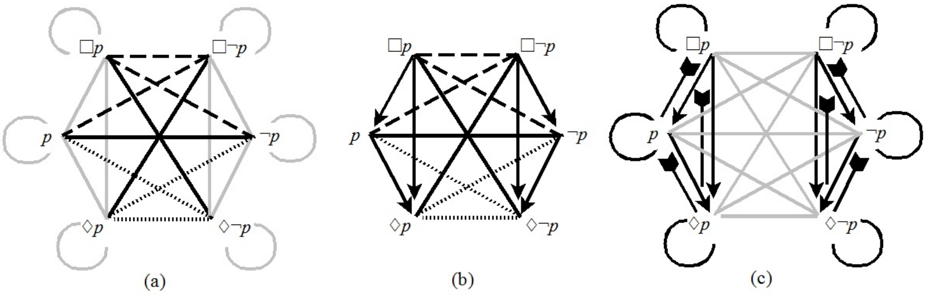
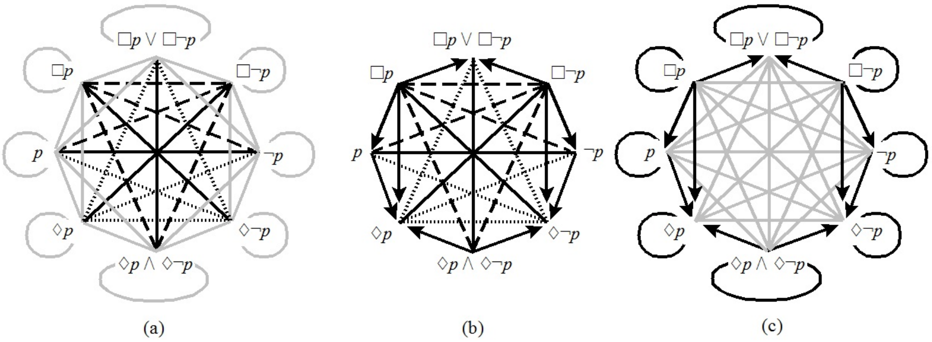


Figure 9.12: The (a) opposition, (b) Aristotelian and (c) implication Béziau octagon for $\{\Box p, \Box\neg p, \Diamond p, \Diamond\neg p, \Box p \vee \neg\Box p, \Diamond p \wedge \Diamond\neg p, p, \neg p\}$ (the RI relations have not been visualized in (c) for the sake of visual clarity)



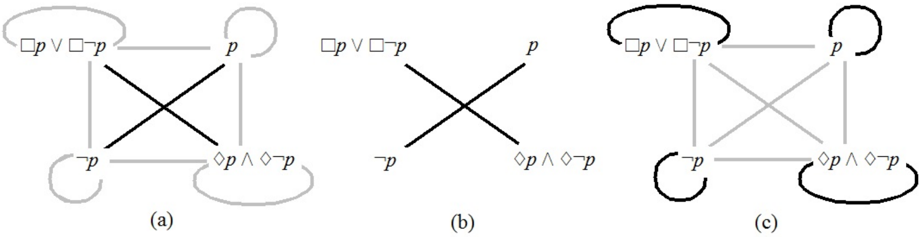
connectedness). For example, we have $NCD(p, \Diamond p \wedge \Diamond\neg p)$ and $NI(p, \Diamond p \wedge \Diamond\neg p)$, and thus $U(p, \Diamond p \wedge \Diamond\neg p)$. Similarly, it holds that $U(p, \Box p \vee \Box\neg p)$, $U(\neg p, \Diamond p \wedge \Diamond\neg p)$ and $U(\neg p, \Box p \vee \Box\neg p)$. Visually speaking, these four formulas are thus connected by four grey NCD relations in the opposition octagon in Figure 9.12(a) and by four grey NI relations in the implication octagon in Figure 9.12(c); hence, they are not connected by any relation in the Aristotelian octagon in Figure 9.12(b).

To summarize, the Aristotelian square and hexagons discussed above do not contain unconnectedness, and thus avoid the combination of the least informative opposition and implication relations. However, this combination *does* occur in the Béziau octagon for $\mathcal{F}_8$, and thus also in every Aristotelian diagram in which this octagon can be embedded (such as the rhombic dodecahedron).

It does not hold in general, however, that larger diagrams contain more unconnectedness. To see this, note that there exist still other 'large' Aristotelian diagrams which do *not* contain any unconnectedness (such as the cube in Moretti (2009a), which consists of 8 formulas). Conversely, there also exist 'small' Aristotelian diagrams which *do* contain unconnectedness. Consider, for example, the three squares for the fragment $\mathcal{F}'_4 = \{\Box p \vee \Box\neg p, \Diamond p \wedge \Diamond\neg p, p, \neg p\}$.[49] The outer

[49] Note that $\mathcal{F}'_4 = \mathcal{F}_8 - \mathcal{F}_4$, i.e. the squares in Figure 9.13 can be seen as the result of 'subtract-

Figure 9.13: The (a) opposition, (b) Aristotelian and (c) implication square for $\{\Box p \vee \neg\Box p, \Diamond p \wedge \Diamond\neg p, p, \neg p\}$



edges of the opposition and implication squares in Figure 9.13(a) and (c) are grey NCD and NI relations, respectively. Hence, there are no Aristotelian relations at the outer edges of the Aristotelian square in Figure 9.13(b), which thus degenerates into "an X of opposition" (Béziau and Payette 2012, p. 13). This abundance of unconnectedness might explain why such degenerated diagrams have rarely been studied in the literature.

9.5.3 Reassessing the Problems of the Aristotelian Geometry

In Subsection 9.3.1 I discussed three problems of the Aristotelian geometry. I will now show that the informativity perspective developed in this section sheds new light on these issues: rather than being 'brute facts' about the Aristotelian geometry, they can be seen as necessary consequences of its informational optimality.

The first problem was that the Aristotelian relations are not mutually exclusive; for example, we have both $C(p \wedge \neg p, p)$ and $LI(p \wedge \neg p, p)$. Recall that any two formulas φ and ψ stand in exactly one opposition relation and exactly one implication relation; hence, if φ an ψ stand in two distinct Aristotelian relations R and S at all, then one of those relations (say R) will be an opposition relation, and the other one (say S) an implication relation. Note that R and S cannot be both the winner of $\{R, S\}$, since otherwise we would have $R <_i^\forall N2(S) <_i^\forall R$, which contradicts the transitivity and irreflexivity of $<_i^\forall$. Hence, if φ and ψ are contingent, then contraposition on Theorem 9.5 yields that R and S cannot be

ing' the classical squares in Figure 9.9 from the corresponding Béziau octagons in Figure 9.12.

both Aristotelian. In other words, φ and ψ can stand in two distinct Aristotelian relations only if at least one of them is non-contingent. This fact was already known (recall Footnote 17); what we discussed here is how it arises out of the informational interplay between the opposition and implication geometries.

The second problem was that the Aristotelian relations are not jointly exhaustive; for example, p and $\Diamond p \wedge \Diamond\neg p$ stand in no Aristotelian relation whatsoever. Recall that unconnectedness is the combination of the least informative opposition and implication relations (NCD and NI). Theorem 9.8 states that two formulas stand in no Aristotelian relation if and only if they are unconnected. In other words, the Aristotelian geometry is indeed not exhaustive, but only inasmuch as this is required by its informational optimality. This means, in particular, that there are no 'fortuitous' failures of exhaustiveness: if two formulas stand in no Aristotelian relation, this can only be because they stand in the least informative opposition and implication relations.[50]

The third and final problem was that the Aristotelian geometry is conceptually confused, because it consists of opposition as well as implication relations. Recall that (modulo the cases of BI and RI) Theorem 9.6 states that all winners are Aristotelian. Since opposition as well as implication relations can be winners, it follows that both kinds of relations belong to the Aristotelian geometry, which thus ends up being hybrid between the opposition and implication geometries. Furthermore, Theorem 9.5 states that a relation is Aristotelian only if it is a winner; in other words, the Aristotelian geometry includes no more relations than is required by informativity considerations.

9.6 Conclusion

In this chapter, I have argued that the classical Aristotelian square of oppositions is highly informative. After distinguishing between the Aristotelian geometry and its concrete diagrams, I introduced two more logical geometries: the opposition and implication geometries. This is a well-motivated move: the new geometries are highly structured (Lemmas 9.1–9.3) and have a canonical correspondence with the binary, truth-functional connectives (Theorem 9.1). I then

[50] Note that the second problem involves contingent formulas (such as p and $\Diamond p \wedge \Diamond\neg p$), and thus occurs both in the Aristotelian geometry and its diagrams. In contrast, we showed above that non-contingency is a necessary condition for the first problem, which thus never occurs in the diagrams (since these contain only contingent formulas).

extended the well-known ‘information as range’-perspective from statements to logical relations, thus obtaining an informativity ordering on the opposition and implication geometries (Theorems 9.3–9.4). This ordering is highly intuitive, and matches well with the geometries’ structural properties (Lemmas 9.5–9.6). I then argued that the Aristotelian geometry is hybrid between the opposition and implication geometries in an informationally optimal way, since it consists entirely and exclusively of winners (Theorems 9.5–9.6). Finally, I studied the notion of unconnectedness (Theorems 9.7–9.8) and found that this minimally informative combination does not occur in the classical Aristotelian square, but does appear in some of its extensions (such as the Béziau octagon).

The following question now arises: what about diagrams such as the Sesmat-Blanché and Sherwood-Czeżowski hexagons? After all, these diagrams are as highly informative as the classical square (they are also Aristotelian diagrams that do not contain any unconnectedness), yet they are much less widely known than the square. In other words, aren’t these hexagons counterexamples to our explanation of the square’s success in terms of its informativity?

Answering this question requires the introduction of one more logical geometry, viz. the *duality geometry*. This geometry is concerned with (the interplay of) internal and external negations on an operator (e.g. $\Diamond = \neg\Box\neg$), and is well-known in linguistics (van Benthem 1991, Löbner 1989, Löbner 1990, Westerståhl 2012) and logic (Demey 2012a, Libert 2012, Veloso et al. 2011). Although the duality geometry is sometimes confused with the Aristotelian geometry (D’Alfonso 2012, Mélès 2012), they are conceptually independent of each other (Löbner 1990, Smessaert 2012a, Westerståhl 2012). It turns out that if the duality geometry is taken into account as well, then the classical square *is* singled out as the most informative diagram (being strictly more informative than all of its extensions, including the hexagons).

It will also be interesting to explore the various connections between the informational account developed here and the exhaustive typology of Aristotelian diagrams developed in Smessaert and Demey (2013a). For example, in this typology we often make use of *bitstrings* (an algebraic representation of the formulas), and some of the informational notions defined here are directly related to bitstring properties (such as length, i.e. number of bit positions). It can be shown, for example, that two formulas are unconnected only if their bitstring representations have a length of at least 4 bit positions.

Appendix

Remark 9.6. The group $\mathbf{G_4} = \langle\{Id, F, N12, FN12\}, \circ\rangle$ is isomorphic to the Klein four-group $\mathbf{V}$. The latter has generators x, y and can be presented as $\langle x, y \,|\, x^2 = 1, y^2 = 1, xy = yx\rangle$. A concrete isomorphism $\iota\colon \mathbf{V} \to \mathbf{G_4}$ is determined by where it sends the generators of $\mathbf{V}$: $\iota(x) = F$ and $\iota(y) = N12$. The Cayley table of $\mathbf{G_4}$ thus looks as follows:

$\circ$	Id	F	$N12$	$FN12$
Id	Id	F	$N12$	$FN12$
F	F	Id	$FN12$	$N12$
$N12$	$N12$	$FN12$	Id	F
$FN12$	$FN12$	$N12$	F	Id

This group acts faithfully on $\mathcal{G}$, and partitions it into six $\mathbf{G_4}$-orbits:

1) $\mathbf{G_4}(CD) = \{CD\}$,
2) $\mathbf{G_4}(C) = \mathbf{G_4}(SC) = \{C, SC\}$,
3) $\mathbf{G_4}(NCD) = \{NCD\}$,
4) $\mathbf{G_4}(BI) = \{BI\}$,
5) $\mathbf{G_4}(LI) = \mathbf{G_4}(RI) = \{LI, RI\}$,
6) $\mathbf{G_4}(NI) = \{NI\}$.

Remark 9.7. The group $\mathbf{G_8} = \langle\{Id, N1, N2, N12, F, FN1, FN2, FN12\}, \circ\rangle$ is isomorphic to the dihedral group of order 8, i.e. $\mathbf{D}_8$. The latter has generators x, y and can be presented as $\langle x, y \,|\, x^4 = 1, y^2 = 1, yxyx = 1\rangle$. A concrete isomorphism $\iota\colon \mathbf{D_8} \to \mathbf{G_8}$ is determined by where it sends the generators of $\mathbf{D_8}$: $\iota(x) = FN2$ and $\iota(y) = F$. The Cayley table of $\mathbf{G_8}$ thus looks as follows:

$\circ$	Id	$N1$	$N2$	$N12$	F	$FN1$	$FN2$	$FN12$
Id	Id	$N1$	$N2$	$N12$	F	$FN1$	$FN2$	$FN12$
$N1$	$N1$	Id	$N12$	$N2$	$FN2$	$FN12$	F	FN
$N2$	$N2$	$N12$	Id	$N1$	$FN1$	F	$FN12$	$FN2$
$N12$	$N12$	$N2$	$N1$	Id	$FN12$	$FN2$	$FN1$	F
F	F	$FN1$	$FN2$	$FN12$	Id	$N1$	$N2$	$N12$
$FN1$	$FN1$	F	$FN12$	$FN2$	$N2$	$N12$	Id	$N1$
$FN2$	$FN2$	$FN12$	F	$FN1$	$N1$	Id	$N12$	$N2$
$FN12$	$FN12$	$FN2$	$FN1$	F	$N12$	$N2$	$N1$	Id

This group acts faithfully on $\mathcal{G}$, and partitions it into three $\mathbf{G_8}$-orbits:

1) $\mathbf{G}_8(CD) = \mathbf{G}_8(BI) = \{CD, BI\}$,
2) $\mathbf{G}_8(C) = \mathbf{G}_8(SC) = \mathbf{G}_8(LI) = \mathbf{G}_8(RI) = \{C, SC, LI, RI\}$,
3) $\mathbf{G}_8(NCD) = \mathbf{G}_8(NI) = \{NCD, NI\}$.

Lemma 9.7. *Consider a binary, truth-functional connective* $\bullet$*. Then for all formulas* $\varphi, \psi \in \mathcal{L}_\mathsf{S}$ *such that* $\mathsf{S} \models \varphi \bullet \psi$*, the following holds:*

$$\textit{for all } 1 \leq i \leq 4 : \textit{ if } \bullet_i = 0 \textit{ then } \mathsf{S} \models \neg\Delta_i(\varphi, \psi).$$

Proof. Suppose that $\bullet_i = 0$. By definition of the propositional function Δ_i, this means that $\mathsf{S} \models \Delta_i(\varphi, \psi) \to \neg(\varphi \bullet \psi)$. Since $\mathsf{S} \models \varphi \bullet \psi$, it follows that $\mathsf{S} \models \neg\Delta_i(\varphi, \psi)$. □

Remark 9.8. Lemma 9.7 can be seen as a partial converse of Theorem 9.1. To see this more clearly, recall Remark 9.1 about the opposition and implication geometries being defined in terms of $\Delta_1 - \Delta_4$. Theorem 9.1 moves *from* an opposition relation and an implication relation, i.e. $\Delta_1 - \Delta_4$, *to* a binary connective. Lemma 9.7 goes exactly in the other direction: it moves *from* a binary connective *to* $\Delta_1 - \Delta_4$.

Of course, Lemma 9.7 is only a *partial* converse of Theorem 9.1, because it states that $\mathsf{S} \models \neg\Delta_i(\varphi, \psi)$ if $i = 0$, but remains silent about the case $i = 1$. Based on Definition 9.6, one might expect that $\mathsf{S} \not\models \neg\Delta_i(\varphi, \psi)$ if $i = 1$, but this does not hold in general. Consider, for example, the binary connective $\vee = (1, 1, 1, 0)$ and the formulas p and $\neg p$. Since $\mathsf{CPL} \models p \vee \neg p$ and $\vee_1 = 1$, one would erroneously conclude that $\mathsf{CPL} \not\models \neg\Delta_1(p, \neg p)$, i.e. $\mathsf{CPL} \not\models \neg(p \wedge \neg p)$.

Definition 9.11. Let S be a logical system as in Definition 9.1. Recall that $\mathcal{G}_\mathsf{S} = \mathcal{OG}_\mathsf{S} \cup \mathcal{IG}_\mathsf{S}$ is the set of all opposition and implication relations for S. The pair $\mathbb{A}_\mathsf{S} := \langle \mathcal{L}_\mathsf{S}, \mathcal{G}_\mathsf{S} \rangle$ is thus a *relational structure*, in the sense of Dunn and Hardegree (2001). Note that S has a notion of *logical equivalence* $\equiv_\mathsf{S} \subseteq \mathcal{L}_\mathsf{S} \times \mathcal{L}_\mathsf{S}$, defined by $\varphi \equiv_\mathsf{S} \psi :\Leftrightarrow \mathsf{S} \models \varphi \leftrightarrow \psi$. The *equivalence class* of $\varphi \in \mathcal{L}_\mathsf{S}$ is defined as $[\varphi]_{\equiv_\mathsf{S}} := \{\psi \in \mathcal{L}_\mathsf{S} \mid \varphi \equiv_\mathsf{S} \psi\}$. This equivalence relation is actually even a *congruence relation* on $\mathbb{A}_\mathsf{S}$ (Dunn and Hardegree 2001, Definition 2.6.2). In the following definition and lemma, the subscript S will be left implicit.

Definition 9.12. Given the relational structure $\mathbb{A} = \langle \mathcal{L}, \mathcal{G} \rangle$ and the congruence relation $\equiv$ on $\mathbb{A}$, we define the *quotient structure* $\mathbb{A}/\equiv\, := \langle \mathcal{L}/\equiv,\ \mathcal{G}/\equiv \rangle$, with $\mathcal{L}/\equiv\, := \{[\varphi] \mid \varphi \in \mathcal{L}\}$, and each relation $R/\equiv\ \in\ \mathcal{G}/\equiv$ defined as follows: $([\varphi], [\psi]) \in R/\equiv\, :\Leftrightarrow\ \exists \psi' \in [\psi] \colon (\varphi, \psi') \in R$ (Dunn and Hardegree 2001, p. 23).

Lemma 9.8. *Since $BI \in \mathcal{G}$, the quotient structure $\mathbb{A}/\equiv$ will also contain the relation $BI/\equiv$. But since $\equiv$ and BI are actually the same relation, $BI/\equiv$ is the* identity relation *on $\mathcal{L}/\equiv$.*

Proof. For any $[\varphi], [\psi] \in \mathcal{L}/\equiv$, we have:

$$\begin{aligned} ([\varphi],[\psi]) \in BI/\equiv \quad &\Leftrightarrow \quad \exists \psi' \in [\psi] \colon (\varphi, \psi') \in BI \\ &\Leftrightarrow \quad \exists \psi' \in \mathcal{L} \colon \psi \equiv \psi' \text{ and } \varphi \equiv \psi' \\ &\Leftrightarrow \quad \varphi \equiv \psi \\ &\Leftrightarrow \quad [\varphi] = [\psi]. \end{aligned}$$

Hence, each $[\varphi]$ stands in the $BI/\equiv$-relation to exactly one element: itself. □

10 Conclusion

The overarching goal of this thesis has been to show that, despite its origins in computer science and game theory, the dynamic turn in epistemic logic also has great philosophical significance. I will now summarize the main results obtained in this thesis, and assess their contribution toward achieving the overarching goal.

The main line of argumentation essentially consists in a sequence of case studies, in which some system of dynamic epistemic logic is applied to a notion or theorem in a philosophically fruitful way. The core chapters of this thesis are thus Chapters 5, 6, 7 and 8, in which dynamic epistemic logic is applied to notions or theorems from game theory, epistemology, cognitive science and logical geometry, respectively. The other chapters provide some more background for these case studies and introduce the technical notions.

Chapter 1 discusses the philosophical and historical background of this thesis. Most importantly, it introduces the distinction between the weak and the strong interpretation of the dynamic turn in epistemic logic, and argues that while the dynamic turn is not philosophically relevant according to the weak interpretation, it is highly philosophical relevant according to the strong interpretation. This distinction is important for the remainder of the thesis, since three of the four case studies on dynamic epistemic logic (viz. those in Part II: Chapters 5, 6 and 7) are clear illustrations of the strong interpretation of the dynamic turn in epistemic logic.

Part I introduces all the technical notions that are needed in the case studies. The three case studies in Part II have in common that they involve not only (dynamic epistemic) logic, but also probability. Hence, it is important to get a clear view of the relationship between (dynamic epistemic) logic and probability theory. There has recently been a vast amount of research on this relationship, which has led to a very interesting, but also quite chaotic literature. Therefore, I have attempted to clarify matters in two consecutive steps. Chapter 2 provides a large-scale overview of the various proposals to combine logic and probability, and shows that they can be organized in a systematical and logically meaningful way. This overview includes discussions of systems such as probabilistic semantics and first-order probabilistic logic, but does not yet focus on probabilistic systems of dynamic epistemic logic. Its importance thus lies in the fact that it sketches the broader context for these probabilistic dynamic epistemic logics: in this thesis, these systems are naturally seen as belonging to the family of dynamic epistemic logics; however, Chapter 2 shows that they can equally naturally be seen as belonging to the family of probabilistic logics.

Next, Chapter 3 focuses on the kind of probabilistic logics that are used most frequently in this thesis, viz. probabilistic epistemic logics (static as well as dynamic). It introduces the syntax and semantics of some important systems in great detail, since the case studies in Part II all involve variations or extensions of these systems. It also discusses the relationship between public announcement and Bayesian conditionalization, and shows how the subtlety of this relationship derives from the ability of probabilistic dynamic epistemic logic to deal with higher-order information.

One of the case studies in Part II (viz. Chapter 6 on the Lockean thesis) involves not only probabilistic Kripke models (which were introduced in Chapter 3), but also epistemic plausibility models. However, in the literature, the latter are often defined in two related, but subtly different ways. Therefore, Chapter 4 provides a detailed introduction to epistemic plausibility models and their model theory. I also use these model-theoretical results to argue that one way of defining these models is superior to the other, since it achieves a better equilibrium between philosophical applicability and mathematical elegance. Finally, although probabilistic Kripke models and epistemic plausibility models capture the agents' soft information (belief) in radically different ways (qualitatively vs. quantitatively), throughout this chapter I emphasize that there are also some fundamental similarities between both kinds of models (for example, the notion of *uniformity*, which states that epistemically indistinguishable states have identical soft information—i.e. identical probability functions or identical plausibility orderings).

Part II presents the three case studies on the strong interpretation of the dynamic turn in epistemic logic. In Chapter 5, I discuss Aumann's celebrated agreeing to disagree theorem from game theory, and argue that Aumann's original formulation fails to fully capture the dynamics behind the agreement theorem (both in its formulation and in its semantic setup). I show how a more natural formulation of the theorem can be obtained in a system of probabilistic dynamic epistemic logic. Furthermore, I show how explicitly representing the dynamics behind the agreement theorem leads to a significant conceptual elucidation concerning the role of common knowledge in the agreement theorem. It turns out that common knowledge is less central to this theorem than is often thought: common knowledge is the result of communication, so if the communication dynamics is explicitly represented in the agreement theorem, there is no need anymore to *assume* common knowledge (as this will now *follow* from the communication protocol).

Next, Chapter 6 discusses the Lockean thesis about belief and degrees of belief. A well-known problem of this thesis is that it yields a notion of belief that is not closed under conjunction. After pointing out that this is a static problem, I examine how the Lockean thesis fares from a dynamic perspective. I compare the notions of high degree of (conditional) belief with the corresponding 'strictly qualitative' notions of (conditional) belief, and show that accepting the Lockean thesis for belief (and a slightly more sophisticated version for conditional belief) leads to a significant and unexpected unification in the dynamic behavior of (conditional) belief and high degree of (conditional) belief with respect to public announcements. Finally, I argue that this technical observation constitutes a methodological and perhaps even a philosophical argument in favor of the Lockean thesis.

The final case study on the strong interpretation of the dynamic turn in epistemic logic is Chapter 7, which discusses the epistemic and cognitive aspects of surprise. After providing a brief overview of existing work on surprise, I argue that the main formal accounts of surprise in logic and artificial intelligence fail to do justice to its essentially dynamic nature. I then propose a new formalization of surprise, using a system of probabilistic dynamic epistemic logic. This system is able to naturally capture the dynamic nature of surprise: this is clear from the logic's semantics as well as its proof theory. Finally, I show that this system is able to capture several key aspects of surprise, such as its role in belief revision and its transitory nature. The former can also be captured by other formalizations, but the latter can only be adequately represented in the current system, since it is a manifestation of the dynamic nature of surprise.

Part III is concerned with logical geometry, both in its relation to the dynamic turn in epstemic logic and as an independent topic of interest. Chapter 8 presents the fourth case study on the dynamic turn: it shows how dynamic epistemic logic gives rise to non-trivial Aristotelian squares and larger diagrams (such as hexagons, octagons, and rhombic dodecahedrons). These diagrams not only extend the scope of logical geometry, but they are also important for its philosophical foundations. It is clear that this application of dynamic epistemic logic is not an illustration of the strong interpretation of the dynamic turn, since it does not have the purpose of uncovering hidden dynamic aspects of some seemingly static notion.

Unlike the case studies in Part II, which were all concerned with topics from well-established fields such as game theory, epistemology and cognitive science, the case study in Chapter 8 is concerned with a much less widely known topic, viz. logical geometry. Hence, to provide some more context to this last case study, Chapter 9 shows how several apparently unrelated notions and theorems in logical geometry can be unified by viewing them from the common perspective of information.

Bibliography

Abadi, Martín and Halpern, Joseph Y. Decidability and expressiveness for first-order logics of probability. *Information and Computation*, 112:1–36, 1994. (Cited on p. 62.)

Adams, Ernest W. *A Primer of Probability Logic*. CSLI Publications, Stanford, CA, 1998. (Cited on p. 42, 43, 45, 47, and 52.)

Adams, Ernest W. and Levine, Howard P. On the uncertainties transmitted from premises to conclusions in deductive inferences. *Synthese*, 30:429–460, 1975. (Cited on p. 40 and 48.)

Ågotnes, Thomas and Alechina, Natasha. The dynamics of syntactic knowledge. *Journal of Logic and Computation*, 17:83–116, 2007. (Cited on p. 30.)

Alchourrón, Carlos, Gärdenfors, Peter, and Makinson, David. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50:510–530, 1985. (Cited on p. 95 and 211.)

Arló Costa, Horacio. Non-adjunctive inference and classical modalities. *Journal of Philosophical Logic*, 34:581–605, 2005. (Cited on p. 55.)

Arló Costa, Horacio. The logic of conditionals. In Zalta, Edward N., editor, *Stanford Encyclopedia of Philosophy*. CSLI Publications, Stanford, CA, 2007. (Cited on p. 47.)

Artemov, Sergei and Kuznets, Roman. Logical omniscience via proof complexity. In Ésik, Zoltán, editor, *Proceedings of Computer Science Logic 2006*, Lecture Notes in Computer Science 4207, pages 135–149. Springer, Berlin, 2006. (Cited on p. 30.)

Aucher, Guillaume. A combined system for update logic and belief revision. Master's thesis, Institute for Logic, Language and Computation, Universiteit van Amsterdam, Amsterdam, 2003. (Cited on p. 121.)

Aumann, Robert. Agreeing to disagree. *Annals of Statistics*, 4:1236–1239, 1976. (Cited on p. 133, 135, 149, and 173.)

Aumann, Robert. Backward induction and common knowledge of rationality. *Games and Economic Behavior*, 8:6–19, 1995. (Cited on p. 21.)

Bacchus, Fahiem. *Representing and Reasoning with Probabilistic Knowledge*. MIT Press, Cambridge, MA, 1990. (Cited on p. 57, 60, 61, and 62.)

Bacharach, Michael. Some extensions of a claim of Aumann in an axiomatic model of knowledge. *Journal of Economic Theory*, 37:167–190, 1985. (Cited on p. 136.)

Balbiani, Philippe, Baltag, Alexandru, van Ditmarsch, Hans P., Herzig, Andreas, Hoshi, Tomohiro, and De Lima, Tiago. 'Knowable' as 'known after an announcement'. *Review of Symbolic Logic*, 1:305–334, 2008. (Cited on p. 29.)

Baltag, Alexandru. Interview. In Hendricks, Vincent F. and Pritchard, Duncan, editors, *Epistemology: 5 Questions*, pages 21–37. Automatic Press, Copenhagen, 2008. (Cited on p. 183.)

Baltag, Alexandru. Research profile. Available online at `http://www.vub.ac.be/CLWF/SS/ResearchProfile.pdf`, 2011. (Cited on p. 183.)

Baltag, Alexandru and Moss, Lawrence S. Logics for epistemic programs. *Synthese*, 139:165–224, 2004. (Cited on p. 24, 95, 107, and 170.)

Baltag, Alexandru and Smets, Sonja. A qualitative theory of dynamic interactive belief revision. In Bonanno, Giacomo, van der Hoek, Wiebe, and Woolridge, Michael, editors, *Texts in Logic and Games*, pages 9–58. Amsterdam University Press, Amsterdam, 2008. (Cited on p. 24, 103, 108, 109, 110, 118, 125, 126, 174, and 175.)

Baltag, Alexandru, Moss, Lawrence S., and Solecki, Slawomir. The logic of common knowledge, public announcements, and private suspicions. In Gilboa, Ithzak, editor, *Proceedings of the 7th Conference on Theoretical*

Aspects of Rationality and Knowledge, pages 43–56. Morgan Kaufmann, Evanston, IL, 1998. (Cited on p. 24, 95, and 107.)

Baltag, Alexandru, Smets, Sonja, and Zvesper, Jonathan A. Keep 'hoping' for rationality. a solution to the backward induction paradox. *Synthese*, 169:303–333, 2009. (Cited on p. 21.)

Barwise, Jon. Three views of common knowledge. In Vardi, Moshe Y., editor, *Proceedings of the 2nd Conference on Theoretical Aspects of Reasoning and Knowledge*. Morgan Kaufmann, Pacific Grove, CA, 1988. (Cited on p. 170.)

Bates, Joseph. The role of emotion in believable agents. *Communications of the ACM*, 37(7):122–125, 1994. (Cited on p. 192.)

Becker, Christian Werner, Kopp, Stefan, and Wachsmuth, Ipke. Simulating the emotion dynamics of a multimodal conversational agent. In André, Elisabeth, Dybkjær, Laila, Minker, Wolfgang, and Heisterkamp, Paul, editors, *Affective Dialogue Systems (ADS 2004)*, Lecture Notes in Computer Science 3068, pages 154–165. Springer, Berlin, 2004. (Cited on p. 192.)

Béziau, Jean-Yves. New light on the square of oppositions and its nameless corner. *Logical Investigations*, 10:218–232, 2003. (Cited on p. 235, 240, 250, 281, and 282.)

Béziau, Jean-Yves and Payette, Gillman. Preface. In Béziau, Jean-Yves and Payette, Gillman, editors, *The Square of Opposition. A General Framework for Cognition*, pages 9–22. Peter Lang, Bern, 2012. (Cited on p. 286.)

Blackburn, Patrick, de Rijke, Maarten, and Venema, Yde. *Modal Logic*. Cambridge University Press, Cambridge, 2001. (Cited on p. 65, 80, 117, 153, and 161.)

Blanché, Robert. Quantity, modality, and other kindred systems of categories. *Mind*, 61:369–375, 1952. (Cited on p. 229 and 250.)

Blanché, Robert. Sur l'opposition des concepts. *Theoria*, 19:89–130, 1953. (Cited on p. 229.)

Blanché, Robert. Opposition et négation. *Revue philosophique de la France et de l'étranger*, 167:187–216, 1957. (Cited on p. 229.)

Blanché, Robert. *Structures Intellectuelles. Essai sur l'organisation systématique des concepts*. Librairie Philosophique J. Vrin, Paris, 1966. (Cited on p. 229 and 250.)

Bocheński, Józef M. *A Precis of Mathematical Logic*. Reidel, Dordrecht, 1959. (Cited on p. 265 and 266.)

Boh, Ivan. *Epistemic Logic in the Later Middle Ages*. Routledge, London, 1993. (Cited on p. 17.)

Boh, Ivan. Four phases of medieval epistemic logic. *Theoria*, 66:129–144, 2000. (Cited on p. 17.)

Bonanno, Giacomo and Dégremont, Cédric. Game theory and logic. In Baltag, Alexandru and Smets, Sonja, editors, *Johan F. A. K. van Benthem on Logical and Informational Dynamics*. Springer, Dordrecht, forthcoming. (Cited on p. 104.)

Bonanno, Giacomo and Nehring, Klaus. Agreeing to disagree: a survey. Manuscript, 1997. (Cited on p. 135, 136, and 141.)

Boole, George. *An Investigation of the Laws of Thought, on which are Founded the Mathematical Theories of Logic and Probabilities*. Walton and Maberly, London, 1854. (Cited on p. 37.)

Boolos, George S., Burgess, John P., and Jeffrey, Richard C. *Computability and Logic (Fifth Edition)*. Cambridge University Press, Cambridge, 2007. (Cited on p. 62.)

Bourbaki, Nicolas. Sur le theorème de Zorn. *Archiv der Mathematik*, 2:434–437, 1949. (Cited on p. 144.)

Boutilier, Craig. Conditional logics of normality: a modal approach. *Artificial Intelligence*, 68:87–154, 1994. (Cited on p. 126.)

Brogaard, Berit and Salerno, Joe. Fitch's paradox of knowability. In Zalta, Edward N., editor, *Stanford Encyclopedia of Philosophy*. CSLI Publications, Stanford, CA, 2009. (Cited on p. 29.)

Burgess, John P. Probability logic. *Journal of Symbolic Logic*, 34:264–274, 1969. (Cited on p. 55.)

Campos-Benítez, Juan M. The medieval modal octagon and the S5 Lewis modal system. In Béziau, Jean-Yves and Jacquette, Dale, editors, *The Square of Opposition. A General Framework for Cognition*, pages 99–116. Springer, Basel, 2012. (Cited on p. 281.)

Carnap, Rudolf. *Meaning and Necessity. A Study in Semantics and Modal Logic*. University of Chicago Press, Chicago, IL, 1947. (Cited on p. 52 and 255.)

Carnap, Rudolf. *Logical Foundations of Probability*. University of Chicago Press, Chicago, IL, 1950. (Cited on p. 37.)

Carnielli, Walter and Pizzi, Claudio. *Modalities and Multimodalities*. Springer, Berlin, 2008. (Cited on p. 247.)

Charlesworth, William R. Instigation and maintenance of curiosity behavior as a function of surprise versus novel and familiar stimuli. *Child Development*, 35:1169–1186, 1964. (Cited on p. 189, 190, and 210.)

Chatti, Saloua. Logical oppositions in Arabic logic: Avicenna and Averroes. In Béziau, Jean-Yves and Jacquette, Dale, editors, *Around and Beyond the Square of Opposition*, pages 21–40. Springer, Basel, 2012. (Cited on p. 247.)

Chatti, Saloua and Schang, Fabien. The cube, the square and the problem of existential import. *History and Philosophy of Logic*, 32:101–132, 2013. (Cited on p. 249 and 250.)

Chellas, Brian F. *Modal Logic. An Introduction*. Cambridge University Press, Cambridge, 1980. (Cited on p. 55.)

Childers, Timothy. *Philosophy and Probability*. Oxford University Press, Oxford, 2013. (Cited on p. 40.)

Chow, Timothy Y. The surprise examination or unexpected hanging paradox. *American Mathematical Monthly*, 105:41–51, 1998. (Cited on p. 187.)

Correia, Manuel. Boethius on the square of opposition. In Béziau, Jean-Yves and Jacquette, Dale, editors, *Around and Beyond the Square of Opposition*, pages 41–52. Springer, Basel, 2012. (Cited on p. 258 and 259.)

Cross, Charles B. From worlds to probabilities: A probabilistic semantics for modal logic. *Journal of Philosophical Logic*, 22:169–192, 1993. (Cited on p. 44.)

Czeżowski, Tadeusz. On certain peculiarities of singular propositions. *Mind*, 64: 392–395, 1955. (Cited on p. 250.)

D'Alfonso, Duilio. The square of opposition and generalized quantifiers. In Béziau, Jean-Yves and Jacquette, Dale, editors, *Around and Beyond the Square of Opposition*, pages 219–227. Springer, Basel, 2012. (Cited on p. 288.)

Davey, Brian A. and Priestley, Hilary A. *Introduction to Lattices and Order (Second Edition)*. Cambridge University Press, Cambridge, 2002. (Cited on p. 218 and 274.)

Davidson, Donald. Rational animals. *Dialectica*, 36:317–327, 1982. (Cited on p. 190.)

De Bona, Glauber, Cozman, Fabio G., and Finger, Marcelo. Towards classifying propositional probabilistic logics, September 17–18. Talk at Progic 2013, Munich, 2013. (Cited on p. 38.)

de Bruin, Boudewijn. Common knowledge of payoff uncertainty in games. *Synthese*, 163:79–97, 2008a. (Cited on p. 21 and 104.)

de Bruin, Boudewijn. Common knowledge of rationality in extensive games. *Notre Dame Journal of Formal Logic*, 49:261–280, 2008b. (Cited on p. 104.)

de Bruin, Boudewijn. *Explaining Games: The Epistemic Programme in Game Theory*. Springer, Dordrecht, 2010. (Cited on p. 22 and 104.)

de Finetti, Bruno. La prévision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7:1–68 (translated as 'Foresight. Its logical laws, its subjective sources', in Henry E. Kyburg, Jr. and Howard E. Smokler, editors, *Studies in Subjective Probability*, pages 53–118, Huntington, NY, Krieger Publishing Company, 1980), 1937. (Cited on p. 37.)

De Morgan, Augustus. *Formal Logic*. Taylor and Walton, London, 1847. (Cited on p. 37.)

de Pater, Wim A. and Vergauwen, Roger. *Logica: Formeel en Informeel*. Leuven University Press, Leuven, 2005. (Cited on p. 233.)

De Raedt, Luc, Frasconi, Paolo, Kersting, Kristian, and Muggleton, Stephen, editors. *Probabilistic Inductive Logic Programming*. Lecture Notes in Computer Science 4911. Springer, New York, NY, 2008. (Cited on p. 41.)

de Vink, Erik P. and Rutten, Jan. Bisimulation for probabilistic transition systems: a coalgebraic approach. *Theoretical Computer Science*, 221:271–293, 1999. (Cited on p. 81.)

Dégremont, Cédric and Roy, Olivier. Agreement theorems in dynamic-epistemic logic. In He, Xiangdong, Horty, John, and Pacuit, Eric, editors, *Logic, Rationality, and Interaction. LORI 2009 Proceedings*, Lecture Notes in Computer Science 5834, pages 105–118. Springer, Berlin, 2009. (Cited on p. 134, 136, and 173.)

Dégremont, Cédric and Roy, Olivier. Agreement theorems in dynamic-epistemic logic. GRIPh Working Paper 0902. Available online at `http://www.rug.nl/filosofie/onderzoek/workingpapers/WorkingPaperRoy.pdf`, 2010. (Cited on p. 120.)

Dégremont, Cédric and Roy, Olivier. Agreement theorems in dynamic-epistemic logic. *Journal of Philosophical Logic*, 41:735–764, 2012. (Cited on p. 134, 136, and 173.)

Demey, Lorenz. Agreeing to disagree in probabilistic dynamic epistemic logic. Master's thesis, Institute for Logic, Language and Computation, Universiteit van Amsterdam, Amsterdam, 2010. (Cited on p. 134, 140, 143, 151, 153, and 161.)

Demey, Lorenz. Some remarks on the model theory of epistemic plausibility models. *Journal of Applied Non-Classical Logics*, 21:375–395, 2011a. (Cited on p. 7.)

Demey, Lorenz. Joint book review of Vincent Hendricks and Olivier Roy (eds.), Epistemic Logic: 5 Questions and Vincent Hendricks and Duncan Pritchard (eds.), Epistemology: 5 Questions. *Tijdschrift voor Filosofie*, 73:596–598, 2011b. (Cited on p. 171.)

Demey, Lorenz. Agreement theorems in probabilistic dynamic epistemic logic. In Grossi, Davide, Minică, Ştefan, Rodenhäuser, Ben, and Smets, Sonja, editors, *Logic and Interactive Rationality Yearbook 2010*, pages 1–26. Institute for Logic, Language and Computation, Amsterdam, 2011c. (Cited on p. 7.)

Demey, Lorenz. Algebraic aspects of duality diagrams. In Cox, Philip T., Plimmer, Beryl, and Rodgers, Peter, editors, *Diagrammatic Representation and Inference*, Lecture Notes in Computer Science 7352, pages 300–302. Springer, Berlin, 2012a. (Cited on p. 288.)

Demey, Lorenz. Narrative and information: Comment on Löwe. In Allo, Patrick and Primiero, Giuseppe, editors, *Proceedings of the Third Workshop in the Philosophy of Information*, pages 29–34. KVAB, Brussels, 2012b. (Cited on p. 267.)

Demey, Lorenz. Structures of oppositions in public announcement logic. In Béziau, Jean-Yves and Jacquette, Dale, editors, *Around and Beyond the Square of Opposition*, pages 313–339. Springer, Basel, 2012c. (Cited on p. 7.)

Demey, Lorenz. Looking for the right notion of epistemic plausibility model. In Van Kerkhove, Bart, Libert, Thierry, Vanpaemel, Geert, and Marage, Pierre, editors, *Logic, Philosophy and History of Science in Belgium II. Proceedings of the Young Researcher Days 2010*, pages 73–78. KVAB, Brussels, 2012d. (Cited on p. 7.)

Demey, Lorenz. De dynamische wending in de epistemische logica. *Submitted*, 2013a. (Cited on p. 7.)

Demey, Lorenz. Surprise in probabilistic dynamic epistemic logic. In Christoff, Zoé, Galeazzi, Paolo, Gierasimczuk, Nina, and Smets, Sonja, editors, *Logic and Interactive Rationality Yearbook 2012*. Institute for Logic, Language and Computation, Amsterdam, 2013b. (Cited on p. 7.)

Demey, Lorenz. Contemporary epistemic logic and the Lockean thesis. *Foundations of Science*, forthcoming a. (Cited on p. 7.)

Demey, Lorenz. The dynamics of surprise. *Logique et Analyse*, forthcoming b. (Cited on p. 7.)

Demey, Lorenz. Ockham on the (in)fallibility of intuitive cognition. *Logical Analysis and History of Philosophy*, forthcoming c. (Cited on p. 21.)

Demey, Lorenz. Agreeing to disagree in probabilistic dynamic epistemic logic. *Synthese*, forthcoming d. (Cited on p. 7.)

Demey, Lorenz and Kooi, Barteld P. Logic and probabilistic update. In Baltag, Alexandru and Smets, Sonja, editors, *Johan F. A. K. van Benthem on Logical and Informational Dynamics*. Springer, Dordrecht, forthcoming. (Cited on p. 7.)

Demey, Lorenz and Sack, Joshua. Epistemic probabilistic logic. In van Ditmarsch, Hans P., Halpern, Joseph Y., van der Hoek, Wiebe, and Kooi, Barteld P., editors, *Handbook of Logics for Knowledge and Belief*. College Publications, London, forthcoming. (Cited on p. 7, 76, and 98.)

Demey, Lorenz and Smessaert, Hans. Logical geometry and Hasse diagrams. Manuscript, 2013a. (Cited on p. 225 and 249.)

Demey, Lorenz and Smessaert, Hans. Visual-logical congruity in Aristotelian diagrams. Manuscript, 2013b. (Cited on p. 225.)

Demey, Lorenz, Kooi, Barteld P., and Sack, Joshua. Logic and probability. In Zalta, Edward N., editor, *Stanford Encyclopedia of Philosophy*. CSLI Publications, Stanford, CA, 2013. (Cited on p. 7 and 37.)

Dempster, Arthur P. A generalization of Bayesian inference. *Journal of the Royal Statistical Society*, 30:205–247, 1968. (Cited on p. 53.)

Douven, Igor and Meijs, Wouter. Measuring coherence. *Synthese*, 156:405–425, 2007. (Cited on p. 167.)

Doyle, John J. The hexagon of relationships. *The Modern Schoolman*, 29:93–97, 1952. (Cited on p. 282.)

Dubois, Didier and Prade, Henri. Accepted beliefs, revision and bipolarity in the possibilistic framework. In Huber, Franz and Schmidt-Petri, Christoph, editors, *Degrees of Belief*, pages 161–184. Springer, Dordrecht, 2009. (Cited on p. 171.)

Dubois, Didier and Prade, Henri. From Blanché's hexagonal organization of concepts to formal concept analysis and possibility theory. *Logica Universalis*, 6:149–169, 2012. (Cited on p. 264.)

Duc, Ho Ngoc. Reasoning about rational, but not logically omniscient, agents. *Journal of Logic and Computation*, 7:633–648, 1997. (Cited on p. 30.)

Dunn, J. Michael and Hardegree, Gary M. *Algebraic Methods in Philosophical Logic*. Oxford University Press, Oxford, 2001. (Cited on p. 290.)

Dutilh Novaes, Catarina. *Formal Languages in Logic. A Philosophical and Cognitive Analysis*. Cambridge University Press, Cambridge, 2012. (Cited on p. 28.)

Eagle, Anthony, editor. *Philosophy of Probability: Contemporary Readings*. Routledge, London, 2010. (Cited on p. 40.)

Edgington, Dorothy. Conditionals. In Zalta, Edward N., editor, *Stanford Encyclopedia of Philosophy*. CSLI Publications, Stanford, CA, 2006. (Cited on p. 47.)

Eells, Ellery and Fitelson, Branden. Measuring confirmation and evidence. *Journal of Philosophy*, 97:663–672, 2000. (Cited on p. 167.)

Eells, Ellery and Skyrms, Brian. *Probability and Conditionals. Belief Revision and Rational Decision*. Cambridge University Press, Cambridge, 1994. (Cited on p. 47.)

El-Nasr, Magy S. Modelling emotion dynamics in intelligent agents. Master's thesis, Texas A&M University, College Station, TX, 1998. (Cited on p. 192.)

Enderton, Herbert. *A Mathematical Introduction to Logic (Second Edition)*. Academic Press, San Diego, CA, 2001. (Cited on p. 262.)

Engel, Konrad. *Sperner Theory*. Cambridge University Press, Cambridge, 1997. (Cited on p. 273.)

Faghihi, Usef, Poirier, Pierre, and Larue, Othalia. Emotional cognitive architectures. In D'Mello, Sidney, Graesser, Arthur, Schuller, Björn, and Martin, Jean-Claude, editors, *Affective Computing and Intelligent Interaction (ACII 2011)*, Lecture Notes in Computer Science 6974, pages 487–496. Springer, Berlin, 2011. (Cited on p. 192.)

Fagin, Ronald and Halpern, Joseph Y. Reasoning about knowledge and probability. *Journal of the ACM*, 41:340–367, 1994. (Cited on p. 76, 78, 83, 192, 198, and 205.)

Fagin, Ronald, Halpern, Joseph Y., and Megiddo, Nimrod. A logic for reasoning about probabilities. *Information and Computation*, 87:78–128, 1990. (Cited on p. 65 and 71.)

Fagin, Ronald, Halpern, Joseph Y., Moses, Yoram, and Vardi, Moshe Y. *Reasoning about Knowledge*. MIT Press, Cambridge, MA, 1995. (Cited on p. 22.)

Feinberg, Yossi. Characterizing common priors in the form of posteriors. *Journal of Economic Theory*, 91:127–179, 2000. (Cited on p. 136.)

Fitch, Frederick. A logical analysis of some value concepts. *Journal of Symbolic Logic*, 28:135–142, 1963. (Cited on p. 29.)

Fitelson, Branden. Inductive logic. In Sarkar, Sahotra and Pfeifer, Jessica, editors, *The Philosophy of Science: An Encyclopedia*, pages 384–394. Routledge, New York, NY, 2006. (Cited on p. 40.)

Fitting, Melvin and Mendelsohn, Richard L. *First-Order Modal Logic*. Kluwer, Dordrecht, 1998. (Cited on p. 217 and 247.)

Foley, Richard. The epistemology of belief and the epistemology of degrees of belief. *American Philosophical Quarterly*, 29:111–121, 1992. (Cited on p. 167 and 171.)

Gamut, L. T. F. *Logic, Language, and Meaning. Volume 1: Introduction to Logic*. University of Chicago Press, Chicago, IL, 1991. (Cited on p. 267.)

Gärdenfors, Peter. Qualitative probability as an intensional logic. *Journal of Philosophical Logic*, 4:171–185, 1975a. (Cited on p. 55.)

Gärdenfors, Peter. Some basic theorems of qualitative probability. *Studia Logica*, 34:257–264, 1975b. (Cited on p. 55.)

Gärdenfors, Peter. *Knowledge in Flux*. MIT Press, Cambridge, MA, 1988. (Cited on p. 95 and 211.)

Geanakoplos, John D. and Polemarchakis, H. M. We can't disagree forever. *Journal of Economic Theory*, 28:192–200, 1982. (Cited on p. 136.)

Geiss, Christel and Geiss, Stefan. An introduction to probability theory. Course notes. Available online at `http://users.jyu.fi/~geiss/scripts/introduction-probability.pdf`, 2009. (Cited on p. 67.)

Georgakopoulos, George, Kavvadias, Dimitris, and Papadimitriou, Christos H. Probabilistic satisfiability. *Journal of Complexity*, 4:1–11, 1988. (Cited on p. 53.)

Gerbrandy, Jelle and Groeneveld, Willem. Reasoning about information change. *Journal of Logic, Language and Information*, 6:147–169, 1997. (Cited on p. 24, 83, and 107.)

Gerla, Giangiacomo. Inferences in probability logic. *Artificial Intelligence*, 70: 33–52, 1994. (Cited on p. 41.)

Gillies, Donald. *Philosophical Theories of Probability*. Routledge, London, 2000. (Cited on p. 40.)

Girard, Patrick. *Modal Logic for Belief and Preference Change*. PhD thesis, Stanford University, Stanford, CA, 2008. (Cited on p. 23 and 121.)

Gochet, Paul and Gribomont, Pascal. Epistemic logic. In Gabbay, Dov M. and Woods, John, editors, *Handbook of the History of Logic, Volume 7*, pages 99–195. Elsevier, Amsterdam, 2006. (Cited on p. 17.)

Goldblatt, Robert. Deduction systems for coalgebras over measurable spaces. *Journal of Logic and Computation*, 20:1069–1100, 2010. (Cited on p. 68.)

Goldman, Alan J. and Tucker, Albert W. Theory of linear programming. In Kuhn, Harold W. and Tucker, Albert W., editors, *Linear Inequalities and Related Systems. Annals of Mathematics Studies 38*, pages 53–98. Princeton University Press, Princeton, NJ, 1956. (Cited on p. 48.)

Goldman, Alvin. What is justified belief? In Pappas, George S., editor, *Justification and Knowledge*, pages 1–23. Reidel, Dordrecht, 1979. (Cited on p. 185.)

Goldman, Alvin. *Knowledge in a Social World*. Oxford University Press, Oxford, 1999. (Cited on p. 170.)

Gombocz, Wolfgang Leopold. Apuleius is better still: a correction to the square of opposition. *Phronesis*, 43:124–131, 1990. (Cited on p. 258.)

Goosens, William K. Alternative axiomatizations of elementary probability theory. *Notre Dame Journal of Formal Logic*, 20:227–239, 1979. (Cited on p. 44.)

Goranko, Valentin and Otto, Martin. Model theory of modal logic. In *Handbook of Modal Logic*, pages 249–330. Elsevier, Amsterdam, 2006. (Cited on p. 108 and 113.)

Grüne-Yanoff, Till and Lehtinen, Aki. Philosophy of game theory. In Mäki, Uskala, editor, *Philosophy of Economics*, pages 531–576. Elsevier, Amsterdam, 2012. (Cited on p. 30.)

Haenni, Rolf and Lehmann, Norbert. Probabilistic argumentation systems: a new perspective on Dempster-Shafer theory. *International Journal of Intelligent Systems*, 18:93–106, 2003. (Cited on p. 53.)

Haenni, Rolf, Romeijn, Jan-Willem, Wheeler, Gregory, and Williamson, Jon. *Probabilistic Logics and Probabilistic Networks*. Springer, Dordrecht, 2011. (Cited on p. 53.)

Hailperin, Theodore. Best possible inequalities for the probability of a logical function of events. *American Mathematical Monthly*, 72:343–359, 1965. (Cited on p. 51.)

Hailperin, Theodore. Probability logic. *Notre Dame Journal of Formal Logic*, 25:198–212, 1984. (Cited on p. 43 and 51.)

Hailperin, Theodore. *Boole's Logic and Probability*. North-Holland, Amsterdam, 1986. (Cited on p. 51.)

Hailperin, Theodore. The development of probability logic from Leibniz to Maccoll. *History and Philosophy of Logic*, 9:131–191, 1988. (Cited on p. 37.)

Hailperin, Theodore. Probability logic in the twentieth century. *History and Philosophy of Logic*, 12:71–110, 1991. (Cited on p. 37.)

Hailperin, Theodore. *Sentential Probability Logic: Origins, Development, Current Status, and Technical Applications*. Lehigh University Press, Bethlehem, PA, 1996. (Cited on p. 37 and 51.)

Hájek, Alan. Probability, logic, and probability logic. In Goble, Lou, editor, *The Blackwell Guide to Philosophical Logic*, pages 362–384. Blackwell, Oxford, 2001. (Cited on p. 37.)

Hájek, Alan. Interpretations of probability. In Zalta, Edward N., editor, *Stanford Encyclopedia of Philosophy*. CSLI Publications, Stanford, CA, 2011. (Cited on p. 40.)

Hájek, Alan and Hartmann, Stephan. Bayesian epistemology. In Dancy, Jonathan, Sosa, Ernest, and Steup, Matthias, editors, *A Companion to Epistemology*, pages 93–106. Blackwell, Oxford, 2010. (Cited on p. 41.)

Hajek, Petr. Fuzzy logic. In Zalta, Edward N., editor, *Stanford Encyclopedia of Philosophy*. CSLI Publications, Stanford, CA, 2010. (Cited on p. 41.)

Halpern, Joseph Y. An analysis of first-order logics of probability. *Artificial Intelligence*, 46:311–350, 1990. (Cited on p. 60 and 62.)

Halpern, Joseph Y. The relationship between knowledge, belief, and certainty. *Annals of Mathematics and Artificial Intelligence*, 4:301–322 (errata appeared in the same journal, 26:59–61, 1999), 1991. (Cited on p. 74.)

Halpern, Joseph Y. Should knowledge entail belief? *Journal of Philosophical Logic*, 25:483–494, 1996. (Cited on p. 107.)

Halpern, Joseph Y. Substantive rationality and backward induction. *Games and Economic Behavior*, 37:425–435, 2001. (Cited on p. 21.)

Halpern, Joseph Y. Characterizing the common prior assumption. *Journal of Economic Theory*, 106:316–355, 2002. (Cited on p. 155.)

Halpern, Joseph Y. *Reasoning about Uncertainty*. MIT Press, Cambridge, MA, 2003. (Cited on p. 41 and 79.)

Halpern, Joseph Y. and Moses, Yoram. Knowledge and common knowledge in a distributed environment. *Journal of the ACM*, 37:549–587, 1990. (Cited on p. 20, 133, and 170.)

Halpern, Joseph Y. and Moses, Yoram. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54:319–379, 1992. (Cited on p. 243.)

Halpern, Joseph Y. and Pucella, Riccardo. Dealing with logical omniscience: Expressiveness and pragmatics. *Artificial Intelligence*, 175:220–235, 2011. (Cited on p. 30.)

Halpern, Joseph Y. and Rabin, Michael O. A logic to reason about likelihood. *Artificial Intelligence*, 32:379–405, 1987. (Cited on p. 54.)

Halpern, Joseph Y., Samet, Dov, and Segev, Ella. Defining knowledge in terms of belief: The modal logic perspective. *Review of Symbolic Logic*, 2:469–487, 2009. (Cited on p. 20 and 168.)

Hamblin, Charles Leonard. The modal “probably”. *Mind*, 68:234–240, 1959. (Cited on p. 54.)

Hansen, Jens Ulrik. A logic-based approach to pluralistic ignorance. In De Vuyst, Jonas and Demey, Lorenz, editors, *Future Directions for Logic. Proceedings of PhDs in Logic III*, pages 67–80. College Publications, London, 2012. (Cited on p. 171.)

Hansen, Pierre and Jaumard, Brigitte. Probabilistic satisfiability. In Kohlas, Jürg and Moral, Serafín, editors, *Handbook of Defeasible Reasoning and Uncertainty Management Systems. Volume 5: Algorithms for Uncertainty and Defeasible Reasoning*, pages 321–367. Kluwer, Dordrecht, 2000. (Cited on p. 53.)

Harel, David, Kozen, Dexter, and Tiuryn, Jerzy. *Dynamic Logic*. MIT Press, Cambridge, MA, 2000. (Cited on p. 192, 195, 230, 267, and 270.)

Harremoës, Peter and Topsœ, Flemming. The quantitative theory of information. In Adriaans, Pieter and van Benthem, Johan, editors, *Philosophy of Information*, pages 171–216. Elsevier, Amsterdam, 2008. (Cited on p. 268.)

Hartmann, Stephan and Sprenger, Jan. Bayesian epistemology. In Bernecker, Sven and Pritchard, Duncan, editors, *Routledge Companion to Epistemology*, pages 609–620. Routledge, London, 2010. (Cited on p. 41.)

Hawthorne, James. Inductive logic. In Zalta, Edward N., editor, *Stanford Encyclopedia of Philosophy*. CSLI Publications, Stanford, CA, 2012. (Cited on p. 40.)

Heifetz, Aviad and Mongin, Philippe. Probability logic for type spaces. *Games and Economic Behavior*, 35:31–53, 2001. (Cited on p. 69.)

Hendricks, Vincent F. Knowledge transmissibility and pluralistic ignorance: A first stab. *Metaphilosophy*, 41:279–291, 2010. (Cited on p. 171.)

Herzig, Andreas and Longin, Dominique. On modal probability and belief. In Nielsen, Thomas D. and Zhang, Nevin L., editors, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty. Proceedings of ECSQARU 2003*, Lecture Notes in Computer Science 2711, pages 62–73. Springer, Berlin, 2003. (Cited on p. 55.)

Heylen, Jan. Modal-epistemic arithmetic and the problem of quantifying in. *Synthese*, 190:89–111, 2013. (Cited on p. 28.)

Hintikka, Jaakko. *Knowledge and Belief. An Introduction to the Logic of the Two Notions*. Cornell University Press, Ithaca, NY, 1962. (Cited on p. 17, 22, 73, 168, and 169.)

Hocutt, Max O. Is epistemic logic possible? *Notre Dame Journal of Formal Logic*, 13:433–453, 1972. (Cited on p. 30.)

Holliday, Wesley H. Dynamic testimonial logic. In He, Xiangdong, Horty, John, and Pacuit, Eric, editors, *Logic, Rationality, and Interaction. LORI 2009 Proceedings*, Lecture Notes in Computer Science 5834, pages 161–179. Springer, Berlin, 2009. (Cited on p. 171.)

Holliday, Wesley H. and Icard III, Thomas F. Moorean phenomena in epistemic logic. In Beklemishev, Lev, Goranko, Valentin, and Shehtman, Valentin, editors, *Advances in Modal Logic, Volume 8*, pages 178–199. College Publications, London, 2010. (Cited on p. 29 and 207.)

Hoover, Douglas N. Probability logic. *Annals of Mathematical Logic*, 14:287–313, 1978. (Cited on p. 60 and 62.)

Horn, Laurence R. *A Natural History of Negation*. University of Chicago Press, Chicago, IL, 1989. (Cited on p. 247.)

Horn, Laurence R. Hamburgers and truth: Why Gricean explanation is Gricean. In Hall, Kira, editor, *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society*, pages 454–471. Berkeley Linguistics Society, Berkeley, CA, 1990. (Cited on p. 251.)

Horn, Laurence R. Contradiction. In Zalta, Edward N., editor, *Stanford Encyclopedia of Philosophy*. CSLI Publications, Stanford, CA, 2010. (Cited on p. 261.)

Horn, Laurence R. Histoire d'*O: Lexical pragmatics and the geometry of opposition. In Béziau, Jean-Yves and Payette, Gillman, editors, *The Square of Opposition. A General Framework for Cognition*, pages 393–426. Peter Lang, Bern, 2012. (Cited on p. 247.)

Horsten, Leon. *Epistemic and Modal-Epistemic Arithmetic*. PhD thesis, KU Leuven, Leuven, 1993. (Cited on p. 28.)

Horsten, Leon. Modal-epistemic variants of Shapiro's system of Epistemic Arithmetic. *Notre Dame Journal of Formal Logic*, 35:284–291, 1994. (Cited on p. 28.)

Howson, Colin. Probability and logic. *Journal of Applied Logic*, 1:151–165, 2003. (Cited on p. 37.)

Howson, Colin. Logic with numbers. *Synthese*, 156:491–512, 2007. (Cited on p. 37.)

Howson, Colin. Can logic be combined with probability? Probably. *Journal of Applied Logic*, 7:177–187, 2009. (Cited on p. 37.)

Huber, Franz. Formal representations of belief. In Zalta, Edward N., editor, *Stanford Encyclopedia of Philosophy*. CSLI Publications, Stanford, CA, 2013. (Cited on p. 41.)

Huber, Franz and Schmidt-Petri, Christoph, editors. *Degrees of Belief*. Springer, Dordrecht, 2009. (Cited on p. 167.)

Hughes, George E. The modal logic of John Buridan. In Corsi, Giovanna, Mangione, Corrado, and Mugnai, Massimo, editors, *Atti del convegno internazionale di storia della logica: le teorie delle modalità*, pages 93–111. CLUEB, Bologna, 1987. (Cited on p. 247 and 281.)

Huitink, Janneke. Modal concord: a case study of Dutch. *Journal of Semantics*, 29:403–437, 2012. (Cited on p. 56.)

Jacoby, Paul. A triangle of opposites for types of propositions in Aristotelian logic. *New Scholasticism*, 24:32–56, 1950. (Cited on p. 250.)

Jacoby, Paul. Contrariety and the triangle of opposites in valid inferences. *New Scholasticism*, 34:141–169, 1960. (Cited on p. 250.)

Jacquette, Dale. Thinking outside the square of opposition box. In Béziau, Jean-Yves and Jacquette, Dale, editors, *Around and Beyond the Square of Opposition*, pages 73–92. Springer, Basel, 2012. (Cited on p. 252 and 281.)

Jaspers, Dany. Logic and colour. *Logica Universalis*, 6:227–248, 2012. (Cited on p. 251.)

Jech, Thomas. *Set Theory (Third Millenium Edition)*. Springer, Berlin, 2002. (Cited on p. 269.)

Jeffrey, Richard C. *The Logic of Decision (Second Edition)*. University of Chicago Press, Chicago, IL, 1983. (Cited on p. 98 and 171.)

Jeffrey, Richard C. *Probability and the Art of Judgment*. Cambridge University Press, Cambridge, 1992. (Cited on p. 37.)

Jonsson, Bengt, Yi, Wang, and Larsen, Kim G. Probabilistic extensions of process algebras. In Bergstra, Jan A., Ponse, Alban, and Smolka, Scott A., editors, *Handbook of Process Algebra*, pages 685–710. Elsevier, Amsterdam, 2001. (Cited on p. 81.)

Kajii, Atsushi and Morris, Stephen. Common p-belief: The general case. *Games and Economic Behavior*, 18:73–82, 1997. (Cited on p. 173.)

Kaneko, Mamuro. Epistemic logics and their game theoretic applications: Introduction. *Economic Theory*, 19:7–62, 2002. (Cited on p. 21.)

Karger, Elizabeth. John Buridan's theory of the logical relations between general modal formulae. In Braakhuis, Henk A. G. and Kneepkens, Corneille H., editors, *Aristotle's Peri Hermeneias in the Later Middle Ages*, pages 429–444. Ingenium, Groningen–Haren, 2003. (Cited on p. 281 and 282.)

Kauffman, Louis H. The mathematics of Charles Sanders Peirce. *Cybernetics & Human Knowing*, 8:79–110, 2001. (Cited on p. 31 and 264.)

Kavvadias, Dimitris and Papadimitriou, Christos H. A linear programming approach to reasoning about probabilities. *Annals of Mathematics and Artificial Intelligence*, 1:189–205, 1990. (Cited on p. 53.)

Keisler, Howard J. Probability quantifiers. In Barwise, Jon and Feferman, Solomon, editors, *Model-Theoretic Logics*, pages 509–556. Springer, New York, NY, 1985. (Cited on p. 60 and 62.)

Kennedy, Chris. Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, 30:1–45, 2007. (Cited on p. 56.)

Khomskii, Yurii. William of Sherwood, singular propositions and the hexagon of opposition. In Béziau, Jean-Yves and Payette, Gillman, editors, *The Square of Opposition. A General Framework for Cognition*, pages 43–60. Peter Lang, Bern, 2012. (Cited on p. 250.)

Knuuttila, Simo. *Modalities in medieval philosophy*. Routledge, London, 1993. (Cited on p. 17.)

Kooi, Barteld P. The Monty Hall dilemma. Master's thesis, Rijksuniversiteit Groningen, Groningen, 1999. (Cited on p. 87.)

Kooi, Barteld P. Probabilistic dynamic epistemic logic. *Journal of Logic, Language and Information*, 12:381–408, 2003. (Cited on p. 80, 87, 88, 92, 198, and 207.)

Koons, Robert. Defeasible reasoning. In Zalta, Edward N., editor, *Stanford Encyclopedia of Philosophy*. CSLI Publications, Stanford, CA, 2013. (Cited on p. 41.)

Korte, Tapio, Maunu, Ari, and Aho, Tuomo. Modal logic from Kant to possible world semantics. In Haaparanta, Leila, editor, *The Development of Modern Logic*, pages 516–550. Oxford University Press, Oxford, 2009. (Cited on p. 27.)

Kowalski, Robert. *Computational Logic and Human Thinking. How to be Artificially Intelligent*. Cambridge University Press, Cambridge, 2011. (Cited on p. 39.)

Kozen, Dexter and Parikh, Rohit. An elementary proof of the completeness of PDL. *Theoretical Computer Science*, 14:113–118, 1981. (Cited on p. 230.)

Kratzer, Angelika. Modality. In von Stechow, Arnim and Wunderlich, Dieter, editors, *Semantics: An International Handbook of Contemporary Research*, pages 639–650. de Gruyter, Berlin, 1991. (Cited on p. 55.)

Kyburg, Henry E. Probability, rationality, and the rule of detachment. In Bar-Hillel, Yehoshua, editor, *Proceedings of the 1964 International Congress for Logic, Methodology, and Philosophy of Science*, pages 301–310. North-Holland, Amsterdam, 1965. (Cited on p. 46.)

Kyburg, Henry E. Uncertainty logics. In Gabbay, Dov M., Hogger, Christopher J., and Robinson, John A., editors, *Handbook of Logic in Artificial Intelligence and Logic Programming*, pages 397–438. Oxford University Press, Oxford, 1994. (Cited on p. 40.)

Lackey, Jennifer and Sosa, Ernest, editors. *The Epistemology of Testimony*. Oxford University Press, Oxford, 2006. (Cited on p. 170.)

Lange, Marc. Is Jeffrey conditionalization defective in virtue of being noncommutative? Remarks on the sameness of sensory experience. *Synthese*, 123: 393–403, 2000. (Cited on p. 171.)

Larsen, Kim G. and Skou, Arne. Bisimulation through probabilistic testing. *Information and Computation*, 94:1–28, 1991. (Cited on p. 81.)

Leblanc, Hugues. Probabilistic semantics for first-order logic. *Zeitschrift für mathematische Logic und Grundlagen der Mathematik*, 25:497–509, 1979. (Cited on p. 44.)

Leblanc, Hugues. Alternatives to standard first-order semantics. In Gabbay, Dov M. and Guenthner, Franz, editors, *Handbook of Philosophical Logic, Volume 1*, pages 189–274. Reidel, Dordrecht, 1983. (Cited on p. 42, 43, and 44.)

Lenzen, Wolfgang. *Recent Work in Epistemic Logic*. North-Holland, Amsterdam, 1978. (Cited on p. 20 and 217.)

Lenzen, Wolfgang. *Glauben, Wissen und Wahrscheinlichkeit: Systeme der epistemischen Logik*. Springer, Berlin, 1980. (Cited on p. 20 and 217.)

Lenzen, Wolfgang. How to square knowledge and belief. In Béziau, Jean-Yves and Jacquette, Dale, editors, *Around and Beyond the Square of Opposition*, pages 305–311. Springer, Basel, 2012. (Cited on p. 217 and 247.)

Lewis, David. *Convention*. Harvard University Press, Cambridge, MA, 1969. (Cited on p. 133 and 168.)

Lewis, David. A subjectivist's guide to objective chance. In Jeffrey, Richard C., editor, *Studies in Inductive Logic and Probability, Volume 2*, pages 263–293. University of California Press, Berkeley, CA, 1980. (Cited on p. 74.)

Libert, Thierry. Hypercubes of duality. In Béziau, Jean-Yves and Jacquette, Dale, editors, *Around and Beyond the Square of Opposition*, pages 293–301. Springer, Basel, 2012. (Cited on p. 288.)

Liu, Fenrong. *Reasoning about Preference Dynamics*. Springer, Dordrecht, 2011. (Cited on p. 23.)

Löbner, Sebastian. German *schon – erst – noch*: An integrated analysis. *Linguistics and Philosophy*, 12:167–212, 1989. (Cited on p. 288.)

Löbner, Sebastian. *Wahr neben Falsch. Duale Operatoren als die Quantoren natürlicher Sprache*. Max Niemeyer Verlag, Tübingen, 1990. (Cited on p. 288.)

Löbner, Sebastian. *Understanding Semantics*. Hodder Arnold, London, 2002. (Cited on p. 267.)

Locke, John. *An Essay Concerning Human Understanding*. Clarendon Press, Oxford, 1975. (Cited on p. 167.)

Londey, David and Johanson, Carmen. Apuleius and the square of opposition. *Phronesis*, 29:165–173, 1984. (Cited on p. 247.)

Lorini, Emiliano. Agents with emotions: A logical perspective. *Association for Logic Programming Newsletter*, 21(2–3):1–9, 2008. (Cited on p. 192.)

Lorini, Emiliano and Castelfranchi, Cristiano. The unexpected aspects of surprise. *International Journal of Pattern Recognition and Artificial Intelligence*, 20:817–833, 2006. (Cited on p. 192.)

Lorini, Emiliano and Castelfranchi, Cristiano. The cognitive structure of surprise: Looking for basic principles. *Topoi*, 26:133–149, 2007. (Cited on p. 104 and 192.)

Lutz, Carsten. Complexity and succinctness of public announcement logic. In Nakashima, Hideyuki, Wellman, Michael P., Weiss, Gerhard, and Stone, Peter, editors, *AAMAS '06: Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 137–143. Association for Computing Machinery, Hakodate, 2006. (Cited on p. 243.)

Luzeaux, Dominique, Sallantin, Jean, and Dartnell, Christopher. Logical extensions of Aristotle's square. *Logica Universalis*, 2:167–187, 2008. (Cited on p. 264.)

Macedo, Luis and Cardoso, Amilcar. Creativity and surprise. In Wiggins, Geraint, editor, *Proceedings of the AISB '01 Symposium on Creativity in Arts and Science*, pages 84–92. The Society for the Study of Artificial Intelligence and Simulation Behaviour, York, 2001a. (Cited on p. 191.)

Macedo, Luis and Cardoso, Amilcar. Modelling forms of surprise in an artificial agent. In Moore, Johanna D. and Stenning, Keith, editors, *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, pages 588–593. Lawrence Erlbaum, Edinburgh, 2001b. (Cited on p. 191.)

Macedo, Luis and Cardoso, Amilcar. Exploration of unknown environments with motivational agents. In Jennings, Nicholas R., Sierra, Charles, Sonenberg, Liz, and Tambe, Milind, editors, *Proceedings of the Third International Joint*

Conference on Autonomous Agents and Multi-Agent Systems, pages 328–335. IEEE Computer Society, New York, NY, 2004. (Cited on p. 191.)

Macedo, Luis, Cardoso, Amilcar, and Reisenzein, Rainer. Modeling forms of surprise in artificial agents: Empirical and theoretical study of surprise functions. In Forbus, Kenneth, Gentner, Dedre, and Regier, Terry, editors, *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, pages 588–593. Lawrence Erlbaum, Mahwah, NJ, 2004. (Cited on p. 191.)

Macedo, Luis, Cardoso, Amilcar, and Reisenzein, Rainer. A surprise-based agent architecture. In Trappl, Robert, editor, *Proceedings of the 18th European Meeting on Cybernetics and Systems Research*, pages 583–588. Austrian Society for Cybernetic Studies, Vienna, 2006. (Cited on p. 191.)

Macedo, Luis, Cardoso, Amilcar, Reisenzein, Rainer, Lorini, Emiliano, and Castelfranchi, Cristiano. Artificial surprise. In Vallverdú, Jordi and Casacuberta, David, editors, *Handbook of Research on Synthetic Emotions and Sociable Robotics: New Applications in Affective Computing and AI*, pages 267–291. IGI Global, Hershey, PA, 2009. (Cited on p. 188 and 192.)

Macedo, Luis, Reisenzein, Rainer, and Cardoso, Amilcar. Surprise and anticipation in learning. In Seel, Norbert M., editor, *Encyclopedia of the Sciences of Learning*, pages 3250–3253. Springer, New York, NY, 2012. (Cited on p. 188.)

Marsella, Stacy C. and Gratch, Jonathan. EMA: A process model of appraisal dynamics. *Cognitive Systems Research*, 10:70–90, 2009. (Cited on p. 192.)

Martens, David B. A late medieval dispute about the conditions for knowledge. *Philosophical Papers*, 40:421–438, 2011. (Cited on p. 17.)

McNamara, Paul. Deontic logic. In Zalta, Edward N., editor, *Stanford Encyclopedia of Philosophy*. CSLI Publications, Stanford, CA, 2010. (Cited on p. 217 and 247.)

Meacham, Christopher J. G. Two mistakes regarding the principal principle. *British Journal for the Philosophy of Science*, 61:407–431, 2010. (Cited on p. 74.)

Mélès, Baptiste. No group of opposition for constructive logic: The intuitionistic and linear cases. In Béziau, Jean-Yves and Jacquette, Dale, editors, *Around and Beyond the Square of Opposition*, pages 201–217. Springer, Basel, 2012. (Cited on p. 247 and 288.)

Meyer, John-Jules and van der Hoek, Wiebe. *Epistemic Logic for AI and Computer Science*. Cambridge University Press, Cambridge, 1995. (Cited on p. 22.)

Meyer, Wulf-Uwe, Reisenzein, Rainer, and Schützwohl, Achim. Towards a process analysis of emotions: The case of surprise. *Motivation and Emotion*, 21: 251–274, 1997. (Cited on p. 189, 190, 191, and 192.)

Milgrom, Paul and Stokey, Nancy. Information, trade and common knowledge. *Journal of Economic Theory*, 26:1327–1347, 1982. (Cited on p. 133.)

Miller, David. A paradox of information. *British Journal for the Philosophy of Science*, 17:59–61, 1966. (Cited on p. 74.)

Miller, Scott A. Contradiction, surprise, and cognitive change: the effects of disconfirmation of belief on conservers and nonconservers. *Journal of Experimental Child Psychology*, 15:47–62, 1973. (Cited on p. 189.)

Monderer, Dov and Samet, Dov. Approximating common knowledge with common beliefs. *Games and Economic Behavior*, 1:170–190, 1989. (Cited on p. 165 and 173.)

Moretti, Alessio. *The Geometry of Logical Opposition*. PhD thesis, University of Neuchâtel, Neuchâtel, 2009a. (Cited on p. 217, 218, 232, 236, 239, 250, 254, 264, and 285.)

Moretti, Alessio. The geometry of standard deontic logic. *Logica Universalis*, 3:19–57, 2009b. (Cited on p. 247.)

Moretti, Alessio. Why the logical hexagon? *Logica Universalis*, 6:69–107, 2012a. (Cited on p. 217, 250, and 254.)

Moretti, Alessio. From the "logical square" to the "logical poly-simplexes". A quick survey of what happened in between. In Béziau, Jean-Yves and Payette, Gillman, editors, *The Square of Opposition. A General Framework for Cognition*, pages 119–156. Peter Lang, Bern, 2012b. (Cited on p. 254.)

Morgan, Charles G. There is a probabilistic semantics for every extension of classical sentence logic. *Journal of Philosophical Logic*, 11:431–442, 1982a. (Cited on p. 44.)

Morgan, Charles G. Simple probabilistic semantics for propositional K, T, B, S4, and S5. *Journal of Philosophical Logic*, 11:443–458, 1982b. (Cited on p. 44.)

Morgan, Charles G. Probabilistic semantics for propositional modal logics. In Leblanc, Hugues, Stern, Raphael, and Gumb, Raymond, editors, *Essays in Epistemology and Semantics*, pages 97–116. Haven Publications, New York, NY, 1983. (Cited on p. 44.)

Morgan, Charles G. and Leblanc, Hugues. Probabilistic semantics for intuitionistic logic. *Notre Dame Journal of Formal Logic*, 24:161–180, 1983. (Cited on p. 44.)

Nagel, Thomas. What is it like to be a bat? *Philosophical Review*, 83:435–450, 1974. (Cited on p. 189.)

Nilsson, Nils J. Probabilistic logic. *Artificial Intelligence*, 28:71–87, 1986. (Cited on p. 51 and 53.)

Nilsson, Nils J. Probabilistic logic revisited. *Artificial Intelligence*, 59:39–42, 1993. (Cited on p. 53.)

Oaksford, Mike and Chater, Nick. *The Probabilistic Mind: Prospects for Bayesian Cognitive Science*. Oxford University Press, Oxford, 2008. (Cited on p. 104.)

Oaksford, Mike and Chater, Nick. *Cognition and Conditionals. Probability and Logic in Human Thinking*. Oxford University Press, Oxford, 2010. (Cited on p. 39.)

Ortony, Andrew and Partridge, Derek. Surprisingness and expectation failure: What's the difference? In McDermott, John, editor, *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, pages 106–108. Morgan Kaufmann, Los Altos, CA, 1987. (Cited on p. 190, 191, and 192.)

Parikh, Rohit and Krasucki, Paul. Communication, consensus and knowledge. *Journal of Economic Theory*, 52:178–189, 1990. (Cited on p. 136, 164, and 165.)

Paris, Jeff B. *The Uncertain Reasoner's Companion*. Cambridge University Press, Cambridge, 1994. (Cited on p. 41.)

Parry, William T. and Hacker, Edward E. *Aristotelian Logic*. State University of New York Press, Albany, NY, 1991. (Cited on p. 232, 233, and 248.)

Parsons, Terrence. The traditional square of opposition. In Zalta, Edward N., editor, *Stanford Encyclopedia of Philosophy*. CSLI Publications, Stanford, CA, 2012. (Cited on p. 247 and 249.)

Pearl, Judea. Probabilistic semantics for nonmonotonic reasoning. In Cummins, Robert and Pollock, John, editors, *Philosophy and AI: Essays at the Interface*, pages 157–188. MIT Press, Cambridge, MA, 1991. (Cited on p. 44.)

Peckhaus, Volker. Algebra of logic, quantification theory, and the square of opposition. In Béziau, Jean-Yves and Payette, Gillman, editors, *The Square of Opposition. A General Framework for Cognition*, pages 25–41. Peter Lang, Bern, 2012. (Cited on p. 247.)

Peirce, Charles S. *Collected Papers. Volume 5: Pragmatism and pragmaticism*. Harvard University Press, Cambridge, MA, 1934. (Cited on p. 189.)

Peirce, Charles S. *Collected Papers. Volume 8: Reviews, correspondence, and bibliography*. Harvard University Press, Cambridge, MA, 1958. (Cited on p. 190.)

Pellissier, Régis. Setting n-opposition. *Logica Universalis*, 2:235–263, 2008. (Cited on p. 229.)

Perea, Andrés. *Epistemic Game Theory. Reasoning and Choice*. Cambridge University Press, Cambridge, 2012. (Cited on p. 22.)

Peterson, Martin. *An Introduction to Decision Theory*. Cambridge University Press, Cambridge, 2009. (Cited on p. 49.)

Pietarinen, Ahti-Veikko. What do epistemic logic and cognitive science have to do with each other? *Cognitive Systems Research*, 4:169–190, 2003. (Cited on p. 104.)

Plaza, Jan. Logics of public communications. In Emrich, Mary L., Pfeifer, M. S., Hadzikadic, Mirsad, and Ras, Zbigniew W., editors, *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*, pages 201–216 (reprinted in: *Synthese*, 158:165–179, 2007). Oak Ridge National Laboratory, Oak Ridge, TN, 1989. (Cited on p. 24, 83, and 107.)

Pritchard, Duncan. The epistemology of testimony. *Philosophical Issues*, 14: 326–348, 2004. (Cited on p. 170.)

Pucella, Riccardo. Knowledge and security. In van Ditmarsch, Hans P., Halpern, Joseph Y., van der Hoek, Wiebe, and Kooi, Barteld P., editors, *Handbook of Logics for Knowledge and Belief*. College Publications, London, forthcoming. (Cited on p. 20.)

Ramsey, Frank P. Truth and probability. In Mellor, David H., editor, *Philosophical Papers*, pages 52–94. Cambridge University Press, Cambridge, 1990. (Cited on p. 37.)

Read, Stephen. John Buridan's theory of consequence and his octagons of opposition. In Béziau, Jean-Yves and Jacquette, Dale, editors, *Around and Beyond the Square of Opposition*, pages 93–110. Springer, Basel, 2012a. (Cited on p. 247 and 281.)

Read, Stephen. Aristotle and Lukasiewicz on existential import. Available online at `www.st-andrews.ac.uk/~slr/Existential_Import.pdf`, 2012b. (Cited on p. 249.)

Reichenbach, Hans. *The Theory of Probability*. University of California Press, Berkeley, CA, 1949. (Cited on p. 43.)

Reisenzein, Rainer. The subjective experience of surprise. In Bless, Herbert and Forgas, Joseph P., editors, *The Message Within: The Role of Subjective Experience in Social Cognition and Behavior*, pages 262–279. Psychology Press, Philadelphia, PA, 2000. (Cited on p. 189.)

Reisenzein, Rainer and Meyer, Wulf-Uwe. Surprise. In Sander, David and Scherer, Klaus R., editors, *Oxford Companion to the Affective Sciences*, pages 386–387. Oxford University Press, Oxford, 2009. (Cited on p. 188.)

Reisenzein, Rainer, Meyer, Wulf-Uwe, and Schützwohl, Achim. Reacting to surprising events: A paradigm for emotion research. In Frijda, Nico, editor, *Proceedings of the 9th Conference of the International Society for Research on Emotions*, pages 292–296. ISRE, Toronto, 1996. (Cited on p. 188.)

Rini, Adriane A. and Cresswell, Max J. *The World-Time Parallel. Tense and Modality in Logic and Metaphysics*. Cambridge University Press, Cambridge, 2012. (Cited on p. 247.)

Roelofsen, Floris. Distributed knowledge. *Journal of Applied Non-Classical Logics*, 17:255–273, 2007. (Cited on p. 170.)

Romeijn, Jan-Willem. Statistics as inductive logic. In Bandyopadhyay, Prasanta S. and Forster, Malcolm, editors, *Handbook for the Philosophy of Science. Volume 7: Philosophy of Statistics*, pages 751–774. Elsevier, Amsterdam, 2011. (Cited on p. 40.)

Romeijn, Jan-Willem. Conditioning and interpretation shifts. *Studia Logica*, 100:583–606, 2012. (Cited on p. 94.)

Rosenhouse, Jason. *The Monty Hall Problem. The Remarkable Story of Math's Most Contentious Brain Teaser*. Oxford University Press, Oxford, 2009. (Cited on p. 87.)

Rotman, Joseph J. *An Introduction to the Theory of Groups (Fourth Edition)*. Springer, New York, NY, 1995. (Cited on p. 260.)

Rumelhart, David E. Schemata and the cognitive system. In Wyer Jr., Robert S. and Srull, Thomas K., editors, *Handbook of Social Cognition*, pages 161–188. Lawrence Erlbaum, Hillsdale, NJ, 1984. (Cited on p. 189.)

Sack, Joshua. Extending probabilistic dynamic epistemic logic. *Synthese*, 169: 241–257, 2009. (Cited on p. 76.)

Salerno, Joe, editor. *New Essays on the Knowability Paradox*. Oxford University Press, Oxford, 2009. (Cited on p. 29.)

Sanford, David H. Contraries and subcontraries. *Noûs*, 2:95–96, 1968. (Cited on p. 249 and 266.)

Sauriol, Pierre. Remarques sur la théorie de l'hexagone logique de Blanché. *Dialogue*, 7:374–390, 1968. (Cited on p. 250 and 264.)

Schang, Fabien. Oppositions and opposites. In Béziau, Jean-Yves and Jacquette, Dale, editors, *Around and Beyond the Square of Opposition*, pages 147–173. Springer, Basel, 2012a. (Cited on p. 263.)

Schang, Fabien. Questions and answers about oppositions. In Béziau, Jean-Yves and Payette, Gillman, editors, *The Square of Opposition. A General Framework for Cognition*, pages 289–320. Peter Lang, Bern, 2012b. (Cited on p. 254.)

Schank, R. *Explaining Patterns: Understanding Mechanically and Creatively*. Lawrence Erlbaum, Hillsdale, NJ, 1986. (Cited on p. 189.)

Scott, Dana. Advice on modal logic. In Lambert, Karel, editor, *Philosophical Problems in Logic*, pages 143–173. Reidel, Dordrecht, 1970. (Cited on p. 22.)

Segerberg, Krister. Qualitative probability in a modal setting. In Fenstad, Jens E., editor, *Proceedings of the Second Scandinavian Logic Symposium*, pages 341–352. North-Holland, Amsterdam, 1971. (Cited on p. 55.)

Sesmat, Auguste. *Logique II. Les Raisonnements. La Syllogistique*. Hermann, Paris, 1951. (Cited on p. 229 and 250.)

Seuren, Pieter. *The Logic of Language. Language from Within, Volume II*. Oxford University Press, Oxford, 2010. (Cited on p. 247, 250, 266, 281, and 282.)

Seuren, Pieter. From logical intuitions to natural logic. In Béziau, Jean-Yves and Payette, Gillman, editors, *The Square of Opposition. A General Framework for Cognition*, pages 231–288. Peter Lang, Bern, 2012a. (Cited on p. 247.)

Seuren, Pieter. Does a leaking O-corner save the square? In Béziau, Jean-Yves and Jacquette, Dale, editors, *Around and Beyond the Square of Opposition*, pages 129–138. Springer, Basel, 2012b. (Cited on p. 247 and 249.)

Shafer, Glenn. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976. (Cited on p. 53.)

Shapiro, Stewart. Epistemic and intuitionistic arithmetic. In Shapiro, Stewart, editor, *Intensional Mathematics*, pages 11–43. North-Holland, Amsterdam, 1985. (Cited on p. 28.)

Shoham, Yoav and Leyton-Brown, Kevin. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, Cambridge, 2009. (Cited on p. 188.)

Sim, Kwang Mong. Epistemic logic and logical omniscience: A survey. *International Journal of Intelligent Systems*, 12:57–81, 1997. (Cited on p. 30.)

Smessaert, Hans. *The Logical Geometry of Comparison and Quantification. A Cross-Categorial Analysis of Dutch Determiners and Aspectual Adverbs*. PhD thesis, KU Leuven, Leuven, 1993. (Cited on p. 217.)

Smessaert, Hans. On the 3D visualisation of logical relations. *Logica Universalis*, 3:303–332, 2009. (Cited on p. 217, 219, 222, 233, 236, 239, 240, 241, 249, 250, and 253.)

Smessaert, Hans. The classical Aristotelian hexagon versus the modern duality hexagon. *Logica Universalis*, 6:171–199, 2012a. (Cited on p. 288.)

Smessaert, Hans. Boolean differences between two hexagonal extensions of the logical square of oppositions. In Cox, Philip T., Plimmer, Beryl, and Rodgers, Peter, editors, *Diagrammatic Representation and Inference*, Lecture Notes in Computer Science 7352, pages 193–199. Springer, Berlin, 2012b. (Cited on p. 250 and 251.)

Smessaert, Hans and Demey, Lorenz. The logical geometry of the Aristotelian rhombic dodecahedron. Manuscript, 2013a. (Cited on p. 241, 251, 282, and 288.)

Smessaert, Hans and Demey, Lorenz. Logical geometries and information in the square of oppositions. *Submitted*, 2013b. (Cited on p. 7 and 222.)

Sokolov, Evgeny N., Spinks, John A., Näätänen, Risto, and Lyytinen, Heikki. *The Orienting Response in Information Processing*. Lawrence Erlbaum, Mahwah, NJ, 2002. (Cited on p. 189.)

Sorensen, Roy. Epistemic paradoxes. In Zalta, Edward N., editor, *Stanford Encyclopedia of Philosophy*. CSLI Publications, Stanford, CA, 2011. (Cited on p. 46.)

Spohn, Wolfgang. A survey of ranking theory. In Huber, Franz and Schmidt-Petri, Christoph, editors, *Degrees of Belief*, pages 185–228. Springer, Dordrecht, 2009. (Cited on p. 171.)

Stalnaker, Robert. On logics of knowledge and belief. *Philosophical Studies*, 128:169–199, 2006. (Cited on p. 110.)

Stevens, Stanley S. On the theory of scales of measurement. *Science*, 103:677—680, 1946. (Cited on p. 54.)

Stiensmeier-Pelster, Joachim, Martini, Alice, and Reisenzein, Rainer. The role of surprise in the attribution process. *Cognition and Emotion*, 9:5–31, 1995. (Cited on p. 189 and 190.)

Suppes, Patrick. Probabilistic inference and the concept of total evidence. In Hintikka, Jaakko and Suppes, Patrick, editors, *Aspects of Inductive Logic*, pages 49–65. Elsevier, Amsterdam, 1965. (Cited on p. 45.)

Szolovits, Peter and Pauker, Stephen G. Categorical and probabilistic reasoning in medical diagnosis. *Artificial Intelligence*, 11:115–144, 1978. (Cited on p. 54.)

Talbott, William. Bayesian epistemology. In Zalta, Edward N., editor, *Stanford Encyclopedia of Philosophy*. CSLI Publications, Stanford, CA, 2008. (Cited on p. 187.)

Tarski, Alfred. Wahrscheinlichkeitslehre und mehrwertige Logik. *Erkenntnis*, 5: 174–175, 1936. (Cited on p. 43.)

Uckelman, Sara. *Modalities in Medieval Logic*. PhD thesis, Institute for Logic, Language and Computation, Universiteit van Amsterdam, Amsterdam, 2011a. (Cited on p. 17.)

Uckelman, Sara. Deceit and indefeasible knowledge: The case of *dubitatio*. *Journal of Applied Non-Classical Logics*, 21:503–519, 2011b. (Cited on p. 27.)

Uckelman, Sara. A dynamic epistemic logic approach to modeling *Obligationes*. In Grossi, Davide, Minică, Ştefan, Rodenhäuser, Ben, and Smets, Sonja, editors, *Logic and Interactive Rationality Yearbook 2010*, pages 147–172. Institute for Logic, Language and Computation, Amsterdam, 2011c. (Cited on p. 27.)

Uckelman, Sara. Medieval *Disputationes de obligationibus* as formal dialogue systems. *Argumentation*, 27:143–166, 2013. (Cited on p. 27.)

van Benthem, Johan. *Modal Logic and Classical Logic*. Bibliopolis, Napoli, 1983. (Cited on p. 79.)

van Benthem, Johan. Linguistic universals in logical semantics. In Zaefferer, Dietmar, editor, *Semantic Universals and Universal Semantics*, Groningen-Amsterdam Studies in Semantics, Volume 12, pages 17–36. Foris, Berlin, 1991. (Cited on p. 288.)

van Benthem, Johan. *Exploring Logical Dynamics*. CSLI Publications, Stanford, CA, 1996. (Cited on p. 28, 133, and 136.)

van Benthem, Johan. Correspondence theory. In Gabbay, Dov M. and Guenthner, Franz, editors, *Handbook of Philosophical Logic, Volume 3 (Second Revised Edition)*, pages 325–408. Kluwer, Dordrecht, 2001a. (Cited on p. 79.)

van Benthem, Johan. Games in dynamic epistemic logic. *Bulletin of Economic Research*, 53:219–248, 2001b. (Cited on p. 104.)

van Benthem, Johan. Extensive games as process models. *Journal of Logic, Language and Information*, 11:289–313, 2002. (Cited on p. 125.)

van Benthem, Johan. Conditional probability meets update logic. *Journal of Logic, Language and Information*, 12:409–421, 2003. (Cited on p. 93.)

van Benthem, Johan. Logic and the dynamics of information. *Minds and Machines*, 13:503–519, 2003. (Cited on p. 23.)

van Benthem, Johan. What one may come to know. *Analysis*, 64:95–105, 2004. (Cited on p. 29.)

van Benthem, Johan. One is a lonely number: Logic and communication. In Chatzidakis, Zoe, Koepke, Peter, and Pohlers, Wolfrahm, editors, *Logic Colloquium '02*, Lecture Notes in Logic 27, pages 95–128. Association for Symbolic Logic & AK Peters, Wellesley, MA, 2006a. (Cited on p. 220.)

van Benthem, Johan. Open problems in logical dynamics. In Gabbay, Dov M., Goncharov, Sergei S., and Zakharyaschev, Michael, editors, *Mathematical Problems from Applied Logic I*, pages 137–192. Springer, Berlin, 2006b. (Cited on p. 220.)

van Benthem, Johan. Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics*, 17:129–155, 2007. (Cited on p. 103, 107, 108, 109, and 174.)

van Benthem, Johan. Rational dynamics and epistemic logic in games. *International Game Theory Review*, 9:13–45, 2007. (Cited on p. 104.)

van Benthem, Johan. Logic and reasoning: Do the facts matter? *Studia Logica*, 88:67–84, 2008. (Cited on p. 104.)

van Benthem, Johan. Actions that make us know. In Salerno, Joe, editor, *New Essays on the Knowability Paradox*, pages 129–146. Oxford University Press, Oxford, 2009. (Cited on p. 29.)

van Benthem, Johan. *Logical Dynamics of Information and Interaction*. Cambridge University Press, Cambridge, 2011. (Cited on p. 25, 28, 103, 109, 133, 188, and 218.)

van Benthem, Johan. A problem concerning qualitative probabilistic update. Manuscript, 2012. (Cited on p. 104.)

van Benthem, Johan and Liu, Fenrong. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logics*, 17:1577–182, 2007. (Cited on p. 23.)

van Benthem, Johan and Martinez, Maricarmen. The stories of logic and information. In Adriaans, Pieter and van Benthem, Johan, editors, *Philosophy of Information*, pages 217–280. Elsevier, Amsterdam, 2008. (Cited on p. 267.)

van Benthem, Johan and Minică, Ştefan. Toward a dynamic logic of questions. In He, Xiangdong, Horty, John, and Pacuit, Eric, editors, *Logic, Rationality, and Interaction. LORI 2009 Proceedings*, Lecture Notes in Computer Science 5834, pages 27–41. Springer, Berlin, 2009. (Cited on p. 138.)

van Benthem, Johan and Minică, Ştefan. Toward a dynamic logic of questions. *Journal of Philosophical Logic*, 41:644–669, 2012. (Cited on p. 138 and 141.)

van Benthem, Johan and Velázquez-Quesada, Fernando R. The dynamics of awareness. *Synthese*, 177 (supplement):5–27, 2010. (Cited on p. 30.)

van Benthem, Johan, van Eijck, Jan, and Kooi, Barteld P. Logics of communication and change. *Information and Computation*, 204:1620–1662, 2006. (Cited on p. 138 and 158.)

van Benthem, Johan, Gerbrandy, Jelle, and Kooi, Barteld P. Dynamic update with probabilities. *Studia Logica*, 93:67–96, 2009. (Cited on p. 95, 96, 98, 103, and 198.)

van Dalen, Dirk. *Logic and Structure (Fourth Edition)*. Springer, Berlin, 2004. (Cited on p. 225.)

van der Auwera, Johan. Modality: The three-layered scalar square. *Journal of Semantics*, 13:181–195, 1996. (Cited on p. 247.)

van der Hoek, Wiebe. Systems for knowledge and belief. *Journal of Logic and Computation*, 3:173–195, 1993. (Cited on p. 20 and 107.)

van der Hoek, Wiebe, van Linder, Bernd, and Meyer, John-Jules. Group knowledge is not always distributed (neither is it always implicit). *Mathematical Social Sciences*, 38:215–240, 1999. (Cited on p. 170.)

van der Meyden, Ron. Two applications of epistemic logic in computer security. In van Benthem, Johan, Gupta, Amitabha, and Parikh, Rohit, editors, *Proof, Computation and Agency – Logic at the Crossroads*, pages 133–144. Springer, Dordrecht, 2011. (Cited on p. 20.)

van Ditmarsch, Hans P. The Russian cards problem. *Studia Logica*, 75:31–62, 2003. (Cited on p. 20.)

van Ditmarsch, Hans P., van der Hoek, Wiebe, and Kooi, Barteld P. *Dynamic Epistemic Logic*. Springer, Dordrecht, 2007. (Cited on p. 20, 83, 91, 161, 176, 198, 207, 218, and 221.)

van Ditmarsch, Hans P., Ruan, Ji, and Verbrugge, Rineke. Sum and product in dynamic epistemic logic. *Journal of Logic and Computation*, 18:563–588, 2008. (Cited on p. 20.)

van Fraassen, Bas. Probabilistic semantics objectified: I. postulates and logics. *Journal of Philosophical Logic*, 10:371–391, 1981. (Cited on p. 44.)

van Fraassen, Bas. Gentlemen's wagers: Relevant logic and probability. *Philosophical Studies*, 43:47–61, 1983. (Cited on p. 44.)

van Fraassen, Bas. Belief and the will. *Journal of Philosophy*, 81:235–256, 1984. (Cited on p. 74.)

Velázquez-Quesada, Fernando R. Inference and update. *Synthese*, 169:283–300, 2009. (Cited on p. 30.)

Veloso, Sheila R. M., Veloso, Paulo A. S., and Veloso, Paula P. A tool for analysing logics. *Electronic Notes in Theoretical Computer Science*, 269:125–137, 2011. (Cited on p. 288.)

Vennekens, Joost, Denecker, Marc, and Bruynooghe, Maurice. CP-logic: A language of causal probabilistic events and its relation to logic programming. *Theory and Practice of Logic Programming*, 9:245–308, 2009. (Cited on p. 41.)

Verbrugge, Rineke, Gärdenfors, Peter, and Szymanik, Jakub. Cognition and logic. In Baltag, Alexandru and Smets, Sonja, editors, *Johan F. A. K. van Benthem on Logical and Informational Dynamics*. Springer, Dordrecht, forthcoming. (Cited on p. 104.)

von Wright, Georg H. *The Logic of Preference*. Edinburgh University Press, Edinburgh, 1963. (Cited on p. 23.)

Walley, Peter. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991. (Cited on p. 51.)

Westerståhl, Dag. Classical vs. modern squares of opposition, and beyond. In Béziau, Jean-Yves and Payette, Gillman, editors, *The Square of Opposition. A General Framework for Cognition*, pages 195–229. Peter Lang, Bern, 2012. (Cited on p. 288.)

Wilce, Alexander. Quantum logic and probability theory. In Zalta, Edward N., editor, *Stanford Encyclopedia of Philosophy*. CSLI Publications, Stanford, CA, 2012. (Cited on p. 41.)

Williams, Michael. *Problems of Knowledge. A Critical Introduction to Epistemology*. Oxford University Press, Oxford, 2001. (Cited on p. 167.)

Williamson, Colwyn. Squares of opposition: Comparisons between syllogistic and propositional logic. *Notre Dame Journal of Formal Logic*, 13:497–500, 1972. (Cited on p. 265.)

Williamson, Jon. Probability logic. In Gabbay, Dov M., Johnson, Ralph H., Ohlbach, Hans J., and Woods, John H., editors, *Handbook of the Logic of Argument and Inference: The Turn Toward the Practical*, pages 397–424. Elsevier, Amsterdam, 2002. (Cited on p. 42.)

Williamson, Timothy. *The Philosophy of Philosophy*. Blackwell, Oxford, 2007. (Cited on p. 28.)

Wittgenstein, Ludwig. *Tractatus Logico-Philosophicus*. Routledge and Kegan Paul, London, 1922. (Cited on p. 209.)

Wooldridge, Michael. *An Introduction to Multiagent Systems*. John Wiley & Sons, West Sussex, 2002. (Cited on p. 188 and 191.)

Yalcin, Seth. Probability operators. *Philosophy Compass*, 5:916–937, 2010. (Cited on p. 55.)

Zeijlstra, Hedde. Modal concord. In Friedman, Tova and Gibson, Masayuki, editors, *Proceedings of SALT 17*, pages 317–332. CLC Publications, Ithaca, NY, 2007. (Cited on p. 56.)

Zellweger, Shea. Untapped potential in Peirce's iconic notation for the sixteen binary connectives. In Houser, Nathan, Roberts, Don D., and Van Evra, James, editors, *Studies in the Logic of Charles Peirce*, pages 334–386. Indiana University Press, Bloomington, IN, 1997. (Cited on p. 264.)