Agreeing to Disagree in Probabilistic Dynamic Epistemic Logic

MSc Thesis (*Afstudeerscriptie*)

written by

Lorenz Demey
(born November 3rd, 1986 in Leuven, Belgium)

under the supervision of **Prof Dr Johan van Benthem** and **Prof Dr Dick de Jongh**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Abstract

Aumann's agreeing to disagree theorem is a central theorem of game theory. This result says that if two agents have a common prior, then they cannot agree (have common knowledge of their posteriors) to disagree (while these posteriors are not identical). This thesis looks at the agreeing to disagree theorem from the perspective of probabilistic dynamic epistemic logic.

The first goal of the thesis is to establish a new connection between game theory and epistemic logic. We prove (local model-based versions and global frame-based versions of) several semantic agreement theorems, and show that these are natural formalizations of Aumann's original result. We also provide axiomatizations of (dynamic) agreement logics, in which the first of these agreement theorems can be derived syntactically.

The second goal is the further technical development of probabilistic dynamic epistemic logic. We mention three examples. First, to model the experiment dynamics, we enrich the probabilistic Kripke models with 'experiment relations', thus establishing a link with the dynamic epistemic logic of questions. Second, to model the communication dynamics, we introduce the notion of a 'dialogue about a proposition φ', which is a particular sequence of public announcements; we show that this sequence always has a fixed point, and that at this fixed point the agents' probabilities for φ have become common knowledge. Thirdly, to make sure that both types of dynamics are well-defined, we introduce the constraint that $\mu_i(w)(w) > 0$ for all states w in any Kripke model, and discuss the technical and methodological consequences of this constraint.

The third goal is to use the technical results for the purpose of clarifying some conceptual issues surrounding the agreement theorem. In particular, we discuss the role of common knowledge (which we claim to be smaller than often thought), and the importance of explicitly representing the experimentation and communication dynamics, which is central in the intuitive motivation behind Aumann's result.

Recently Dégremont and Roy have formalized Aumann's agreement theorem in the context of epistemic plausibility models. Our fourth and final goal is to provide a detailed comparison between their approach and the one developed in this thesis, focusing on the representation of the agents' soft information (quantitatively versus qualitatively).

Acknowledgments

First of all, I wish to thank my supervisors Johan van Benthem and Dick de Jongh. It was Johan's recent work on dynamic probabilistic epistemic logic that got me interested in this topic in the first place. During our meetings and in his e-mails, he often made small remarks that took me hours to explore all the details and consequences of. He also left me a great deal of freedom to take my work in the directions I considered most interesting, and thus gave me the feeling that I was, and still am, developing into an autonomous logician, capable of independent research. It has been a great experience to be able to work together with Dick de Jongh. I will never forget the marathon session we spent solving a small detail in a completeness proof. During this afternoon, I learned more about *doing* logic than I ever could have in any regular class. I also wish to thank Dick for being a helpful academic mentor, and for allowing me to pursue the 'Logic and Mathematics' track of the master's program, even though my background was mainly in philosophy.

I wish to thank Eric Pacuit for his day-to-day supervision. Our weekly meetings helped me to clarify many conceptual issues, which I hope is reflected in this final version of the thesis. I also enjoyed our research project on neighborhood semantics for public announcement logic, and I look forward to continue working with him as the supervisor of my PhD project.

I wish to thank Cédric Dégremont for accepting to be on my thesis committee and to read the entire thesis. I also wish to thank him because it was his talk on agreement theorems in epistemic plausibility models (joint work with Olivier Roy) that formed the concrete inspiration for me to explore agreement theorems in a probabilistic setting, and that thus eventually led to this thesis.

I presented early versions of parts of this thesis at the Logic and Interactive Rationality (LIRA) Seminar in Amsterdam, and at the Formal Epistemology Project (FEP) Seminar in Leuven. I wish to thank the organizers of these seminars, Davide Grossi and Jake Chandler, respectively, for giving me the opportunity to present my work. I also wish to thank the audiences of both seminars for their suggestions and questions.

I wish to thank the Formal Epistemology Project and NUFFIC's Huygens Scholarship Programme, which funded the second year of my stay at the ILLC, and thus allowed me to do research and write this thesis without having to worry about financial matters.

I wish to thank my colleagues from Leuven: Jake Chandler, Richard Dietz,

Contents

Chapter 1

Introduction

This introductory chapter consists of three sections. In Section 1.1 we present the main research goals that will be addressed in the present thesis, and discuss their importance. In Section 1.2 we make some remarks about the broader context in which the thesis topic is to be situated. Finally, in Section 1.3 we provide a detailed overview of the thesis.

1.1   Research goals

The overarching aim of this thesis is to study Aumann's celebrated 'agreeing to disagree' theorem [2] from the perspective of epistemic logic, in particular *probabilistic dynamic epistemic logic* (PDEL). The agreement theorem (and the related no-trade theorem [29, 30]) 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 [8, 20, 22, 27, 29, 30, 32]. By bringing the agreement theorem to epistemic logic, this thesis thus establishes a new connection between the perspectives on knowledge (and related epistemic phenomena) that are provided by game theory on the one hand and epistemic logic on the other hand.

   This project also has definite advantages for both epistemic logic and game theory as separate disciplines. Probabilistic extensions of (dynamic) epistemic logic are a relatively recent development. Using (both the semantic and syntactic sides of) this framework to model the agreement theorem provides a clear illustration of its expressive strength. Furthermore, to adequately capture agreement results, further developments of this framework were necessary. Kooi [26] discusses the definition of the updated probability function $\mu_i^\varphi(w)$ (after a public announcement of φ) in the case that $\mu_i(w)(\llbracket\varphi\rrbracket^\mathbb{M}) = 0$ (and the updated probability function can thus not be defined by Bayesian conditionalization, because that would involve division by 0). In this thesis we propose a new solution to this issue, viz. the frame condition that $\mu_i(w)(w) > 0$ for all states w. We

will show that imposing this condition —in interaction with the usual truth precondition of public announcements— solves the problem, while not hurting the other properties of the framework (for example, it is preserved by all of the updates considered in this thesis, and it gives rise to a frame-correspondence result). Furthermore, we introduce (relativized) common knowledge, a feature which was left as future work in [26]. To capture the experiments mentioned in the intuitive explanation of the agreement theorem, we have extended the notion of probabilistic Kripke model, thus establishing a connection with the dynamic epistemic logic of questions [45]. Finally, we have modeled the communication between the agents as a series of public announcements that reaches a fixed point, which is also still relatively unexplored territory.

Bringing the agreement theorem to epistemic logic has advantages for game theory as well. Game theorists often use logic as a valuable tool for conceptual clarification: it provides a fine-grained analysis of some key concepts. This is certainly also the case for the agreement theorem. Based on our technical work, we will re-assess the role and importance of common knowledge, and —especially— of the underlying dynamics. We will even argue in Chapter 7 that the original formulation of the theorem is fundamentally flawed, and thus conceptually inferior to the new, logical formulation. The usefulness of logic as a clarificatory tool for game theorists was perfectly expressed by the game theorist Michael Bacharach, who we quote at length:

> Game theory is full of deep puzzles, and there is often disagreement about proposed solutions to them. The puzzlement and disagreement are neither empirical nor mathematical but, rather, concern the meanings of fundamental concepts [. . .] and the soundness of certain arguments [. . .] Logic appears to be an appropriate tool for game theory both because these conceptual obscurities involve notions such as reasoning, knowledge and counterfactuality which are part of the stock-in-trade of logic, and because it is a prime function of logic to establish the validity or invalidity of disputed arguments. [3, p. 17]

Aumann's original agreement theorem is a *static* result. The motivation behind this theorem, however, involves a lot of *dynamics*. This dynamics is mainly of two types: an experimentation phase and a communication phase. Much subsequent work on the agreement theorems has focused on the communication phase, while completely neglecting the experimentation phase. In this paper we will explicitly model both types of dynamics. This leads to a conceptually clearer formulation of the agreement theorem. Methodologically speaking, our formalization can be seen as an illustration of Johan van Benthem's program of 'dynamifying' logic. Our aim, however, is broader than providing a case study of this particular research program: we will argue that the dynamification of the agreement theorem is strictly *necessary* to obtain its most natural (non-convoluted) formulation.

Recently, Cédric Dégremont and Olivier Roy [13] have brought Aumann's agreement theorem (and some extensions) already to epistemic logic. They,

however, formalized it using epistemic plausibility models. Because these models do not have any probabilistic components, Dégremont and Roy's agreement theorem is a *qualitative* one. In this thesis, however, we will introduce probabilistic Kripke models [14, 26] (enriched with 'experiment relations', similar to the issue relations of the dynamic epistemic logic of questions [45]), which will facilitate the formulation of *quantitative* agreement theorems. Dégremont and Roy's work is very similar in spirit to ours; therefore the final aim of the thesis is to provide a detailed comparison between their approach and the approach developed here.

We conclude this section by summarizing the goals of the present thesis:

1. establishing a new link between epistemic logic and game theory

2. further technical development of probabilistic dynamic epistemic logic

3. conceptual clarification of the agreement theorem (in particular the role and importance of common knowledge and of the underlying dynamics)

4. comparison of our approach with that of Dégremont and Roy

1.2   Broader context

In this section, we make some remarks about the broader context of this thesis. First, we present the following three considerations:

First consideration. There exist several links between game theory and epistemic logic. Epistemic logic is typically used as a tool to investigate the epistemic foundations of various game-theoretical concepts (cf. our remarks and the quotation by Bacharach in the previous section). Typical examples include [40, 43]. Concrete game-theoretical topics that have been studied from the perspective of game theory include agreeing to disagree [13] and backward induction [7, 21, 38].[1]

Second consideration. Probability plays an important role in game theory. Probability is used to model the agents' uncertainty, and incoming information is processed by means of Bayesian conditionalization. Some of the main game-theoretical results are proved in a probabilistic setting, e.g. Aumann's agreeing to disagree theorem [2]. Furthermore, various game-theoretical notions are defined using probabilities; e.g. the notions of mixed strategy and expected payoff [33].

Third consideration. There has been much work recently on adding probabilities to epistemic logic. Much of this work was inspired by Fagin and Halpern's work on (static) epistemic logic [14]. Later, also probabilistic versions of dynamic epistemic logics were introduced by Kooi, van Benthem and Gerbrandy [26, 41, 44].

---

[1]For another perspective on the use of (modal) epistemic logic in game theory (considered either as an idealized descriptive or as a normative discipline), we refer the reader to [10].

When we now look back at the first consideration in light of the second and third ones, then it is striking that there has been almost no work on extending the connections between epistemic logic and game theory into probabilistic realms. To mention just one, particularly relevant, example: Dégremont and Roy [13] analyze Aumann's (probabilistic) agreement theorem in dynamic epistemic logic, but they turn it into a qualitative, rather than a probabilistic, result. Game theory certainly contains some important probabilistic notions (third consideration), and epistemic logic has the probabilistic tools available (second consideration) to analyze these notions.

From this broader perspective, the work presented in this thesis can be seen as a first example of how also the probabilistic aspects of game theory can be analyzed in (probabilistic) epistemic logic. Obviously, from this broader perspective much more work is needed in this area. For now, we mention just one concrete example: the logical analysis of mixed strategies might require the introduction of probabilistic Kripke models that do not (only) assign probabilities to *states*, but (also) to *relations* between states.[2]

1.3 Overview of the thesis

This thesis is organized as follows.

Chapter 2 provides an introduction to Aumann's original agreement theorem and highlights those features that will become particularly important in later chapters. We discuss both the formal expression and a more intuitive explanation/motivation of Aumann's result. This *formal/intuitive* dichotomy has a methodological analogue, viz. *static/dynamic*, which will be one of the red threads throughout the thesis. We conclude the chapter with some preliminary remarks about the research program of dynamifying logics.

In Chapter 3 we introduce the semantic setup of probabilistic dynamic epistemic logic. We define (enriched) probabilistic Kripke frames and models, and we introduce three ways of updating them: (1) carrying out experiments, (2) public announcement of a formula φ, and (3) a *dialogue* about a formula φ, i.e. a sequence of public annoucements that reaches a fixed point after finitely many steps. In parallel, we introduce the various (static and dynamic) languages that will be interpreted on these models, and we define their formal semantics.

Chapter 4 is the central chapter of this thesis. It brings together the material introduced in the previous two chapters, by formulating and proving several agreement theorems in the context of probabilistic Kripke models/frames. We prove a dynamic agreement theorem which makes only the experimentation dynamics explicit, and one which makes both the experimentation and the communication dynamics explicit. Both theorems have a local, model-based version and a global, frame-based version. Note that we do not have any *static* agreement theorems.

---

[2]Melvin Fitting makes a similar distinction in his work on many-valued modal logic. The first approach turns the *states* into many-valued entities; the second approach turns the *relations* between states into many-valued entities [15].

In Chapter 5 we move from the modeling perspective to the metatheoretical perspective. We first discuss a syntactic expressivity problem, and introduce two ways of solving it: incorporating notions from hybrid logic, and restricting to binary experiments. We argue that the second strategy is to be preferred on both technical and methodological grounds. Next, we establish characterization results for the conditions of the (frame-based) agreement theorems established in the previous chapter. These characterization results are then used as guidelines to construct several logical systems that are able to capture the semantic reasoning required to prove these agreement theorems. Finally, we turn to the metatheoretical properties of these various logics, and establish their soundness and completeness.

Chapter 6 provides a detailed comparison between the approach developed in this thesis and the one developed by Dégremont and Roy. We point out some important similarities between both approaches, but we also emphasize several technical and conceptual differences.

In Chapter 7 we turn to some more philosophical or methodological questions that were raised in earlier chapters. In particular, we focus on three topics. First of all, we discuss the perspective on agreeing to disagree via axiomatizations (i.e. the perspective developed in this thesis), and contrast it with another perspective, which aims to derive agreement results as theorems of weaker logical systems. Secondly, we discuss the importance of common knowledge for agreeing to disagree results. On the basis of some concrete technical results established in earlier chapters, we will argue that common knowledge is (at least conceptually speaking) not so central for agreeing to disagree results as is often thought. Our third, and most important, point is that the only 'natural' agreement theorems are all dynamic in nature, and that static theorems (such as Aumann's original one) are only possible at the expense of a convoluted semantic setup. The dynamification of the agreement theorem is thus more than merely an application or case study of a particular research program; rather, we will argue, it arises out of a general desire for conceptual clarity.

Chapter 8 provides an overview of all the results obtained in this thesis. The term 'result' is meant here to include both formal-logical theorems and philosophical/methodological theses. Furthermore, it collects and discusses some interesting open questions that arose throughout this thesis, and that will be the subject of future research.

Appendices A, B, C and D contain technical material (extra definitions, lemmas, etc.) that is used to establish the metatheoretical properties in Chapter 5. We have chosen to include this large amount of material in separate appendices, so that it would not interrupt the natural build-up of the main text.

Chapter 2

Analyzing the structure of the agreement theorem

In this chapter we introduce Aumann's original agreement theorem, and point out some of its features that will play a crucial role in our formalization. In Section 2.1 we discuss the formal expression of the theorem, and then, in Section 2.2, we discuss a more intuitive way of explaining or motivating it. Finally, in Section 2.3 we make some methodological remarks about the relation between the formal result on the one hand and its intuitive explanation on the other, and we connect this with van Benthem's general program of dynamifying logics.

2.1 Aumann's original agreement theorem

Aumann originally expressed the '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. [2, p. 1236]

The name by which this theorem is best known derives from another way of saying the same thing: 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).

A static and probabilistic theorem. It is clear that, when phrased in this way, the agreement theorem is a *static* result. We are not using the term 'static' in any technical sense here, but merely want to draw attention to the theorem's concrete formal structure: it is a conditional statement that can, as such, be written in a propositional (epistemic) language without any dynamic operators:

$$\big(\mathsf{equalpriors} \wedge C(\mathsf{posteriors})\big) \rightarrow \mathsf{equalposteriors}$$

So although this theorem describes what is/will be the case *if* there is common knowledge of the posteriors, it doesn't say anything about *how* this common

knowledge is to be achieved (i.e. how the conditional's antecedent is to be fulfilled). Furthermore, the theorem talks about 'priors' and 'posteriors', which are temporally relative terms: something is prior/posterior *relative to* some event. The theorem, however, does not mention any such event. We will return to these issues in the next section.

Finally we remark that the original agreement theorem is a *probabilistic* (quantitative) result. The models it talks about are probability spaces (that carry some additional structure). The priors and posteriors mentioned in the theorem are prior/posterior *probability measures*. Furthermore, these posteriors are calculated from the priors by means of Bayesian conditionalization, the most important probabilistic update rule.

2.2 The intuitive motivation behind the agreement theorem

We have just introduced the agreement theorem as it was originally formulated and proved by Aumann [2]. On the final page of his paper, Aumann motivates this theorem by sketching an informal scenario that embodies the intuitions behind it. (A similar explanatory scenario is described more extensively by Bonanno and Nehring [11, Section 2].) Roughly speaking, the scenario sketched by Aumann can be represented as follows:
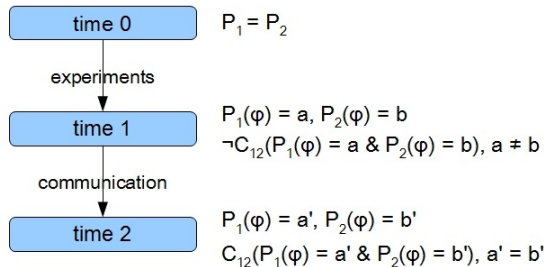
Intuitive scenario. We are considering two agents, 1 and 2. At the initial stage (time 0), they have the same probability distribution ($P_1 = P_2$). Then, both agents separate and perform an experiment: agent 1 performs experiment 1, and agent 2 performs experiment 2.[1] (In Aumann's scenario, these experiments involve coin-tossing.) Immediately after both agents have performed their experiments (time 1), their 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. 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 does not know how agent 2's probability

---

[1]The 'structure' of both experiments is supposed to be common knowledge between both agents. We will return to this point later.

function has changed. A symmetric argument applies to agent 2. We summarize this by saying that at time 1, 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. For example, agent 1 tells agent 2 that $P_1(\varphi) = a$. On the basis of this new information, agent 2 changes her probability function. At a certain point in the conversation, the agents obtain common knowledge of their probabilities. At this point (time 2) both agents' probability functions have changed again ($P_1(\varphi) = a'$ and $P_2(\varphi) = b'$), and these probabilities have become common knowledge between both agents. Since both agents had the same prior ($P_1 = P_2$ at time 0) and their posteriors have become common knowledge, Aumann's theorem now says that these probabilities have to coincide: $a' = b'$, i.e. $P_1(\varphi) = P_2(\varphi)$.[2]

Conclusion. This scenario clearly involves a lot of dynamic phenomena. Consider again the scheme above: it represents three points in time, and two dynamic processes: the experiments to get from time 0 to time 1, and the communication (or dialogue) to get from time 1 to time 2. The intuitive motivation behind the agreement theorem thus involves two broad types of dynamics: (1) the *experiments* and (2) the *communication*.

2.3   A first methodological remark

In the previous two sections, we have introduced Aumann's agreement theorem and the intuitive motivation behind it. Although the theorem itself is a static result, the intuitive explanation turned out to involve a lot of dynamics. This situation seems to be a good illustration of one of the key points in the general 'philosophy' of dynamic logic, which Johan van Benthem has formulated thus:

> the motivation for standard logics often contains procedural elements present in textbook presentations —and one can make this implicit dynamics explicit [39, p. 17]

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. Within the game theory literature, Geanakoplos and Polemarchakis [19] have focused on the communication part. Similarly, Dégremont and Roy [13] present not only a static version of the agreement theorem, but also two dynamic agreement theorems, in which the communication part is modeled explicitly (either as conditionalizing or as public announcements).

---

[2] Here is Aumann's own description of the final part of the scenario: "If the players inform each other of these posteriors, further revision may be called for. Our result implies that the process of exchanging information on the posteriors for [φ] will continue until these posteriors are equal." [2, p. 1238].

Still, both of these papers do not treat the experimentation phase: times 0 and 1 of the diagram above are still left 'entangled'.[3] In this thesis, however, we will explicitly formalize *both* types of dynamics. Although the communication is technically speaking the most interesting phenomenon (it will be modeled as a series of public announcements that reaches a fixed point), modeling the experiments will turn out to yield the biggest conceptual advantage. We will return to these methodological remarks in Chapter 7.

---

[3]Actually, to the best of our knowledge, the experimentation dynamics is not treated *anywhere* in the vast literature on agreeing to disagree.

Chapter 3

The general setup of PDEL

We will now introduce the general semantic setup of probabilistic dynamic epistemic logic. This setup will be used in Chapter 4 to formalize and prove various dynamic agreement theorems. In Section 3.1 we define (enriched) probabilistic Kripke models and the static epistemic language $\mathcal{L}$ that can be interpreted on such models. In Sections 3.2 and 3.3, we define two ways of updating these models, viz. carrying out experiments and performing (a series of) public announcements,[1] and we extend the language $\mathcal{L}$ with dynamic operators to talk about these model updates.

3.1   Probabilistic Kripke models

We first introduce (enriched) probabilistic Kripke frames and models. We focus on the two agent-case (this will suffice for the statement of the agreement theorems); generalizations to any (finite) number of agents are straightforward. We also fix a countably infinite set *Prop* of proposition letters.

Definition 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, and μ_1 and μ_2 assign to each world $w \in W$ a probability mass function $\mu_i(w) : W \to [0,1]$ (so $\sum_{v \in W} \mu_i(w)(v) = 1$). We also require (i) that $\mu_i(w)(w) > 0$ for all $w \in W$, and (ii) that $\mu_i(w)(v) = 0$ for $(w,v) \notin R_i$.

Definition 2. An *(enriched) probabilistic Kripke model* is a tuple $\mathbb{M} = \langle \mathbb{F}, V \rangle$, where $\mathbb{F} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2 \rangle$ is an (enriched) probabilistic Kripke frame and $V : Prop \to \wp(W)$ is a valuation.

The probabilistic Kripke models (and frames) defined above are called 'enriched' to distinguish them from the ones used in [14, 26]: our models contain

---

[1] It should be clear that these two ways of updating probabilistic Kripke models correspond to the two types of dynamics involved in the intuitive explanation of the agreement theorem, cf. Section 2.2.

the equivalence relations E_i (whose function will be clarified below), whereas the models used in [14, 26] do not. However, the models used in the remainder of this thesis are always the enriched ones defined above; therefore we will henceforth omit the extra qualifier and simply talk about 'probabilistic Kripke models'.

Discussion. We make some comments on the different components of these 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 w and v. Intuitively, we 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 our framework as the *issue relations* do in dynamic epistemic logics of questions [45].[2]

The probability mass function $\mu_i(w)$ represents agent i's subjective probabilities (at state w). 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 definition of $\mu_i(w)$ is lifted to any set $X \subseteq W$ by putting $\mu_i(w)(X) := \sum_{x \in X} \mu_i(w)(x)$.

Conditions (i) and (ii). We now make some comments on conditions (i) and (ii) of Definition 1. Condition (i) says that $\mu_i(w)(w) > 0$ for all $w \in W$. Intuitively, this 'liveness' condition requires that the agents do not assign probability 0 to their present world. At the end of this section (after the object language and its semantics have formally been introduced), we will see that this condition corresponds to the sentence $p \rightarrow P_i(p) > 0$, i.e. the agents assign non-zero probability to truths. This seems a fairly reasonable requirement for rational agents. The main reasons for including this condition are, however, of a more technical nature. In the next section we will be dealing with several ways of updating a probabilistic Kripke model, and all of these update rules say that the agents' probabilities change via some form of Bayesian conditionalization. This requires, however, that $\mu_i(w)(X) > 0$ for several sets $X \subseteq W$ (because otherwise we cannot divide by $\mu_i(w)(X)$, and the Bayesian conditionalization cannot take place). Condition (i) is an easy way to ensure we will always have $\mu_i(w)(X) > 0$ for all the relevant sets X. This will be discussed in more detail in the next section.[3] Finally, note that an assumption similar to our condition

---

[2]From a modeling perspective, this means that we choose not to model the experiments as 'events' that are independent of the (static) epistemic model, and that can be performed on/in that model. In other words, we do not represent the experiments as action models, which can then be 'multiplied' with an epistemic model to yield a new, updated epistemic model, but rather, we model the experiments as being part of the model itself.

[3]The case $\mu_i(w)(X) = 0$ in the update rules has been one of the problems that haunted this thesis for several months. Many other solutions to this issue (using $\mu_i(w)(X) > 0$ as a precondition, leaving the probability distribution unchanged, etc.) are unsatisfactory for technical as well as conceptual reasons. We think that the solution ultimately proposed (imposing condition (i) on all probabilistic Kripke frames) delivers a lot of good news for a relatively small price. Many thanks to Johan van Benthem for some stimulating discussions about this topic.

(i) is also made by Aumann when setting the stage for his original agreement theorem.[4]

Condition (ii) says that $\mu_i(w)(v) = 0$ for $(w, v) \notin R_i$. In words: at state w agent i has to assign probability 0 to all states v that she can epistemically distinguish from w (i.e. that she knows not to be the actual world).[5] At the end of this section, we will see that this condition corresponds to the sentence $K_i p \to P_i(p) = 1$, i.e. the agents assign probability 1 to all the propositions that they know. This seems to be a very reasonable demand for rational agents.[6] Technically speaking, condition (ii) leads to the following easy, but very useful lemma.[7]

Lemma 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 $\mathbb{M}$. For any set $X \subseteq W$ it holds that $\mu_i(w)(X \cap R_i[w]) = \mu_i(w)(X)$.

Proof. Condition (ii) that $\mu_i(w)(v) = 0$ for $(w, v) \notin R_i$ of Definition 1 is used at step $*$ in the following computation:

$$
\begin{aligned}
\mu_i(w)(X) &= \sum_{x \in X} \mu_i(w)(x) \\
&= \sum_{\substack{x \in X \\ x \in R_i[w]}} \mu_i(w)(x) \;+\; \sum_{\substack{x \in X \\ x \notin R_i[w]}} \mu_i(w)(x) \\
&\overset{*}{=} \sum_{\substack{x \in X \\ x \in R_i[w]}} \mu_i(w)(x) \;+\; \qquad 0 \\
&= \mu_i(w)(X \cap R_i[w])
\end{aligned}
$$

$\square$

The language. We now introduce the (static) language $\mathcal{L}$ by means of the following Backus-Naur form:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K_i\varphi \mid R_i\varphi \mid C^\varphi\varphi \mid X^\varphi\varphi \mid a_1 P_i(\varphi_1) + \cdots + a_n P_i(\varphi_n) \geq k$$

(where $i \in \{1, 2\}$, $1 \leq n < \omega$ and $a_1, \ldots, a_n, k \in \mathbb{Q}$). We only allow rational numbers as values for $a_1, \ldots, a_n, k$ in order to keep the language countable.[8] This issue will return in the proofs of the correspondence results in Section 5.2.

---

[4] Literally: "$\mathcal{P}_1$ and $\mathcal{P}_2$ [are] partitions of Ω whose join $\mathcal{P}_1 \vee \mathcal{P}_2$ *consists of nonnull events*" [2, p. 1236, my emphasis].

[5] Note that conditions (i) and (ii) are 'compatible' with each other. Condition (i) requires that $\mu_i(w)(w) > 0$ and condition (ii) requires that $\mu_i(w)(v) = 0$ if $(w, v) \notin R_i$. Since R_i is an equivalence relation, we always have $(w, w) \in R_i$, so condition (ii) can never require that $\mu_i(w)(w) = 0$, which would contradict condition (i).

[6] Fagin and Halpern [14] even call this the consistency requirement.

[7] For any binary relation $\mathcal{R} \subseteq W \times W$, we abbreviate $\mathcal{R}[w] := \{v \in W \mid (w, v) \in \mathcal{R}\}$. Furthermore, we will write $\mathcal{R}^*$ for the reflexive transitive closure of $\mathcal{R}$ and $\mathcal{R}^+$ for the transitive closure of $\mathcal{R}$.

[8] Also [14, 26] only allow rational numbers in formulas. However, none of them give any motivation for this choice; presumably at least one of their reasons is the one mentioned in the main text.

Intuitive interpretation. 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 [46].

Knowledge and relativized common knowledge also have 'post-experimental' counterparts: $R_i\varphi$ and $X^\varphi\psi$.[9] 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 section.

Common abbreviations. Ordinary common knowledge can be defined as $C\varphi := C^\top\varphi$ (here, and in the remainder of this thesis, $\top$ is any propositional tautology). Analogously, ordinary post-experimental common knowledge can be defined in terms of relativized post-experimental common knowledge by putting $X\varphi := X^\top\varphi$. We define 'general knowledge' by putting $E\varphi := K_1\varphi \wedge K_2\varphi$ ('everybody knows that φ'), and 'general post-experimental knowledge' by putting $F\varphi := R_1\varphi \wedge R_2\varphi$ ('after the experiments, everybody knows that φ was the case before the experiments').

i-probability formulas. Formulas of the form $a_1 P_i(\varphi_1) + \cdots + a_n P_i(\varphi_n) \geq k$ will be called *i-probability formulas*. Note that we do not allow mixed agent indices in such formulas; e.g. $P_1(\varphi) + P_2(\psi) \geq k$ is *not* a well-formed formula. Intuitively, $P_i(\varphi) \geq k$ says that agent i assigns probability at least k to φ. There are two reasons for allowing summation and multiplication by rationals: (i) this extra expressivity is useful when establishing completeness results, and (ii) more importantly, it allows us to express comparative judgments such as 'agent i thinks that φ is 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:

$$\sum_{\ell=1}^n a_\ell P_i(\varphi_\ell) \geq k \qquad \text{for} \quad a_1 P_i(\varphi_1) + \cdots + a_n P_i(\varphi_n) \geq k$$

$$a_1 P_i(\varphi_1) \geq a_2 P_i(\varphi_2) \quad \text{for} \quad a_1 P_i(\varphi_1) + (-a_2)P_i(\varphi_2) \geq 0$$

---

[9]Hence we have 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 $\mathcal{L}$. Our main reason for not using another letter for the post-experimental knowledge operator is to ensure uniformity of notation with [45]. We trust that the meaning of R_i will always be clear from the context.

$$\sum_{\ell=1}^n a_\ell P_i(\varphi_\ell) \leq k \qquad \text{for} \quad \sum_{\ell=1}^n -a_\ell P_i(\varphi_\ell) \geq -k$$

$$\sum_{\ell=1}^n a_\ell P_i(\varphi_\ell) < k \qquad \text{for} \quad \neg\left(\sum_{\ell=1}^n a_\ell P_i(\varphi_\ell) \geq k \right)$$

$$\sum_{\ell=1}^n a_\ell P_i(\varphi_\ell) > k \qquad \text{for} \quad \neg\left(\sum_{\ell=1}^n a_\ell P_i(\varphi_\ell) \leq k \right)$$

$$\sum_{\ell=1}^n a_\ell P_i(\varphi_\ell) = k \qquad \text{for} \quad \left(\sum_{\ell=1}^n a_\ell P_i(\varphi_\ell) \geq k \right) \wedge \left(\sum_{\ell=1}^n a_\ell P_i(\varphi_\ell) \leq k \right)$$

$$a_1 = a_2 \qquad \text{for} \quad a_1 P_1(\top) + (-a_2)P_1(\top) = 0$$

There is an obvious difference in generality between atomic proposition letters p and arbitrary formulas φ. A similar distinction can be made on the level of i-probability formulas (which will be used in Section 5.2):

Definition 4. Consider an i-probability formula $\varphi := \sum_{\ell=1}^n a_\ell P_i(\varphi_\ell) \geq k$. We will say that φ is *atomic* iff all φ_ℓ are propositional atoms. In other words, φ is an atomic i-probability formula iff it is of the form $\sum_{\ell=1}^n a_\ell P_i(p_\ell) \geq k$.

Formal semantics. Consider an arbitrary 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 thesis, we will often abbreviate $R := R_1 \cup R_2$, $R^e := (R_1 \cap E_1) \cup (R_2 \cap E_2)$, and $[\![\varphi]\!]^{\mathbb{M}} := \{v \in W \mid \mathbb{M}, v \models \varphi\}$. The semantics of $\mathcal{L}$ is inductively defined as follows:

$$
\begin{array}{lll}
\mathbb{M}, w \models p & \text{iff} & w \in V(p) \\
\mathbb{M}, w \models \neg\varphi & \text{iff} & \mathbb{M}, w \not\models \varphi \\
\mathbb{M}, w \models \varphi \wedge \psi & \text{iff} & \mathbb{M}, w \models \varphi \text{ and } \mathbb{M}, w \models \psi \\
\mathbb{M}, w \models K_i\varphi & \text{iff} & \forall v \in W : (w,v) \in R_i \Rightarrow \mathbb{M}, v \models \varphi \\
\mathbb{M}, w \models C^\varphi\psi & \text{iff} & \forall v \in W : (w,v) \in \left(R \cap (W \times [\![\varphi]\!]^{\mathbb{M}})\right)^+ \Rightarrow \mathbb{M}, v \models \psi \\
\mathbb{M}, w \models R_i\varphi & \text{iff} & \forall v \in W : (w,v) \in R_i \cap E_i \Rightarrow \mathbb{M}, v \models \varphi \\
\mathbb{M}, w \models X^\varphi\psi & \text{iff} & \forall v \in W : (w,v) \in \left(R^e \cap (W \times [\![\varphi]\!]^{\mathbb{M}})\right)^+ \Rightarrow \mathbb{M}, v \models \varphi \\
\mathbb{M}, w \models \sum_{\ell=1}^n a_\ell P_i(\varphi_\ell) \geq k & \text{iff} & \sum_{\ell=1}^n a_\ell \mu_i(w)([\![\varphi_\ell]\!]^{\mathbb{M}}) \geq k
\end{array}
$$

Combining the syntactic definitions of the ordinary common knowledge operators that we introduced above with the definition of the semantics, we get the expected semantics for these operators:

$$
\begin{array}{lll}
\mathbb{M}, w \models C\varphi & \text{iff} & \forall v \in W : (w,v) \in R^* \Rightarrow \mathbb{M}, v \models \varphi \\
\mathbb{M}, w \models X\varphi & \text{iff} & \forall v \in W : (w,v) \in (R^e)^* \Rightarrow \mathbb{M}, v \models \varphi
\end{array}
$$

We have defined $\mathbb{M}, w \models \varphi$ for all formulas $\varphi \in \mathcal{L}$. As usual, we write $\mathbb{M} \models \varphi$ iff $\mathbb{M}, w \models \varphi$ for all $w \in W$. For a probabilistic Kripke frame $\mathbb{F}$ (with domain W), we write $\mathbb{F} \models \varphi$ iff $\langle \mathbb{F}, V \rangle \models \varphi$ for all valuations $V : Prop \to \wp(W)$. Finally, if $\mathcal{C}$ is a class of probabilistic Kripke frames, then we write $\mathcal{C} \models \varphi$ iff $\mathbb{F} \models \varphi$ for

every frame $\mathbb{F} \in \mathcal{C}$.

Characterization results for conditions (i) and (ii). As promised earlier in this section, we wil now provide frame correspondence results for conditions (i) and (ii) of Definition 1.

Lemma 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:

$$\text{for all } w \in W : \mu_i(w)(w) > 0 \quad \text{iff} \quad \mathbb{F} \models p \to P_i(p) > 0$$

Proof. We first prove the $\Rightarrow$-direction. Let $V \colon Prop \to \wp(W)$ be an arbitrary valuation and $w \in W$ an arbitrary state. Suppose that $\langle \mathbb{F}, V \rangle, w \models p$. Hence $w \in \llbracket p \rrbracket^{\langle \mathbb{F}, V \rangle}$, and thus $\mu_i(w)(\llbracket p \rrbracket^{\langle \mathbb{F}, V \rangle}) \geq \mu_i(w)(w) > 0$. Hence it follows that $\mathbb{M}, w \models P_i(p) > 0$.

  We now prove the $\Leftarrow$-direction by contraposition. Suppose that there exists a state $w \in W$ such that $\mu_i(w)(w) = 0$. Define a valuation V by putting $V(p) := \{w\}$. It now follows that $\langle \mathbb{F}, V \rangle, w \models p$ and that $\langle \mathbb{F}, V \rangle, w \not\models P_i(p) > 0$. $\square$

Lemma 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 we have:

$$\text{for all } w, v \in W : \text{if } (w, v) \notin R_i \text{ then } \mu_i(w)(v) = 0 \quad \text{iff} \quad \mathbb{F} \models K_i p \to P_i(p) = 1$$

Proof. We first prove the $\Rightarrow$-direction. Let $V \colon Prop \to \wp(W)$ be an arbitrary valuation and $w \in W$ an arbitrary state. Suppose that $\langle \mathbb{F}, V \rangle, w \models K_i p$. Hence $R_i[w] \subseteq \llbracket p \rrbracket^{\langle \mathbb{F}, V \rangle}$, and thus $\llbracket p \rrbracket^{\langle \mathbb{F}, V \rangle} \cap R_i[w] = R_i[w] = W \cap R_i[w]$. Applying Lemma 3 at the $*$-labeled steps, we get that $\mu_i(w)(\llbracket p \rrbracket^{\langle \mathbb{F}, V \rangle}) \overset{*}{=} \mu_i(w)(\llbracket p \rrbracket^{\langle \mathbb{F}, V \rangle} \cap R_i[w]) = \mu_i(w)(W \cap R_i[w]) \overset{*}{=} \mu_i(w)(W) = 1$, and hence $\langle \mathbb{F}, V \rangle, w \models P_i(p) = 1$.

  We now prove the $\Leftarrow$-direction by contraposition. Suppose that there exist states $w, v \in W$ such that $(w, v) \notin R_i$ and yet $\mu_i(w)(v) > 0$. Define a valuation V by putting $V(p) := R_i[w]$. It now easily follows that $\langle \mathbb{F}, V \rangle, w \models K_i p$. Note that $\mu_i(w)(R_i[w]) = 1 - \mu_i(w)(W - R_i[w])$. Since $v \in W - R_i[w]$ and $\mu_i(w)(v) > 0$, it follows that $\mu_i(w)(W - R_i[w]) > 0$, and thus $\mu_i(w)(R_i[w]) < 1$. Hence $\langle \mathbb{F}, V \rangle, w \not\models P_i(p) = 1$. $\square$

3.2 Dynamics: the experimentation phase

We will now model the first type of dynamics described in Section 2.2, viz. carrying out the experiments. Syntactically, this involves adding a dynamic operator to the language $\mathcal{L}$; semantically, we will define a particular way of updating probabilistic Kripke models and argue that this formal definition correctly captures the intuitive idea of carrying out an experiment.

  Syntactically, we add a new dynamic operator [EXP] to the language $\mathcal{L}$, thus obtaining the language $\mathcal{L}([EXP])$, which has the following Backus-Naur form:

$$\varphi \quad ::= \quad p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K_i\varphi \mid R_i\varphi \mid C^\varphi\varphi \mid X^\varphi\varphi \mid$$
$$a_1 P_i(\varphi_1) + \cdots + a_n P_i(\varphi_n) \geq k \mid [\text{EXP}]\,\varphi$$

The [EXP]-operator says that both agents perform their experiments; hence, $[\text{EXP}]\,\varphi$ 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 defined immediately afterwards.

$$\mathbb{M}, w \models [\text{EXP}]\,\varphi \quad \text{iff} \quad \mathbb{M}^e, w \models \varphi$$

Definition 7. 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:

- $W^e := W$

- $R_i^e := R_i \cap E_i$

- $E_i^e := E_i$

- for all $w \in W^e$, put $\mu_i^e(w) \colon W^e \to [0,1] \colon v \mapsto \mu_i^e(w)(v) := \frac{\mu_i(w)(\{v\} \cap E_i[w])}{\mu_i(w)(E_i[w])}$

- for all $p \in Prop$, put $V^e(p) := V(p)$

Recall that we abbreviated $R^e = (R_1 \cap E_1) \cup (R_2 \cap E_2)$ in the previous section. Applying Definition 7, this can now be rewritten as $R^e = R_1^e \cup R_2^e$, which is structurally analogous to our other abbreviation: $R = R_1 \cup R_2$.

Discussion. We will now justify our definition of the model update operation $\mathbb{M} \mapsto \mathbb{M}^e$ by showing that it nicely captures the intuitive idea of carrying out an experiment. 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. Symmetric remarks hold for experiment 2. Hence, after the experiments, agent i cannot distinguish between states w and v iff (i) she could not distinguish between them *before* the experiments, i.e. $(w, v) \in R_i$, *and* (ii) agent i's experiment does not differentiate between w and v, i.e. $(w, v) \in E_i$.[10] This closely resembles Bonanno and Nehring's [11] description of the experiments as imposing a partition on the model.

We 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

---

[10]We already discussed the analogy between carrying out an experiment and asking a question. Our modeling of the experiments as intersecting R_i with E_i is analogous to the 'resolve' action in the dynamic epistemic logic of questions (cf. [45, Definition 6]): carrying out an experiment means getting an answer to a question posed to nature.

her experiment. This captures the idea that the agents process new information by means of "Bayesian updating (which they agree to be the correct way to update probabilities)" [11, p. 4]. One might worry whether this division is always defined: what if $\mu_i(w)(E_i[w]) = 0$? It is at this point, however, that the power of condition (i) in probabilistic Kripke models/frames becomes clear. Recall that this condition says that $\mu_i(w)(w) > 0$ (for all $w \in W$). Since E_i is reflexive, we get that $w \in E_i[w]$, and hence $\mu_i(w)(E_i[w]) \geq \mu_i(w)(w) > 0$. We thus do not have to worry about $\mu_i(w)(E_i[w]) = 0$.

Finally, since carrying out the experiments does not change the objective facts, the valuation function is left untouched.

Definition 7 matches well with our intuitive idea of what experiments are and how they influence the agents' knowledge and probabilities. To illustrate this, we now discuss in detail an intuitive scenario, and how this can be formalized in the setup introduced thus far.

Example 8. Consider the following scenario. Agent 1 does not 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 $W := \{w, v\}$, $R_1 := W \times W$, $E_1 = \{(w, w), (v, v)\}$, $\mu_1(w)(w) = \mu_1(w)(v) = \frac{1}{2}$, and $V(p) = \{w\}$ (we do not care about the definitions of $\mu_1(v), R_2, E_2$ and μ_2). It is easy to see that this model is a faithful representation of the above scenario. Consider, for example:

$$\mathbb{M}, w \models \ \neg K_1 p \wedge \neg K_1 \neg p$$
$$\mathbb{M}, w \models \ P_1(p) = \tfrac{1}{2} \wedge P_1(\neg p) = \tfrac{1}{2}$$

Now suppose that the agents carry out their experiments, i.e. consider the updated model $\mathbb{M}^e$. Applying Definition 7, we see that $R_1^e = \{(w, w), (v, v)\}$ and that

$$\mu_1^e(w)(w) = \frac{\mu_1(w)(\{w\} \cap E_1[w])}{\mu_1(w)(E_1[w])} = \frac{\mu_1(w)(w)}{\mu_1(w)(w)} = 1$$
$$\mu_1^e(w)(v) = \frac{\mu_1(w)(\{v\} \cap E_1[w])}{\mu_1(w)(E_1[w])} = \frac{\mu_1(w)(\emptyset)}{\mu_1(w)(w)} = 0$$

Hence:

$$\mathbb{M}, w \models \ [\text{EXP}] \, K_1 p$$
$$\mathbb{M}, w \models \ [\text{EXP}](P_1(p) = 1 \wedge P_1(\neg p) = 0)$$

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 7 seems to be a natural way of representing the experimentation dynamics: it makes the intuitively right 'predictions' about the agents' knowledge and probabilities.

Some useful lemmas. Since probabilistic Kripke models are the most general semantic notion of this thesis (they do not form a subclass of some more general class of models), the class of probabilistic Kripke models should be closed under all the model update operations.[11] The following lemma shows that this is the case for the model update operation $\mathbb{M} \mapsto \mathbb{M}^e$.

Lemma 9. If $\mathbb{M} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V \rangle$ is a probabilistic Kripke model, then $\mathbb{M}^e$ is a probabilistic Kripke model as well.

Proof. First of all, note that $W^e = W$ is finite. Since both R_i and E_i are equivalence relations on W, it is easily verified that also $R_i^e = R_i \cap E_i$ and $E_i^e = E_i$ are equivalence relations on $W^e = W$.

We now check that $\mu_i^e(w)$ is a probability mass function on W^e (for all $w \in W^e$). Consider $w \in W^e$ arbitrary. Note that $\mu_i^e(w)(v) = \frac{\mu_i(w)(\{v\} \cap E_i[w])}{\mu_i(w)(E_i[w])} \in [0,1]$ for any $v \in W^e = W$, and furthermore, we compute:

$$
\begin{aligned}
\sum_{v \in W^e} \mu_i^e(w)(v) &= \sum_{v \in W} \frac{\mu_i(w)(\{v\} \cap E_i[w])}{\mu_i(w)(E_i[w])} \\
&= \sum_{\substack{v \in W \\ v \in E_i[w]}} \frac{\mu_i(w)(\{v\} \cap E_i[w])}{\mu_i(w)(E_i[w])} + \sum_{\substack{v \in W \\ v \notin E_i[w]}} \frac{\mu_i(w)(\{v\} \cap E_i[w])}{\mu_i(w)(E_i[w])} \\
&= \sum_{\substack{v \in W \\ v \in E_i[w]}} \frac{\mu_i(w)(v)}{\mu_i(w)(E_i[w])} + \sum_{\substack{v \in W \\ v \notin E_i[w]}} \frac{\mu_i(w)(\emptyset)}{\mu_i(w)(E_i[w])} \\
&= \frac{\mu_i(w)(E_i[w])}{\mu_i(w)(E_i[w])} + 0 \\
&= 1
\end{aligned}
$$

Finally, we check that $\mathbb{M}^e$ still satisfies conditions (i) and (ii) of Definition 1. Note that for any $w \in W^e$, it holds that $\mu_i(w)(w) > 0$ (by condition (i) for the model $\mathbb{M}$), so

$$
\mu_i^e(w)(w) = \frac{\mu_i(w)(\{w\} \cap E_i[w])}{\mu_i(w)(E_i[w])} = \frac{\mu_i(w)(w)}{\mu_i(w)(E_i[w])} > 0
$$

and thus condition (i) is fulfilled. Furthermore, note that if $(w,v) \notin R_i^e$, then $(w,v) \notin R_i$ or $(w,v) \notin E_i$. If $(w,v) \notin R_i$, then $\mu_i(w)(v) = 0$ (by condition (ii)

---

[11] One motivation for this desideratum is the following. The semantics $\mathbb{M}, w \models \varphi$ is defined only for probabilistic Kripke models $\mathbb{M}$. If the class of probabilistic Kripke models were not closed under, for example, $\mathbb{M} \mapsto \mathbb{M}^e$, then a formula such as $[\text{EXP}] \varphi$ would be uninterpretable: $\mathbb{M}, w \models [\text{EXP}] \varphi$ iff $\mathbb{M}^e, w \models \varphi$, but the latter is undefined because $\mathbb{M}^e$ is itself not a probabilistic Kripke model...

for the model $\mathbb{M}$) and thus

$$\mu_i^e(w)(v) = \frac{\mu_i(w)(\{v\} \cap E_i[w])}{\mu_i(w)(E_i[w])} \leq \frac{\mu_i(w)(v)}{\mu_i(w)(E_i[w])} = 0$$

On the other hand, if $(w,v) \notin E_i$, then $\{v\} \cap E_i[w] = \emptyset$, and thus

$$\mu_i^e(w)(v) = \frac{\mu_i(w)(\{v\} \cap E_i[w])}{\mu_i(w)(E_i[w])} = \frac{\mu_i(w)(\emptyset)}{\mu_i(w)(E_i[w])} = 0$$

This shows that condition (ii) is fulfilled. $\qquad\square$

Just like $\mu_i(w)$, also $\mu_i^e(w)$ can be lifted to the level of sets, by putting $\mu_i^e(w)(X) := \sum_{x \in X} \mu_i^e(w)(x)$. Lemma 10 provides another way of expressing $\mu_i^e(w)(X)$, which clearly exhibits the fact that carrying out the experiments is related to Bayesian conditionalization. This lemma will often be used tacitly in the remainder of the thesis.

Lemma 10. 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, $w \in W^e$ an arbitrary state, and $X \subseteq W^e$ an arbitrary set. Then

$$\mu_i^e(w)(X) = \frac{\mu_i(w)(X \cap E_i[w])}{\mu_i(w)(E_i[w])}$$

Proof. We calculate:

$$
\begin{aligned}
\mu_i^e(w)(X) &= \sum_{x \in X} \mu_i^e(w)(x) \\
&= \sum_{x \in X} \frac{\mu_i(w)(\{x\} \cap E_i[w])}{\mu_i(w)(E_i[w])} \\
&= \sum_{\substack{x \in X \\ x \in E_i[w]}} \frac{\mu_i(w)(\{x\} \cap E_i[w])}{\mu_i(w)(E_i[w])} + \sum_{\substack{x \in X \\ x \notin E_i[w]}} \frac{\mu_i(w)(\{x\} \cap E_i[w])}{\mu_i(w)(E_i[w])} \\
&= \sum_{\substack{x \in X \\ x \in E_i[w]}} \frac{\mu_i(w)(x)}{\mu_i(w)(E_i[w])} + \sum_{\substack{x \in X \\ x \notin E_i[w]}} \frac{\mu_i(w)(\emptyset)}{\mu_i(w)(E_i[w])} \\
&= \sum_{x \in X \cap E_i[w]} \frac{\mu_i(w)(x)}{\mu_i(w)(E_i[w])} + 0 \\
&= \frac{\mu_i(w)(X \cap E_i[w])}{\mu_i(w)(E_i[w])}
\end{aligned}
$$

$\qquad\square$

Recall that carrying out an experiment is analogous to asking a question to nature. Lemma 11 says that once this question has been answered, the agents cannot gain any new (epistemic or probabilistic) information by asking it again. In other words: it does not make sense to carry out the experiments twice. This lemma will be used in Appendix C to establish the soundness of the formula $[\mathrm{EXP}]\,[\mathrm{EXP}]\,\varphi \leftrightarrow [\mathrm{EXP}]\,\varphi$.

Lemma 11. 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. Then $(\mathbb{M}^e)^e = \mathbb{M}^e$.

Proof. We establish the identity of $\mathbb{M}^e$ and $(\mathbb{M}^e)^e$ componentwise. First of all, note that $(W^e)^e = W^e = W$, that $(E_i^e)^e = E_i^e = E_i$ (for $i = 1, 2$), and that $(V^e)^e = V^e = V$. Furthermore, note that $(R_i^e)^e = R_i^e \cap E_i^e = R_i \cap E_i \cap E_i = R_i \cap E_i = R_i^e$ (for $i = 1, 2$). It remains to be proved that $(\mu_i^e)^e = \mu_i^e$ (for $i = 1, 2$). Consider arbitrary states $w, v \in (W^e)^e = W^e$ and $i \in \{1, 2\}$; we will prove that $(\mu_i^e)^e(w)(v) = \mu_i^e(w)(v)$. Recall that $E_i^e = E_i$, and hence $E_i^e[w] \cap E_i[w] = E_i[w]$. Now we compute:

$$
\begin{aligned}
(\mu_i^e)^e(w)(v) \quad &= \quad \frac{\mu_i^e(w)(\{v\} \cap E_i^e[w])}{\mu_i^e(w)(E_i^e[w])} \quad = \quad \frac{\frac{\mu_i(w)(\{v\} \cap E_i^e[w] \cap E_i[w])}{\mu_i(w)(E_i[w])}}{\frac{\mu_i(w)(E_i^e[w] \cap E_i[w])}{\mu_i(w)(E_i[w])}} \\[2ex]
&= \quad \frac{\mu_i(w)(\{v\} \cap E_i[w])}{\mu_i(w)(E_i[w])} \quad = \quad \mu_i^e(w)(v)
\end{aligned}
$$

$\square$

Remark 12. We already noted that Lemma 11 intuitively says that it does not make sense for the agents to carry out their experiments twice. Technically, this can be interpreted as saying that nature does not 'play tricks' on the agents: they perform their experiments, i.e. they pose their questions to nature, and nature answers these questions in a 'normal' fashion (i.e. not by means of Moore-like sentences).

Remark 13. Methodologically, Lemma 11 can be used to justify our omission of *repeated experiments*. It is perfectly conceivable that the agents, rather than each performing just one experiment, each perform an entire sequence of (different) experiments. Each experiment further refines the partition (or experiment relation). Hence, agent i's entire sequence of experiments can be regarded as imposing one very fine-grained partition on the set of possible states, i.e. it can be regarded as *one* very detailed experiment. Because we allow E_i to be *any* equivalence relation (no matter how fine-grained it is),[12] it can also capture these very detailed experiments —in other words, it can also capture sequences of experiments.

A remark on language. As we saw in Chapter 2, Aumann's theorem is often phrased in terms of 'priors' and 'posteriors'. In our formal language, however, we can express statements about probabilities only by formulas of the form '$P_i(\varphi) \geq k$'. Whether the symbol P_i is interpreted as agent i's prior or as her posterior, depends on whether or not it is in the scope of an [EXP]-operator. For example, $P_i(\varphi) = k$ can be read as 'agent i's prior for φ is k', whereas [EXP] $P_i(\varphi) = \ell$ can be read as 'agent i's posterior for φ is ℓ'.

3.3   Dynamics: the communication phase

We will now model the second type of dynamics described in Section 2.2, viz. the communication phase. Informally, we treat the communication as a *dialogue about* φ, i.e. a sequence in which the agents each repeatedly communicate the

---

[12] We will return to this point in Subection 5.1.3.

subjective probability they assign to φ (at that point in the dialogue). These communications are modeled as public announcements.[13]

In Section 3.3.1 we provide a formal definition of public announcement. Then, in Section 3.3.2, we will move to *dialogues*, i.e. particular sequences of public announcements, and show that these always reach a fixed point after finitely many steps.

3.3.1 Public announcements

We first introduce single public announcements. Syntactically, we add a new dynamic operator $[!\cdot]$ to the language $\mathcal{L}([\text{EXP}])$, thus obtaining the language $\mathcal{L}([\text{EXP}], [!\cdot])$, which has the following Backus-Naur form:

$$\begin{aligned} \varphi \quad ::= \quad & p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K_i\varphi \mid R_i\varphi \mid C^\varphi\varphi \mid X^\varphi\varphi \mid \\ & a_1 P_i(\varphi_1) + \cdots + a_n P_i(\varphi_n) \geq k \mid [\text{EXP}]\,\varphi \mid [!\varphi]\varphi \end{aligned}$$

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:

$$\mathbb{M}, w \models [!\varphi]\psi \quad \text{iff} \quad (\text{if } \mathbb{M}, w \models \varphi \text{ then } \mathbb{M}^\varphi, w \models \psi)$$

Definition 14. 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. 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:

- $W^\varphi := [\![\varphi]\!]^{\mathbb{M}} = \{w \in W \mid \mathbb{M}, w \models \varphi\}$

- $R_i^\varphi := R_i \cap ([\![\varphi]\!]^{\mathbb{M}} \times [\![\varphi]\!]^{\mathbb{M}})$

- $E_i^\varphi := E_i \cap ([\![\varphi]\!]^{\mathbb{M}} \times [\![\varphi]\!]^{\mathbb{M}})$

- for all $w \in W^\varphi$, put $\mu_i^\varphi(w) \colon W^\varphi \to [0,1] \colon v \mapsto \mu_i^\varphi(w)(v) := \frac{\mu_i(w)(\{v\} \cap [\![\varphi]\!]^{\mathbb{M}})}{\mu_i(w)([\![\varphi]\!]^{\mathbb{M}})}$

- for all $p \in Prop$, put $V^\varphi(p) := V(p) \cap [\![\varphi]\!]^{\mathbb{M}}$

Discussion. We will now justify our definition of the model update operation $\mathbb{M} \mapsto \mathbb{M}^\varphi$ by showing that it nicely captures the intuitive idea of the public announcement of a formula φ. As usual, the main effect of the public announcement of φ is that all $\neg\varphi$-states get deleted. The other components, R_i, E_i and V, change accordingly. For example, $R_i^\varphi = R_i \cap ([\![\varphi]\!]^{\mathbb{M}} \times [\![\varphi]\!]^{\mathbb{M}})$ —in words: two states are epistemically indistinguishable for agent i *after* the announcement

---

[13]Especially in the two-agent case, it is very plausible to model the dialogue as a series of public announcements. In n-agent cases ($n > 2$), this decision loses some of its plausibility, as the agents can then, for example, communicate their probabilities *pairwise* [34, 35].

of φ iff they were indistinguishable for i *before* the announcement and the announcement does not discriminate between them (in the sense that they both verify the announced formula φ).

We now turn to the probabilistic component. The definition of $\mu_i^\varphi(w)$ can be rewritten in terms of conditional probabilities: $\mu_i^\varphi(w)(x) = \mu_i(w)(x \mid [\![\varphi]\!]^{\mathbb{M}})$; i.e. agent i conditionalizes on (the information conveyed by) the formula that was publicly announced. This idea can also be expressed in the object language, by means of the following formula (cf. [26, p. 394]):[14]

$$\varphi \longrightarrow \big([!\varphi]P_i(\psi) = k \leftrightarrow P_i([!\varphi]\psi \mid \varphi) = k\big)$$

It is easy to check that this formula is true on all probabilistic Kripke models. The antecedent mentions the truthfulness precondition of public announcements. The consequent says that public announcement is related to Bayesian conditionalization (modulo dynamic effects): agent i's probability for ψ *after* the public announcement of φ is the same as her probability *before* the announcement for $[!\varphi]\psi$, conditional on φ.

The worry that arose for μ_i^e in the previous section now arises again for μ_i^φ. Note that for any states $w, v \in W^\varphi$, we defined $\mu_i^\varphi(w)(v)$ as the fraction

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

—but is this fraction always defined? What if $\mu_i(w)([\![\varphi]\!]^{\mathbb{M}}) = 0$? This issue is (again) solved by relying on the power of condition (i) in probabilistic Kripke models/frames (although this time in interaction with the truth precondition of public announcements). Recall that this condition says that $\mu_i(w)(w) > 0$ (for all $w \in W$). Since we are dealing with a state $w \in W^\varphi = [\![\varphi]\!]^{\mathbb{M}}$, we get that $\mu_i(w)([\![\varphi]\!]^{\mathbb{M}}) \geq \mu_i(w)(w) > 0$. We thus do not have to worry about $\mu_i(w)([\![\varphi]\!]^{\mathbb{M}}) = 0$.

Definition 14 seems to match well with our intuitive idea of what a public announcement of φ is, and how it influences the agents' knowledge and probabilities. We will now illustrate this.

Example 15. Consider the following scenario. Agent 1 does not 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. (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 $W := \{w, v\}$, $R_1 := W \times W$, $\mu_1(w)(w) = \mu_1(w)(v) = \frac{1}{2}$, and $V(p) = \{w\}$ (we do not care about the definitions of $E_1, \mu_1(v), R_2, E_2$, and μ_2). It is easy to see that this model is a faithful representation of the above scenario. Consider, for example:

---

[14]We use $P_i([!\varphi]\psi \mid \varphi) = k$ as an abbreviation for $P_i(\varphi \wedge [!\varphi]\psi) = kP_i(\varphi)$.

$$\mathbb{M}, w \models \quad \neg K_1 p \wedge \neg K_1 \neg p$$
$$\mathbb{M}, w \models \quad P_1(p) = \tfrac{1}{2} \wedge P_1(\neg p) = \tfrac{1}{2}$$

Now suppose that p is publicly announced (this is indeed possible, since p was assumed to be true at the actual state), i.e. consider the updated model $\mathbb{M}^p$. Applying Definition 14, we see that $W^p = \{w\}, R_1^p = \{(w,w)\}$ and that

$$\mu_1^p(w)(\llbracket p \rrbracket^{\mathbb{M}^p}) = \mu_1^p(w)(w) = \frac{\mu_1(w)(\{w\} \cap \llbracket p \rrbracket^{\mathbb{M}})}{\mu_1(w)(\llbracket p \rrbracket^{\mathbb{M}})} = \frac{\mu_1(w)(w)}{\mu_1(w)(w)} = 1$$
$$\mu_1^p(w)(\llbracket \neg p \rrbracket^{\mathbb{M}^p}) = \mu_1^p(w)(\llbracket p \rrbracket^{\mathbb{M}} - \llbracket p \rrbracket^{\mathbb{M}}) = \mu_1^p(w)(\emptyset) = 0$$

Since $\mathbb{M}, w \models p$, the precondition of $[!p]$ is fulfilled, and hence:

$$\mathbb{M}, w \models \quad [!p] K_1 p$$
$$\mathbb{M}, w \models \quad [!p](P_1(p) = 1 \wedge P_1(\neg p) = 0)$$

So after the public announcement of p, 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 14 seems to be a natural way of representing public announcements: it makes the intuitively right 'predictions' about the agents' knowledge and probabilities.

Experiments versus public announcements. The previous example shows that there is an important similarity between the experiment update ($\mathbb{M} \mapsto \mathbb{M}^e$) and the public announcement of a formula φ ($\mathbb{M} \mapsto \mathbb{M}^\varphi$). Both types of dynamics tell the agents to update their probabilities in the same way, viz. by Bayesian conditionalization. With the experiments, the agents conditionalize on the outcome of the experiment; with public announcements, they conditionalize on the content of the announcement.

The difference between both types of dynamics lies in the epistemic component. With the experiments, the agents gain knowledge in a *semi-private* way: each agent learns the outcome of her own experiment, but not that of the other agent's experiment (but she does know the *structure* of the other agent's experiment —hence *semi*-private). With public announcements, as can be expected from the name, the agents gain knowledge in a fully *public* way.

Some useful lemmas. In the previous section we already showed that the class of probabilistic Kripke models is closed under the experiment update ($\mathbb{M} \mapsto \mathbb{M}^e$). We now show that it is also closed under the public announcement update ($\mathbb{M} \mapsto \mathbb{M}^\varphi$).

Lemma 16. If $\mathbb{M} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V \rangle$ is a probabilistic Kripke model and $\varphi \in \mathcal{L}([\mathrm{EXP}], [!\cdot])$ a formula, then $\mathbb{M}^\varphi$ is a probabilistic Kripke model as well.

Proof. First of all, note that $W^\varphi \subseteq W$ is finite. It is easily verified that R_i^φ and E_i^φ are equivalence relations on W^φ.

We now check that $\mu_i^\varphi(w)$ is a probability mass function on W^φ (for all $w \in W^\varphi$). Consider $w \in W^\varphi = [\![\varphi]\!]^{\mathbb{M}}$ arbitrary. Note that $\mu_i^\varphi(w)(v) = \frac{\mu_i(w)(\{v\}\cap[\![\varphi]\!]^{\mathbb{M}})}{\mu_i(w)([\![\varphi]\!]^{\mathbb{M}})} \in [0,1]$ for any $v \in W^\varphi$, and furthermore, we compute:

$$
\begin{aligned}
\sum_{v\in W^\varphi}\mu_i^\varphi(w)(v) &= \sum_{v\in[\![\varphi]\!]^{\mathbb{M}}}\frac{\mu_i(w)(\{v\}\cap[\![\varphi]\!]^{\mathbb{M}})}{\mu_i(w)([\![\varphi]\!]^{\mathbb{M}})}\\
&= \sum_{v\in[\![\varphi]\!]^{\mathbb{M}}}\frac{\mu_i(w)(v)}{\mu_i(w)([\![\varphi]\!]^{\mathbb{M}})}\\
&= \frac{\mu_i(w)([\![\varphi]\!]^{\mathbb{M}})}{\mu_i(w)([\![\varphi]\!]^{\mathbb{M}})} \qquad = \quad 1
\end{aligned}
$$

Finally, we check that $\mathbb{M}^\varphi$ still satisfies conditions (i) and (ii) of Definition 1. Note that for any $w \in W^\varphi = [\![\varphi]\!]^{\mathbb{M}} \subseteq W$, it holds that $\mu_i(w)(w) > 0$ (by condition (i) for the model $\mathbb{M}$), so

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

and thus condition (i) is fulfilled. Furthermore, note that if $w, v \in W^\varphi = [\![\varphi]\!]^{\mathbb{M}}$ and $(w,v) \notin R_i^\varphi$, then $(w,v) \notin R_i$, so $\mu_i(w)(v) = 0$ (by condition (ii) for the model $\mathbb{M}$) and thus

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

This shows that condition (ii) is fulfilled. $\qquad\square$

Just like $\mu_i(w)$, also $\mu_i^\varphi(w)$ can be lifted to the level of sets, by putting $\mu_i^\varphi(w)(X) := \sum_{x\in X}\mu_i^\varphi(w)(x)$. Lemma 17 provides another way of expressing $\mu_i^\varphi(w)(X)$, which clearly exhibits the fact that public announcement is related to Bayesian conditionalization.

Lemma 17. 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, φ an arbitrary formula, $w \in W^\varphi$ an arbitrary state, and $X \subseteq W^\varphi$ an arbitrary set. Then

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

Proof. Completely analogous to the proof of Lemma 10. $\qquad\square$

The following lemma says that an uninformative sentence (i.e. one which is true at every state in the model) does not change the model at all.

Lemma 18. 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 φ an arbitrary formula. Suppose that $[\![\varphi]\!]^{\mathbb{M}} = W$. Then $\mathbb{M}^\varphi = \mathbb{M}$.

Proof. We check the identity of the models $\mathbb{M}^\varphi$ and $\mathbb{M}$ componentwise. For the models' sets of states, note that $W^\varphi = [\![\varphi]\!]^\mathbb{M} = W$. For the models' epistemic accessibility relations, note that $R_i^\varphi = R_i \cap ([\![\varphi]\!]^\mathbb{M} \times [\![\varphi]\!]^\mathbb{M}) = R_i \cap (W \times W) = R_i$. And similarly $E_i^\varphi = E_i$, and $V^\varphi(p) = V(p)$ for all $p \in Prop$. Finally, we turn the models' probability functions. For any $w, v \in W^\varphi = W$, we have that

$$\mu_i^\varphi(w)(v) = \frac{\mu_i(w)(\{v\} \cap [\![\varphi]\!]^\mathbb{M})}{\mu_i(w)([\![\varphi]\!]^\mathbb{M})} = \frac{\mu_i(w)(\{v\} \cap W)}{\mu_i(w)(W)} = \frac{\mu_i(w)(v)}{1} = \mu_i(w)(v)$$

$\square$

The following lemma provides reduction axioms for the public announcement operator $[!\varphi]$. This shows that the language $\mathcal{L}([! \cdot])$ is no more expressive than $\mathcal{L}$, i.e. for every sentence $\varphi \in \mathcal{L}([! \cdot])$ these reduction axioms allow us to find a sentence $\varphi' \in \mathcal{L}$ such that $\varphi \leftrightarrow \varphi'$ is true on all probabilistic Kripke models.

Lemma 19. The following sentences are sound with respect to the class of probabilistic Kripke models:

$$
\begin{array}{lll}
[!\varphi]p & \leftrightarrow & \varphi \to p \quad (\text{for } p \in Prop) \\
[!\varphi]\neg\psi & \leftrightarrow & \varphi \to \neg[!\varphi]\psi \\
[!\varphi](\psi \wedge \chi) & \leftrightarrow & [!\varphi]\psi \wedge [!\varphi]\chi \\
[!\varphi]K_i\psi & \leftrightarrow & \varphi \to K_i[!\varphi]\psi \\
[!\varphi]R_i\psi & \leftrightarrow & \varphi \to R_i[!\varphi]\psi \\
[!\varphi]C^\psi\chi & \leftrightarrow & \varphi \to C^{\varphi \wedge [!\varphi]\psi}[!\varphi]\chi \\
[!\varphi]X^\psi\chi & \leftrightarrow & \varphi \to X^{\varphi \wedge [!\varphi]\psi}[!\varphi]\chi \\
[!\varphi]\sum_\ell a_\ell P_i(\varphi_\ell) \geq k & \leftrightarrow & \varphi \to \sum_\ell a_\ell P_i(\varphi \wedge [!\varphi]\varphi_\ell) \geq k P_i(\varphi)
\end{array}
$$

Proof. We prove the soundness of the reduction axioms for $K_i\psi$, $X^\psi\chi$, and $\sum_\ell a_\ell P_i(\varphi_\ell) \geq k$, and leave the details of the other cases to the reader. Let $\mathbb{M} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V \rangle$ be an arbitrary Kripke model, and $w \in W$ an arbitrary state. For $K_i\psi$ we have:

$$
\begin{array}{lll}
\mathbb{M}, w \models [!\varphi]K_i\psi & \text{iff} & \text{if } \mathbb{M}, w \models \varphi \text{ then } \mathbb{M}^\varphi, w \models K_i\psi \\
& \text{iff} & \text{if } \mathbb{M}, w \models \varphi \text{ then } \forall v \in W^\varphi : (w, v) \in R_i^\varphi \Rightarrow \mathbb{M}^\varphi, v \models \psi \\
& \text{iff} & \text{if } \mathbb{M}, w \models \varphi \text{ then } \forall v \in W : (w, v) \in R_i \Rightarrow \\
& & \qquad\qquad\qquad\qquad (\mathbb{M}, v \models \varphi \Rightarrow \mathbb{M}^\varphi, v \models \psi) \\
& \text{iff} & \mathbb{M}, w \models \varphi \to K_i[!\varphi]\psi
\end{array}
$$

For $X^\psi\chi$ we have:

$$
\begin{array}{lll}
\mathbb{M}, w \models [!\varphi]X^\psi\chi & \text{iff} & \text{if } \mathbb{M}, w \models \varphi \text{ then } \mathbb{M}^\varphi, w \models X^\psi\chi \\
& \text{iff} & \text{if } \mathbb{M}, w \models \varphi \text{ then } \forall v \in W^\varphi : (w, v) \in \big((R^\varphi)^e \cap (W^\varphi \times [\![\psi]\!]^{\mathbb{M}^\varphi})\big)^+ \\
& & \qquad\qquad\qquad\qquad\qquad \Rightarrow \mathbb{M}^\varphi, v \models \psi \\
& \text{iff} & \text{if } \mathbb{M}, w \models \varphi \text{ then } \forall v \in W : (w, v) \in \big(R^e \cap (W \times [\![\varphi \wedge [!\varphi]\psi]\!]^\mathbb{M})\big)^+ \\
& & \qquad\qquad\qquad\qquad\qquad \Rightarrow (\mathbb{M}, v \models \varphi \Rightarrow \mathbb{M}^\varphi, v \models \psi) \\
& \text{iff} & \mathbb{M}, w \models \varphi \to X^{\varphi \wedge [!\varphi]\psi}[!\varphi]\chi
\end{array}
$$

29

And finally, for $\sum_\ell a_\ell P_i(\varphi_\ell) \geq k$ we have:

$\mathbb{M}, w \models [!\varphi] \sum_\ell a_\ell P_i(\varphi_\ell) \geq k$

$\quad$ iff $\quad$ if $\mathbb{M}, w \models \varphi$ then $\mathbb{M}^\varphi, w \models \sum_\ell a_\ell P_i(\varphi_\ell) \geq k$

$\quad$ iff $\quad$ if $\mathbb{M}, w \models \varphi$ then $\sum_\ell a_\ell \mu_i^\varphi(w)(\llbracket \varphi_\ell \rrbracket^{\mathbb{M}^\varphi}) \geq k$

$\quad$ iff $\quad$ if $\mathbb{M}, w \models \varphi$ then $\sum_\ell a_\ell \frac{\mu_i(w)(\llbracket \varphi_\ell \rrbracket^{\mathbb{M}^\varphi} \cap \llbracket \varphi \rrbracket^{\mathbb{M}})}{\mu_i(w)(\llbracket \varphi \rrbracket^{\mathbb{M}})} \geq k$

$\quad$ iff $\quad$ if $\mathbb{M}, w \models \varphi$ then $\sum_\ell a_\ell \mu_i(w)(\llbracket \varphi \wedge [!\varphi]\varphi_\ell \rrbracket^{\mathbb{M}}) \geq k\mu_i(w)(\llbracket \varphi \rrbracket^{\mathbb{M}})$

$\quad$ iff $\quad$ $\mathbb{M}, w \models \varphi \rightarrow \sum_\ell a_\ell P_i(\varphi \wedge [!\varphi]\varphi_\ell) \geq kP_i(\varphi)$

$\hfill \square$

Repeated announcements. In this subsection we have discussed single public announcements. In the next subsection we will move to (a specific kind of) sequences of announcements. First, however, we prove two lemmas about repeated announcements in general. The first one says that two consecutive public announcements can always be contracted into one, more complex announcement.

Lemma 20. 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 $w \in W$ an arbitrary state. Then

$$\mathbb{M}, w \models [!\varphi_1][!\varphi_2]\psi \leftrightarrow [!(\varphi_1 \wedge [!\varphi_1]\varphi_2)]\psi$$

for all $\psi \in \mathcal{L}([!\,\cdot])$.

Proof. Let $\psi \in \mathcal{L}([!\,\cdot])$ arbitrary. We claim that $(\mathbb{M}^{\varphi_1})^{\varphi_2} = \mathbb{M}^{\varphi_1 \wedge [!\varphi_1]\varphi_2}$. This claim justifies step $(*)$ in the following chain of equivalences:

$\mathbb{M}, w \models [!\varphi_1][!\varphi_2]\psi$

$\quad \Leftrightarrow \quad$ if $\mathbb{M}, w \models \varphi_1$ then $\mathbb{M}^{\varphi_1}, w \models [!\varphi_2]\psi$

$\quad \Leftrightarrow \quad$ if $\mathbb{M}, w \models \varphi_1$ then (if $\mathbb{M}^{\varphi_1}, w \models \varphi_2$ then $(\mathbb{M}^{\varphi_1})^{\varphi_2}, w \models \psi$)

$\quad \Leftrightarrow \quad$ if $(\mathbb{M}, w \models \varphi_1$ and $\mathbb{M}^{\varphi_1}, w \models \varphi_2)$ then $(\mathbb{M}^{\varphi_1})^{\varphi_2}, w \models \psi$

$\quad \overset{*}{\Leftrightarrow} \quad$ if $\mathbb{M}, w \models \varphi_1 \wedge [!\varphi_1]\varphi_2$ then $\mathbb{M}^{\varphi_1 \wedge [!\varphi_1]\varphi_2}, w \models \psi$

$\quad \Leftrightarrow \quad$ $\mathbb{M}, w \models [!(\varphi_1 \wedge [!\varphi_1]\varphi_2)]\psi$

We will now prove the claim that $(\mathbb{M}^{\varphi_1})^{\varphi_2} = \mathbb{M}^{\varphi_1 \wedge [!\varphi]\varphi_2}$ by checking the identity of these models componentwise. For the models' *sets of states*, note that $(W^{\varphi_1})^{\varphi_2} = \llbracket \varphi_2 \rrbracket^{\mathbb{M}^{\varphi_1}} = \llbracket \varphi_1 \wedge [!\varphi_1]\varphi_2 \rrbracket^{\mathbb{M}} = W^{\varphi_1 \wedge [!\varphi_1]\varphi_2}$. For the models' *epistemic accessibility relations*, note that

$$(R_i^{\varphi_1})^{\varphi_2} \;=\; R_i^{\varphi_1} \cap (\llbracket \varphi_2 \rrbracket^{\mathbb{M}^{\varphi_1}} \times \llbracket \varphi_2 \rrbracket^{\mathbb{M}^{\varphi_1}})$$
$$=\; R_i \cap (\llbracket \varphi_1 \rrbracket^{\mathbb{M}} \times \llbracket \varphi_1 \rrbracket^{\mathbb{M}}) \cap (\llbracket \varphi_2 \rrbracket^{\mathbb{M}^{\varphi_1}} \times \llbracket \varphi_2 \rrbracket^{\mathbb{M}^{\varphi_1}})$$
$$=\; R_i \cap (\llbracket \varphi_1 \wedge [!\varphi_1]\varphi_2 \rrbracket^{\mathbb{M}} \times \llbracket \varphi_1 \wedge [!\varphi_1]\varphi_2 \rrbracket^{\mathbb{M}})$$
$$=\; R_i^{\varphi_1 \wedge [!\varphi_1]\varphi_2}$$

and similarly we show that $(E_i^{\varphi_1})^{\varphi_2} = E_i^{\varphi_1 \wedge [!\varphi_1]\varphi_2}$ and that $(V^{\varphi_1})^{\varphi_2}(p) = V^{\varphi_1 \wedge [!\varphi_1]}(p)$ for all $p \in Prop$. Finally, we turn to the models' *probability functions*. Let $w \in (W^{\varphi_1})^{\varphi_2} = W^{\varphi_1 \wedge [!\varphi_1]\varphi_2}$ arbitrary; then we have:

$$(\mu_i^{\varphi_1})^{\varphi_2}(w)(v) \;=\; \frac{\mu_i^{\varphi_1}(w)(\{v\} \cap \llbracket \varphi_2 \rrbracket^{\mathbb{M}^{\varphi_1}})}{\mu_i^{\varphi_1}(w)(\llbracket \varphi_2 \rrbracket^{\mathbb{M}^{\varphi_1}})}$$
$$=\; \frac{\frac{\mu_i(w)(\{v\} \cap \llbracket \varphi_2 \rrbracket^{\mathbb{M}^{\varphi_1}} \cap \llbracket \varphi_1 \rrbracket^{\mathbb{M}})}{\mu_i(w)(\llbracket \varphi_1 \rrbracket^{\mathbb{M}})}}{\frac{\mu_i(w)(\llbracket \varphi_2 \rrbracket^{\mathbb{M}^{\varphi_1}} \cap \llbracket \varphi_1 \rrbracket^{\mathbb{M}})}{\mu_i(w)(\llbracket \varphi_1 \rrbracket^{\mathbb{M}})}}$$
$$=\; \frac{\mu_i(w)(\{v\} \cap \llbracket \varphi_1 \wedge [!\varphi_1]\varphi_2 \rrbracket^{\mathbb{M}})}{\mu_i(w)(\llbracket \varphi_1 \wedge [!\varphi_1]\varphi_2 \rrbracket^{\mathbb{M}})}$$
$$=\; \mu_i^{\varphi_1 \wedge [!\varphi_1]\varphi_2}(w)(v)$$

$\square$

The final lemma is a generalization of the previous one. It says that *any* number of consecutive public announcements can be contracted into one, more complex announcement.

Lemma 21. 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 $w \in W$ an arbitrary state. Then for all $n \geq 1$, there is a formula Φ_n such that

$$\mathbb{M}, w \models [!\varphi_1] \cdots [!\varphi_n]\psi \leftrightarrow [!\Phi_n]\psi$$

for all $\psi \in \mathcal{L}([!\,\cdot])$ and also

$$\mathbb{M}, w \models \varphi_1 \wedge [!\varphi_1]\varphi_2 \wedge [!\varphi_1][!\varphi_2]\varphi_3 \wedge \cdots \wedge [!\varphi_1] \cdots [!\varphi_{n-1}]\varphi_n \leftrightarrow \Phi_n$$

Proof. We prove this by induction on n. For the base case, $n = 1$, we need to show that there exists a formula Φ_1 such that (for all $\psi \in \mathcal{L}([!\,\cdot])$):

$$\mathbb{M}, w \models [!\varphi_1]\psi \leftrightarrow [!\Phi_1]\psi$$

and

$$\mathbb{M}, w \models \varphi_1 \leftrightarrow \Phi_1$$

This is trivial: simply put $\Phi_1 := \varphi_1$. Now the induction case. The induction hypothesis is that there exists a formula Φ_n such that

$$\mathbb{M}, w \models [!\varphi_1] \cdots [!\varphi_n]\psi \leftrightarrow [!\Phi_n]\psi \tag{3.1}$$

31

for all $\psi \in \mathcal{L}([!\,\cdot])$ and

$$\mathbb{M}, w \models \varphi_1 \wedge [!\varphi_1]\varphi_2 \wedge [!\varphi_1][!\varphi_2]\varphi_3 \wedge \cdots \wedge [!\varphi_1] \cdots [!\varphi_{n-1}]\varphi_n \leftrightarrow \Phi_n \qquad (3.2)$$

Now define $\Phi_{n+1} := \Phi_n \wedge [!\Phi_n]\varphi_{n+1}$; we will show that this sentence has the desired properties. First, note that for any $\psi \in \mathcal{L}([!\,\cdot])$, also $[!\varphi_{n+1}]\psi \in \mathcal{L}([!\,\cdot])$ and thus it follows from (3.2) that

$$\mathbb{M}, w \models [!\varphi_1] \cdots [!\varphi_n][!\varphi_{n+1}]\psi \leftrightarrow [!\Phi_n][!\varphi_{n+1}]\psi$$

and thus by Lemma 20 it follows that also

$$\mathbb{M}, w \models [!\varphi_1] \cdots [!\varphi_n][!\varphi_{n+1}]\psi \leftrightarrow [!(\Phi_n \wedge [!\Phi_n]\varphi_{n+1})]\psi$$

Second, note that from (3.2) it follows that

$$\mathbb{M}, w \models \varphi_1 \wedge [!\varphi_1]\varphi_2 \wedge \cdots \wedge [!\varphi_1] \cdots [!\varphi_{n-1}]\varphi_n \wedge [!\varphi_1] \cdots [!\varphi_n]\varphi_{n+1} \leftrightarrow \Phi_n \wedge [!\varphi_1] \cdots [!\varphi_n]\varphi_{n+1}$$

so by a final application of (3.1) (putting $\psi := \varphi_{n+1}$) we get that

$$\mathbb{M}, w \models \varphi_1 \wedge [!\varphi_1]\varphi_2 \wedge \cdots \wedge [!\varphi_1] \cdots [!\varphi_n]\varphi_{n+1} \leftrightarrow \Phi_n \wedge [!\Phi_n]\varphi_{n+1}$$

$\square$

3.3.2 Dialogues

We will now move from a *single* public announcement to *sequences* of public announcements. We will focus on one particular type of such a sequence, 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). We 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 $\mu_1(w)(\llbracket\varphi\rrbracket^{\mathbb{M}}) = a$ and $\mu_2(w)(\llbracket\varphi\rrbracket^{\mathbb{M}}) = b$; i.e. such that $\mathbb{M}, w \models P_1(\varphi) = a \wedge P_2(\varphi) = b$. We now define the sentence $d(\mathbb{M}, w, \varphi)$ as follows:

$$d(\mathbb{M}, w, \varphi) := P_1(\varphi) = a \wedge P_2(\varphi) = b$$

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

$$\mathbb{M}, w \models d(\mathbb{M}, w, \varphi) \qquad (3.3)$$

Rationals versus reals. Note that we have moved outside the official object language here, because we are writing formulas like $P_1(\varphi) = a \wedge P_2(\varphi) = b$, with *real* numbers a, b, whereas the official object language only contains *rational* numbers (cf. supra).

One might try to avoid this problem by making sure that $\mu_i(w)(\llbracket\varphi\rrbracket^{\mathbb{M}})$ is *itself* a rational number, so that we can write $P_i(\varphi) = a$ (for $a := \mu_i(w)(\llbracket\varphi\rrbracket^{\mathbb{M}}) \in \mathbb{Q}$), without moving out of the official object language. However, because this needs

to hold for *all* states w and formulas φ, this is not a trivial requirement. The easiest way to satisfy it is by requiring that $\mu_i(w)$ is a rational probability mass function ($\mu_i(w)\colon W \to [0,1] \cap \mathbb{Q}$) for each state $w \in W$. Since W is finite, it follows for every $X \subseteq W$ that also

$$\mu_i(w)(X) = \sum_{x \in X} \mu_i(w)(x) \in \mathbb{Q}$$

because a finite sum of rational numbers is itself also rational.

Note that this solution cannot be generalized to models with infinite domains, because infinite sums of rational numbers might themselves be irrational. A well-known example is:

$$\sum_{n \geq 0} \frac{1}{n!} = e \in \mathbb{R} - \mathbb{Q}$$

Therefore we choose to tacitly extend the language, and allow expressions of the form $\sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) \geq k$, with $a_\ell \in \mathbb{Q}$, but $k \in \mathbb{R}$. To keep this particular choice independent of the rest of the material presented in this thesis, we nowhere make use of this language extension (in particular, we do not use it in the proofs of the lemmas in Section 5.2).

A single step in the dialogue. We will now model a single step in the dialogue. Such a single step 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 $\mathbb{M}$ (cf. Definition 14). Formally:

$$f_{w,\varphi}(\mathbb{M}) := \mathbb{M}^{d(\mathbb{M},w,\varphi)} \tag{3.4}$$

The following lemma guarantees that it makes sense to reiterate $f_{w,\varphi}$, i.e. that expressions such as $f_{w,\varphi}^n(\mathbb{M})$ make sense for all $n \geq 1$.

Lemma 22. If $\mathbb{M}$ belongs to the domain of $f_{w,\varphi}$, then so does $f_{w,\varphi}(\mathbb{M})$.

Proof. Suppose that $\mathbb{M}$ belongs to the domain of $f_{w,\varphi}$, i.e. $\mathbb{M}$ is a probabilistic Kripke model that contains the state w. By Lemma 16, also $f_{w,\varphi}(\mathbb{M}) = \mathbb{M}^{d(\mathbb{M},w,\varphi)}$ is a probabilistic Kripke model. Furthermore, by (3.3) we have that $w \in [\![d(\mathbb{M}, w, \varphi)]\!]^{\mathbb{M}} = W^{d(\mathbb{M},w,\varphi)}$, which by (3.4) is the set of states of the model $f_{w,\varphi}(\mathbb{M})$. $\qquad \square$

Illustration. Consider a probabilistic Kripke model $\mathbb{M}$ that contains the state w. Unraveling the definitions, we see that

$$f_{w,\varphi}^2(\mathbb{M}) = f_{w,\varphi}(f_{w,\varphi}(\mathbb{M})) = \left(\mathbb{M}^{d(\mathbb{M},w,\varphi)}\right)^{d(\mathbb{M}^{d(\mathbb{M},w,\varphi)},w,\varphi)}$$

At the first stage $(\mathbb{M}, w)$, the agents announce the probabilities they assign to φ at that stage of the dialogue: $d(\mathbb{M}, w, \varphi)$. Because of this announcement, the agents' probability distributions can change (cf. Definition 14). At the next stage $(\mathbb{M}^{d(\mathbb{M}, w, \varphi)})$, the agents again announce the probabilities they assign to φ at *that* stage of the dialogue: $d(\mathbb{M}^{d(\mathbb{M}, w, \varphi)}, w, \varphi)$.

The dialogue. We are now ready to model the entire dialogue about φ, 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 Lemma 22, it makes sense to repeatedly apply $f_{w,\varphi}$ to $\mathbb{M}$. Hence, we obtain a sequence which looks as follows:

$$\mathbb{M} \mapsto f_{w,\varphi}(\mathbb{M}) \mapsto f_{w,\varphi}^2(\mathbb{M}) \mapsto f_{w,\varphi}^3(\mathbb{M}) \mapsto f_{w,\varphi}^4(\mathbb{M}) \mapsto \cdots$$

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.

Lemma 23. Consider a probabilistic Kripke model $\mathbb{M}$ that contains the state w. Then there exists an $n \in \mathbb{N}$ such that

$$f_{w,\varphi}^n(\mathbb{M}) = f_{w,\varphi}^{n+1}(\mathbb{M})$$

Proof. For any probabilistic Kripke model $\mathbb{K}$, we write $|\mathbb{K}|$ for the number of states that $\mathbb{K}$ contains. For a reductio, suppose that

$$\text{for all } n \in \mathbb{N} : f_{w,\varphi}^n(\mathbb{M}) \neq f_{w,\varphi}^{n+1}(\mathbb{M})$$

We claim (for any $n \in \mathbb{N}$) that if $f_{w,\varphi}^n(\mathbb{M}) \neq f_{w,\varphi}^{n+1}(\mathbb{M})$, then $|f_{w,\varphi}^n(\mathbb{M})| > |f_{w,\varphi}^{n+1}(\mathbb{M})|$. Hence it follows that

$$\text{for all } n \in \mathbb{N} : |f_{w,\varphi}^n(\mathbb{M})| > |f_{w,\varphi}^{n+1}(\mathbb{M})|$$

Hence we find that

$$|\mathbb{M}| > |f_{w,\varphi}(\mathbb{M})| > |f_{w,\varphi}^2(\mathbb{M})| > |f_{w,\varphi}^3(\mathbb{M})| > \cdots$$

But this implies that $|\mathbb{M}|$ is infinitely large, which contradicts the fact that probabilistic Kripke models have, by definition, *finite* sets of states.

We now prove the claim that if $f_{w,\varphi}^n(\mathbb{M}) \neq f_{w,\varphi}^{n+1}(\mathbb{M})$, then $|f_{w,\varphi}^n(\mathbb{M})| > |f_{w,\varphi}^{n+1}(\mathbb{M})|$ (for any $n \in \mathbb{N}$). Let $n \in \mathbb{N}$ arbitrary and suppose that $f_{w,\varphi}^n(\mathbb{M}) \neq f_{w,\varphi}^{n+1}(\mathbb{M})$. We will show that $|f_{w,\varphi}^n(\mathbb{M})| > |f_{w,\varphi}^{n+1}(\mathbb{M})|$.

We first make an easy observation. Let W^n denote the set of states of $f_{w,\varphi}^n(\mathbb{M})$, and W^{n+1} the set of states of $f_{w,\varphi}^{n+1}(\mathbb{M})$. Note that

$$W^{n+1} = (W^n)^{d(f_{w,\varphi}^n(\mathbb{M}), w, \varphi)} = [\![d(f_{w,\varphi(\mathbb{M})}^n, w, \varphi)]\!]^{f_{w,\varphi}^n(\mathbb{M})} \subseteq W^n$$

and thus $|f_{w,\varphi}^n(\mathbb{M})| \geq |f_{w,\varphi}^{n+1}(\mathbb{M})|$. Hence it suffices to show that $|f_{w,\varphi}^n(\mathbb{M})| \neq |f_{w,\varphi}^{n+1}(\mathbb{M})|$.

For a reductio, suppose that $|f_{w,\varphi}^n(\mathbb{M})| = |f_{w,\varphi}^{n+1}(\mathbb{M})|$. This means that by applying $f_{w,\varphi}$, no states were deleted from $f_{w,\varphi}^n(\mathbb{M})$. Since exactly those states are deleted where $\delta := d(f_{w,\varphi}^n(\mathbb{M}), w, \varphi)$ is *false*, this means that δ is true at all states of $f_{w,\varphi}^n(\mathbb{M})$, i.e. $[\![\delta]\!]^{f_{w,\varphi}^n(\mathbb{M})} = W^n$. It now follows by Lemma 18 that $f_{w,\varphi}^{n+1}(\mathbb{M}) = f_{w,\varphi}(f_{w,\varphi}^n(\mathbb{M})) = \left(f_{w,\varphi}^n(\mathbb{M})\right)^\delta = f_{w,\varphi}^n(\mathbb{M})$, contradiction. $\qquad\square$

Remark 24. From a mathematical perspective, Lemma 23 is rather trivial: the proof is entirely based on the simple observation that since there are only finitely many states in the beginning (and no new states can be created), the state-deleting process must stop after finitely many steps. Despite its mathematical triviality, this lemma has important conceptual consequences; cf. Lemma 37 and Theorem 38 in Section 4.2.

Language and semantics. We are now ready to provide an exact definition of how we model the communication dynamics. Syntactically, we add the $[\mathrm{DIAL}(\,\cdot\,)]$-operator to the language $\mathcal{L}([\mathrm{EXP}], [!\,\cdot\,])$, thus obtaining the language $\mathcal{L}([\mathrm{EXP}], [!\,\cdot\,], [\mathrm{DIAL}(\,\cdot\,)])$, which has the following Backus-Naur form:

$$\varphi \quad ::= \quad p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K_i\varphi \mid R_i\varphi \mid C^\varphi\varphi \mid X^\varphi\varphi \mid$$
$$a_1 P_i(\varphi_1) + \cdots + a_n P_i(\varphi_n) \geq k \mid [\mathrm{EXP}]\varphi \mid [!\varphi]\varphi \mid [\mathrm{DIAL}(\varphi)]\varphi$$

The $[\mathrm{DIAL}(\varphi)]$-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 23 guarantees that such a fixed point will indeed always be reached after finitely many steps). Hence, $[\mathrm{DIAL}(\varphi)]\psi$ 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 $[\mathrm{DIAL}(\varphi)]$ involves going to the fixed point model $\mathbb{M}^{dial_w(\varphi)}$, which is defined immediately afterwards.

$$\mathbb{M}, w \models [\mathrm{DIAL}(\varphi)]\psi \quad \text{iff} \quad \mathbb{M}^{dial_w(\varphi)}, w \models \psi$$

Definition 25. 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 $\varphi \in \mathcal{L}([\mathrm{EXP}], [\mathrm{DIAL}(\,\cdot\,)])$ an arbitrary formula. Then we define

$$\mathbb{M}^{dial_w(\varphi)} := f_{w,\varphi}^n(\mathbb{M})$$

—where n is the least number such that

$$f_{w,\varphi}^n(\mathbb{M}) = f_{w,\varphi}^{n+1}(\mathbb{M})$$

(this number is guaranteed to exist, because of Lemma 23).

Remark 26. Recall that we assume public announcements to be *truthful*; hence, the semantics of $[!\varphi]$ involves the precondition that the formula being announced is true. Furthermore, we have modeled dialogues about φ as a sequence of public announcements. However, the semantics of $[\mathrm{DIAL}(\varphi)]$ does not involve any preconditions.

The reason for this is that the formulas being announced throughout the sequence are true *by definition*, cf. (3.3). 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 $\mathbb{K}$), we can safely leave the truth precondition out.

Lemma 27. If $\mathbb{M} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V \rangle$ is a probabilistic Kripke model, $w \in W$ a state, and $\varphi \in \mathcal{L}([\mathrm{EXP}], [\mathrm{DIAL}(\,\cdot\,)])$ a formula, then $\mathbb{M}^{dial_w(\varphi)}$ is a probabilistic Kripke model as well.

Proof. By Definition 25, $\mathbb{M}^{w, dial_w(\varphi)}$ is obtained out of $\mathbb{M}$ by means of a finite series of public announcements. By Lemma 16 the class of probabilistic Kripke models is closed under the public announcement update. Hence $\mathbb{M}^{dial_w(\varphi)}$ is a probabilistic Kripke model. $\qquad\square$

Contracting a dialogue. We will now discuss the possibility of representing a dialogue about φ, which is a series of public announcements, as one single public announcement.

Consider an arbitrary probabilistic Kripke model $\mathbb{M}$ that contains the state w. Let n be the number at which $f_{w,\varphi}$ reaches a fixed point. Hence $\mathbb{M}^{dial_w(\varphi)} = f_{w,\varphi}^n(\mathbb{M})$, by Definition 25. Recall (3.3) and (3.4), which say that $\mathbb{K}, w \models d(\mathbb{K}, w, \varphi)$ and that $f_{w,\varphi}(\mathbb{K}) = \mathbb{K}^{d(\mathbb{K},w,\varphi)}$, for any probabilistic Kripke model $\mathbb{K}$ that contains w. Hence:

$$
\begin{aligned}
\mathbb{M}, w \models [\mathrm{DIAL}(\varphi)]\psi \quad &\Leftrightarrow \quad \mathbb{M}^{dial_w(\varphi)}, w \models \psi \\
&\Leftrightarrow \quad f_{w,\varphi}^n(\mathbb{M}), w \models \psi \\
&\Leftrightarrow \quad \mathbb{M}, w \models [!d(\mathbb{M}, w, \varphi)][!d(f_{w,\varphi}(\mathbb{M}), w, \varphi)]\cdots \\
&\qquad\qquad \cdots [!d(f_{w,\varphi}^{n-1}(\mathbb{M}), w, \varphi)]\psi
\end{aligned}
$$

Remark 28. Note that this chain of equivalences does *not* mean that $[\mathrm{DIAL}(\varphi)]$ is *definable* in terms of public announcements, because the formulas being announced are dependent on the concrete model $\mathbb{M}$ and state w. (Furthermore, also the *number* of public announcements, viz. n, is dependent on the concrete model $\mathbb{M}$.)

The following lemma applies the 'contraction methodology' (developed in general at the end of the previous subsection) to the special case of a dialogue about φ. It will be used in the proof of the second agreement theorem (Theorem 38) in Chapter 4.

Lemma 29. Let $\mathbb{M}$ be an arbitrary probabilistic Kripke model containing the state w, and let φ be a formula. Then there exists a sentence $\Phi(\mathbb{M}, w, \varphi)$ such that for all $\psi \in \mathcal{L}([! \cdot])$:

$$\mathbb{M}, w \models [\mathrm{DIAL}(\varphi)]\psi \leftrightarrow [!\Phi(\mathbb{M}, w, \varphi)]\psi$$

and also:

$$\mathbb{M}, w \models \Phi(\mathbb{M}, w, \varphi)$$

Proof. Let n be the number at which $f_{w,\varphi}$ reaches a fixed point. By the chain of equivalences established above, it holds for all $\psi \in \mathcal{L}([! \cdot])$ that

$$\mathbb{M}, w \models [\mathrm{DIAL}(\varphi)]\psi \leftrightarrow [!d(\mathbb{M}, w, \varphi)] \cdots [!d(f_{w,\varphi}^{n-1}(\mathbb{M}), w, \varphi)]\psi \qquad (3.5)$$

So the formulas being announced are $d(f_{w,\varphi}^{\ell}(\mathbb{M}), w, \varphi)$ (for $0 \leq \ell < n$). By Lemma 21 there exists a sentence $\Phi(\mathbb{M}, w, \varphi)$ such that

$$\mathbb{M}, w \models [!d(\mathbb{M}, w, \varphi)] \cdots [!d(f_{w,\varphi}^{n-1}(\mathbb{M}), w, \varphi)]\psi \leftrightarrow [!\Phi(\mathbb{M}, w, \varphi)]\psi \qquad (3.6)$$

for all $\psi \in \mathcal{L}([! \cdot])$ and

$$\begin{aligned} \mathbb{M}, w \models \quad & d(\mathbb{M}, w, \varphi) \wedge [!d(\mathbb{M}, w, \varphi)]d(f_{w,\varphi}(\mathbb{M}), w, \varphi) \wedge \cdots \\ & \cdots \wedge [!d(\mathbb{M}, w, \varphi)] \cdots [!d(f_{w,\varphi}^{n-2}(\mathbb{M}), w, \varphi)]d(f_{w,\varphi}^{n-1}(\mathbb{M}), w, \varphi) \leftrightarrow \Phi(\mathbb{M}, w, \varphi) \end{aligned}$$

All the conjuncts of the left-hand-side of this sentence are true, by definition of $d(\mathbb{K}, w, \varphi)$ (for any probabilistic Kripke model $\mathbb{K}$); see also (3.3) and Remark 26. Hence it follows that

$$\mathbb{M}, w \models \Phi(\mathbb{M}, w, \varphi)$$

Furthermore, by combining (3.5) and (3.6) we immediately get for all $\psi \in \mathcal{L}([! \cdot])$ that

$$\mathbb{M}, w \models [\mathrm{DIAL}(\varphi)]\psi \leftrightarrow [!\Phi(\mathbb{M}, w, \varphi)]\psi$$

$$\square$$

Chapter 4

Agreement theorems in PDEL

We have now arrived at the main chapter of this thesis. Using the semantic setup introduced in the previous chapter, we will formulate and prove various dynamic agreement theorems in probabilistic dynamic epistemic logic. In Section 4.1 we discuss agreement theorems that make the experimentation dynamics explicit, but still leave the communication implicit. In Section 4.2 we build on this and formulate agreement theorems that make both the experimentation *and* the communication dynamics explicit.

4.1 Agreement theorems in PDEL: only experimentation

Before turning to the first agreement theorem in probabilistic dynamic epistemic logic, we formulate and prove two easy auxiliary lemmas. These lemmas are used to streamline the proof of the actual agreement theorem, so that we can focus on the main strategy of the proof rather than on the small technical details.

Lemma 30. 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 $i \in \{1, 2\}$. We first show that $R^*[w] = \bigcup_{x \in R^*[w]} R_i[x]$, by proving that both sets are included in each other. For the $\subseteq$-direction, note that for any $v \in R^*[w]$, we have (by the reflexivity of R_i) that $v \in R_i[v] \subseteq \bigcup_{x \in R^*[w]} R_i[x]$. For the $\supseteq$-direction, note that for any $v \in \bigcup_{x \in R^*[w]} R_i[x]$, we know that there exists an $x \in R^*[w]$ such that $v \in R_i[x]$, from which it follows that $v \in R^*[w]$.

Since W is finite (cf. Definition 1), also $R^*[w]$ is finite, say $R^*[w] = \{v_1, \ldots, v_n\}$ (note that since R^* is reflexive, $w = v_\ell$ for some $1 \le \ell \le n$). Since $R^*[w] = \bigcup_{x \in R^*[w]} R_i[x]$, this means that we can write $R^*[w] = R_i[v_1] \cup \cdots \cup R_i[v_n]$. Since R_i is an equivalence relation, we know for $1 \le \ell, m \le n$ that either $R_i[v_\ell] = R_i[v_m]$ or $R_i[v_\ell] \cap R_i[v_m] = \emptyset$. By systematically deleting the 'redundant' $R_i[v_\ell]$ (i.e. if $\ell \ne m$ and $R_i[v_\ell] = R_i[v_m]$, then delete exactly one of $R_i[v_\ell]$ and $R_i[v_m]$; keep repeating until stabilization), we obtain a (finite) partition of $R^*[w]$ in cells of the form $R_i[v_\ell]$, as desired. $\qquad\square$

Lemma 31. 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 element Y_ℓ of the partition it holds that $\mu_i(w)(Y_\ell) > 0$ and that

$$\frac{\mu_i(w)(X \cap Y_\ell)}{\mu_i(w)(Y_\ell)} = a \tag{4.1}$$

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 $0 < \mu_i(w)(Y_1) \le \mu_i(w)(Y)$. This proves the first part. For the second part, note that

$$
\begin{aligned}
\mu_i(w)(X \cap Y) &= \mu_i(w)\big(X \cap \bigcup_{\ell=1}^m Y_\ell\big) \\
&= \mu_i(w)\big(\bigcup_{\ell=1}^m (X \cap Y_\ell)\big) \quad \text{(set theory)} \\
&= \sum_{\ell=1}^m \mu_i(w)(X \cap Y_\ell) \quad \text{(partition)} \\
&= \sum_{\ell=1}^m a \cdot \mu_i(w)(Y_\ell) \quad (4.1) \\
&= a \cdot \mu_i(w)\big(\bigcup_{\ell=1}^m Y_\ell\big) \quad \text{(partition)} \\
&= a \cdot \mu_i(w)(Y)
\end{aligned}
$$

Since $\mu_i(w)(Y) > 0$ it follows that $\frac{\mu_i(w)(X \cap Y)}{\mu_i(w)(Y)} = a$. $\qquad\square$

We are now ready to formulate and prove the first agreement theorem for probabilistic dynamic epistemic logic:

Theorem 32. 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 $\mathbb{M}, w \models [\text{EXP}]\, C(P_1(\varphi) = a \wedge P_2(\varphi) = b)$; we will show that $\mathbb{M}, w \models a = b$. Since $a = b$ abbreviates $aP_1(\top) + (-b)P_1(\top) = 0$, this means that we have to prove that $a = b$.

By Lemma 9, $\mathbb{M}^e$ is a probabilistic Kripke model. Applying Lemma 30 to $\mathbb{M}^e$ (for agent 1), we 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 any ℓ between 1 and 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 $\mu_1^e(v_\ell)(\llbracket\varphi\rrbracket^{\mathbb{M}^e}) = a$ (†). 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:

$$
\begin{aligned}
a &= \mu_1^e(v_\ell)(\llbracket\varphi\rrbracket^{\mathbb{M}^e}) & \text{(†)}\\
&= \frac{\mu_1(v_\ell)(\llbracket\varphi\rrbracket^{\mathbb{M}^e} \cap E_1[v_\ell])}{\mu_1(v_\ell)(E_1[v_\ell])} & \text{(Def. 7)}\\
&= \frac{\mu_1(v_\ell)(\llbracket\varphi\rrbracket^{\mathbb{M}^e} \cap E_1[v_\ell] \cap R_1[v_\ell])}{\mu_1(v_\ell)(E_1[v_\ell] \cap R_1[v_\ell])} & \text{(Lemma 3)}\\
&= \frac{\mu_1(w)(\llbracket\varphi\rrbracket^{\mathbb{M}^e} \cap R_1^e[v_\ell])}{\mu_1(w)(R_1^e[v_\ell])} & \text{(‡)}
\end{aligned}
$$

(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 ℓ was chosen arbitrarily, this holds for all $1 \leq \ell \leq m$. By Lemma 31 it now follows that $\mu_1(w)\big((R^e)^*[w]\big) > 0$ and

$$
\frac{\mu_1(w)\big(\llbracket\varphi\rrbracket^{\mathbb{M}^e} \cap (R^e)^*[w]\big)}{\mu_1(w)\big((R^e)^*[w]\big)} = a \tag{4.2}
$$

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

$$
\frac{\mu_2(w)\big(\llbracket\varphi\rrbracket^{\mathbb{M}^e} \cap (R^e)^*[w]\big)}{\mu_2(w)\big((R^e)^*[w]\big)} = b \tag{4.3}
$$

Now recall condition 1 of this theorem: $\mu_1(w) = \mu_2(w)$. Hence (4.2) and (4.3) together imply that $a = b$. $\qquad\square$

Remark 33. Upon close inspection of the proof, it should be clear that condition 2 can be replaced with a slightly weaker version, which does not quantify over $R^*[w]$, but over $(R^e)^*[w]$. The reasons for using the stronger condition in the main presentation of the theorem are conceptual in nature and will be discussed in Chapter 5; cf. Remark 47.

Remark 34. The reader familiar with Aumann's paper [2] will probaby have noticed that the proof of our 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; cf. Section 7.3). We think this shows that probabilistic Kripke models are a natural setting in which to formalize (dynamic) agreement theorems.

Intuitive interpretation. The theorem is essentially a sentence of the formal language $\mathcal{L}([EXP])$, 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 says 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. In other words: the second piece of dynamics described in Section 2.2, the communication, is not yet explicitly represented. We will return to this in the next section.

The two conditions. The two conditions required to prove the agreement theorem are fairly weak. We now discuss their technical import and intuitive interpretation. 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 $x, y \in W$. Our theorem shows that this assumption can be weakened: the local version $(\mu_i(x) = \mu_i(w)$ for all $x \in R^*[w])$ suffices. In Section 5.2 we will show that we do not even need common knowledge to characterize this property: individual knowledge suffices.

Local versus global. It should be noted that Theorem 32 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, also the reference to the concrete state w can 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 35. 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:

$$\mathbb{F} \models [EXP]\, C(P_1(\varphi) = a \wedge P_2(\varphi) = b) \rightarrow a = b$$

Proof. Let $V \colon Prop \rightarrow \wp(W)$ be an arbitrary valuation on $\mathbb{F}$ and $w \in W$ an arbitrary state. Since the conditions of this theorem are simply the 'de-localized' versions of the two conditions of Theorem 32, it follows immediately by that theorem that $\langle \mathbb{F}, V \rangle, w \models [EXP]\, C(P_1(\varphi) = a \wedge P_2(\varphi) = b) \rightarrow a = b$. $\qquad \square$

Remark 36. Because this theorem is an immediate generalization of Theorem 32, also Remark 33 can be generalized. The second condition of Theorem 35 can be weakened by replacing the condition that $(w, v) \in R^*$ with the condition that $(w, v) \in (R^e)^*$. As was already mentioned in Remark 33, the conceptual reasons for using the stronger condition will be discussed in Chapter 5; cf. Remark 47.

4.2   Agreement theorems in PDEL: experimentation *and* communication

We 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, we need to prove one more auxiliary lemma. Intuitively, this lemma says that after a dialogue about φ, the agents' probabilities for φ have become common knowledge.

Lemma 37. 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

$$\mathbb{M}, w \models [\mathrm{DIAL}(\varphi)]\big((P_1(\varphi) = a \wedge P_2(\varphi) = b) \to C(P_1(\varphi) = a \wedge P_2(\varphi) = b)\big)$$

Proof. Suppose that

$$\mathbb{M}^{dial_w(\varphi)}, w \models P_1(\varphi) = a \wedge P_2(\varphi) = b$$

Note that this means that

$$\delta := d(\mathbb{M}^{dial_w(\varphi)}, w, \varphi) \; = \; \big(P_1(\varphi) = a \wedge P_2(\varphi) = b\big)$$

Let n be the least natural 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 23). Note that $\mathbb{M}^{dial_w(\varphi)} = f^n_{w,\varphi}(\mathbb{M})$, by Definition 25. Since $f^n_{w,\varphi}(\mathbb{M}) = f^{n+1}_{w,\varphi}(\mathbb{M})$ we have that $[\![\delta]\!]^{f^n_{w,\varphi}(\mathbb{M})}$ is the entire set of states of $f^n_{w,\varphi}(\mathbb{M})$, i.e.: $f^n_{w,\varphi}(\mathbb{M}) \models \delta$. From this it trivially follows that $f^n_{w,\varphi}(\mathbb{M}), w \models C\delta$, i.e.

$$\mathbb{M}^{dial_w(\varphi)}, w \models C(P_1(\varphi) = a \wedge P_2(\varphi) = b)$$

$\square$

We are now ready to formulate and prove the second agreement theorem for probabilistic dynamic epistemic logic, which explicitly represents both the experimentation and the communication dynamics:

Theorem 38. 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:

$$\mathbb{M}, w \models [\text{EXP}][\text{DIAL}(\varphi)](P_1(\varphi) = a \wedge P_2(\varphi) = b) \rightarrow a = b$$

Proof. Assume that $\mathbb{M}, w \models [\text{EXP}][\text{DIAL}(\varphi)](P_1(\varphi) = a \wedge P_2(\varphi) = b)$; we will show that $\mathbb{M}, w \models a = b$. Since $a = b$ abbreviates $aP_1(\top) + (-b)P_1(\top) = 0$, this means that we have to prove that $a = b$.

By Lemma 29 there exists a sentence $\Phi(\mathbb{M}^e, w, \varphi)$ such that

$$\mathbb{M}^e, w \models \Phi(\mathbb{M}^e, w, \varphi) \tag{4.4}$$

$$\mathbb{M}^e, w \models [\text{DIAL}(\varphi)]\psi \leftrightarrow [!\Phi(\mathbb{M}^e, w, \varphi)]\psi \qquad (\text{for } \psi \in \mathcal{L}([! \cdot])) \tag{4.5}$$

Since $\mathbb{M}, w \models [\text{EXP}][\text{DIAL}(\varphi)](P_1(\varphi) = a \wedge P_2(\varphi) = b)$, it follows that

$$\mathbb{M}^e, w \models [\text{DIAL}(\varphi)](P_1(\varphi) = a \wedge P_2(\varphi) = b)$$

By Lemma 9, we know that $\mathbb{M}^e$ is still a probabilistic Kripke model that contains w. Hence, by Lemma 37 it follows that

$$\mathbb{M}^e, w \models [\text{DIAL}(\varphi)]C(P_1(\varphi) = a \wedge P_2(\varphi) = b)$$

It now follows by (4.5) that

$$\mathbb{M}^e, w \models [!\Phi(\mathbb{M}^e, w, \varphi)]C(P_1(\varphi) = a \wedge P_2(\varphi) = b)$$

By (4.4) it follows that

$$(\mathbb{M}^e)^{\Phi(\mathbb{M}^e, w, \varphi)}, w \models C(P_1(\varphi) = a \wedge P_2(\varphi) = b) \tag{4.6}$$

We will abbreviate $\mathcal{R}_i := (R_i^e)^{\Phi(\mathbb{M}^e, w, \varphi)}$ and $\mathcal{R} := \mathcal{R}_1 \cup \mathcal{R}_2$.

By Lemmas 9 and 16, $(\mathbb{M}^e)^{\Phi(\mathbb{M}^e, w, \varphi)}$ is a probabilistic Kripke model. Hence, Lemma 30 (applied to agent 1) allows us to express $\mathcal{R}^*[w] = \mathcal{R}_1[v_1] \cup \cdots \cup \mathcal{R}_1[v_m]$, with all the $\mathcal{R}_1[v_\ell]$ pairwise disjoint. Now consider any ℓ between 1 and m. Since $\mathcal{R}_1$ is reflexive, we have $v_\ell \in \mathcal{R}_1[v_\ell] \subseteq \mathcal{R}^*[w]$. By (4.6) this implies that $(\mathbb{M}^e)^{\Phi(\mathbb{M}^e, w, \varphi)}, v_\ell \models P_1(\varphi) = a \wedge P_2(\varphi) = b$, and hence

$$(\mu_1^e)^{\Phi(\mathbb{M}^e, w, \varphi)}(v_\ell)(X) = a \tag{4.7}$$

(We abbreviate $X := [\![\varphi]\!]^{(\mathbb{M}^e)^{\Phi(\mathbb{M}^e, w, \varphi)}}$.) Note that

$$v_\ell \in \mathcal{R}^*[w] \subseteq (W^e)^{\Phi(\mathbb{M}^e, w, \varphi)} = [\![\Phi(\mathbb{M}^e, w, \varphi)]\!]^{\mathbb{M}^e} \tag{4.8}$$

Hence we also have that

$$\mathcal{R}_1[v_\ell] = (R_1^e)^{\Phi(\mathbb{M}^e, w, \varphi)}[v_\ell] = R_1^e[v_\ell] \cap [\![\Phi(\mathbb{M}^e, w, \varphi)]\!]^{\mathbb{M}^e} = R_1[v_\ell] \cap E_1[v_\ell] \cap [\![\Phi(\mathbb{M}^e, w, \varphi)]\!]^{\mathbb{M}^e} \tag{4.9}$$

Finally, note that since $v_\ell \in \mathcal{R}^*[w] \subseteq R^*[w]$, condition 2 applies to v_ℓ, and thus

$$\mu_i(w) = \mu_i(v_\ell) \tag{4.10}$$

We now put everything together, and find that

$$a \quad = (\mu_1^e)^{\Phi(\mathbb{M}^e, w, \varphi)}(v_\ell)(X) \tag{4.7}$$

$$= \frac{\mu_1^e(v_\ell)(X \cap [\![\Phi(\mathbb{M}^e, w, \varphi)]\!]^{\mathbb{M}^e})}{\mu_1^e(v_\ell)([\![\Phi(\mathbb{M}^e, w, \varphi)]\!]^{\mathbb{M}^e})} \tag{4.8}$$

$$= \frac{\mu_1(v_\ell)(X \cap [\![\Phi(\mathbb{M}^e, w, \varphi)]\!]^{\mathbb{M}^e} \cap E_1[v_\ell])}{\mu_1(v_\ell)([\![\Phi(\mathbb{M}^e, w, \varphi)]\!]^{\mathbb{M}^e} \cap E_1[v_\ell])} \tag{Def. 7}$$

$$= \frac{\mu_1(v_\ell)(X \cap [\![\Phi(\mathbb{M}^e, w, \varphi)]\!]^{\mathbb{M}^e} \cap E_1[v_\ell] \cap R_1[v_\ell])}{\mu_1(v_\ell)([\![\Phi(\mathbb{M}^e, w, \varphi)]\!]^{\mathbb{M}^e} \cap E_1[v_\ell] \cap R_1[v_\ell])} \tag{Lemma 3}$$

$$= \frac{\mu_1(v_\ell)(X \cap \mathcal{R}_1[v_\ell])}{\mu_1(v_\ell)(\mathcal{R}_1[v_\ell])} \tag{4.9}$$

$$= \frac{\mu_1(w)(X \cap \mathcal{R}_1[v_\ell])}{\mu_1(w)(\mathcal{R}_1[v_\ell])} \tag{4.10}$$

As ℓ was chosen arbitrarily, this holds for all $1 \le \ell \le m$. By Lemma 31 it now follows that

$$\frac{\mu_1(w)\big(X \cap \mathcal{R}^*[w]\big)}{\mu_1(w)\big(\mathcal{R}^*[w]\big)} = a \tag{4.11}$$

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

$$\frac{\mu_2(w)\big(X \cap \mathcal{R}^*[w]\big)}{\mu_2(w)\big(\mathcal{R}^*[w]\big)} = b \tag{4.12}$$

Now recall condition 1 of this theorem: $\mu_1(w) = \mu_2(w)$. Hence (4.11) and (4.12) together imply that $a = b$. □

Remark 39. Note that Remark 33 also applies here: condition 2 can be replaced with a slightly weaker version, which does not quantify over $R^*[w]$, but over $\mathcal{R}^*[w]$.

Intuitive interpretation. The theorem is essentially a sentence of the formal language $\mathcal{L}([\text{EXP}], [\text{DIAL}(\cdot)])$, and 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.

Remark 40. In the first agreement theorem, we said that if the agents have common knowledge of their posteriors, then these posteriors have to be identical. However, we said nothing about how this common knowledge is to be achieved, i.e. we did not say anything about the communication. Now, however, we *do* explicitly represent the communication dynamics, and we thus no longer need the common knowledge operator in the formulation of the theorem: the existence of common knowledge can now be derived as the result of the communication.

Local versus global, bis. We again obtain a *global frame version* of the agreement theorem by 'de-localizing' the assumptions:

Theorem 41. 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:

$$\mathbb{F} \models [\text{EXP}] \, [\text{DIAL}(\varphi)](P_1(\varphi) = a \land P_2(\varphi) = b) \rightarrow a = b$$

Proof. Let $V : Prop \rightarrow \wp(W)$ be an arbitrary valuation on $\mathbb{F}$ and $w \in W$ an arbitrary state. Since the conditions of this theorem are simply the 'de-localized' versions of the two conditions of Theorem 38, it follows immediately by that theorem that $\langle \mathbb{F}, V \rangle, w \models [\text{EXP}] \, [\text{DIAL}(\varphi)](P_1(\varphi) = a \land P_2(\varphi) = b) \rightarrow a = b$. $\square$

Chapter 5

Metatheory

In the previous chapter we have formulated and proved various dynamic agreement theorems in probabilistic dynamic epistemic logic. So far, all results have been strictly *semantical* in nature: they speak about various sentences holding at models/frames. In this chapter we will focus more on the *syntactic* side.[1] In Section 5.1 we describe a technical difficulty related to the syntactic perspective on probabilistic epistemic logic in general. In Subsections 5.1.1 and 5.1.2 we propose two solutions: incorporating notions from hybrid logic and restricting to binary experiments, respectively. In Subsection 5.1.3 we argue that the second strategy is to be preferred for methodological as well as technical reasons. In Section 5.2 we provide characterization results for all the agreement theorems that were proved in Chapter 4. These characterization results are then used in Section 5.3 to obtain various axiomatizations. Finally, in Section 5.4 we show that these axiomatizations are sound and complete.

5.1 A difficulty about expressivity

Our 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 $\mathcal{L}([\text{EXP}])$. We 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 $P_i(\varphi) \geq k$:

$$\mathbb{M}, w \models P_i(\varphi) \geq k \quad \text{iff} \quad \mu_i(w)(\llbracket \varphi \rrbracket^{\mathbb{M}}) \geq k$$

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

---

[1]We refer the reader to [28, Section 1] for more comments on the syntax/semantics interface and its relevance to game theorists.

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).

In modeling the experiments, we defined $\mu_i^e(w)(v)$ in terms of $\mu_i(w)(\{v\} \cap E_i[w])$. 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, consider what happens when we try to find an [EXP]-reduction axiom for $P_i(\varphi) \geq k$:[2]

$$
\begin{aligned}
\mathbb{M}, w \models [\text{EXP}]\, P_i(\varphi) \geq k \quad &\text{iff} \quad \mathbb{M}^e, w \models P_i(\varphi) \geq k \\
&\text{iff} \quad \mu_i^e(w)([\![\varphi]\!]^{\mathbb{M}^e}) \geq k \\[2ex]
&\text{iff} \quad \frac{\mu_i(w)([\![\varphi]\!]^{\mathbb{M}^e} \cap E_i[w])}{\mu_i(w)(E_i[w])} \geq k \\[2ex]
&\text{iff} \quad \frac{\mu_i(w)([\![[\text{EXP}]\,\varphi]\!]^{\mathbb{M}} \cap E_i[w])}{\mu_i(w)(E_i[w])} \geq k
\end{aligned}
$$

Although we only have probability mass functions $\mu_i(w)$ on the final line, we cannot go any further, because $E_i[w]$ and/or $[\![[\text{EXP}]\,\varphi]\!]^{\mathbb{M}} \cap E_i[w]$ might not be expressible in $\mathcal{L}([\text{EXP}])$.

We have analyzed the expressivity problem and shown that it boils down to $E_i[w]$ being potentially an undefinable set. Hence, to solve the problem we need to make sure that $E_i[w]$ is always definable by means of some formula. We will now discuss two different ways of doing exactly this. Subsection 5.1.1 introduces a solution that incorporates elements from hybrid logic. Subsection 5.1.2 involves restricting to *binary* experiments. In Subsection 5.1.3 we evaluate both solutions and argue that the second one is to be preferred to the first one, on both technical and methodological grounds.

5.1.1 Solution 1: hybrid logic

The first solution is based on hybrid logic. We will not provide a general introduction to hybrid logic in all its details, but rather introduce those elements that are really needed (in the context of probabilistic Kripke models). A more general introduction to hybrid logic can be found in [1].

We introduce a set *Nom* of *nominals*. These nominals behave exactly like proposition letters (elements of *Prop*), except that the valuation V of a model $\langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V \rangle$ always maps them to *singleton* sets —so if $n \in Nom$, then $V(n) = \{w\}$ for some state $w \in W$. Because a nominal n is satisfied at exactly one state w, the nominal n can function as a *name* for the state w.

Next, we extend the language with operators Q_i, for $i \in \{1, 2\}$.[3] In a probabilistic Kripke model $\mathbb{M} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V \rangle$, the operator Q_i is interpreted as follows:

---

[2] Ultimately, of course, we will want to obtain an [EXP]-reduction axiom not just for $P_i(\varphi) \geq k$, but for *arbitrary* i-probability formulas $a_1 P_i(\varphi_1) + \cdots + a_n P_i(\varphi_n) \geq k$. This general reduction axiom will be established in Sections 5.3 and 5.4. Our aim here, however, is merely to illustrate the syntactic expressivity issue; for this purpose, the restricted case $P_i(\varphi) \geq k$ already suffices.

[3] These operators are also introduced in the dynamic epistemic logic of questions [45].

$$\mathbb{M}, w \models Q_i\varphi \text{ iff } \forall v \in W : (w, v) \in E_i \Rightarrow \mathbb{M}, v \models \varphi$$

Recall that K_i is the universal modality for R_i and that R_i is the universal modality for $R_i \cap E_i$. Despite their close semantic interconnections, it is a well-known fact from modal logic that the R_i-operator is not definable in terms of the K_i- and Q_i-operators. In particular, the following is *not* valid: $\hat{R}_i\varphi \leftrightarrow (\hat{K}_i\varphi \wedge \hat{Q}_i\varphi)$ (where $\hat{\Box}$ abbreviates the *dual* of the operator $\Box$, i.e. $\hat{\Box} := \neg\Box\neg$). In hybrid logic, however, R_i *is* definable in terms of K_i and Q_i:

$$\hat{R}_i n \leftrightarrow (\hat{K}_i n \wedge \hat{Q}_i n) \qquad \text{(where } n \in Nom)$$

Now 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 nominal n that names w (so $V(n) = \{w\}$). Because E_i is an equivalence relation and thus symmetric, it is easy to see that

$$E_i[w] = [\![\hat{Q}_i n]\!]^{\mathbb{M}} \tag{5.1}$$

This means that the set $E_i[w]$ is now definable, as was desired. However, the formula that defines $E_i[w]$ contains the nominal n that names w. Since we want *all* sets $E_i[w]$ (for $w \in W$) to be definable, this means that we need to assume that for each state w (in each model $\mathbb{M}$) there exists a nominal $n_w \in Nom$ that names that state w. Using more tools from hybrid logic, however, we will be able to achieve the same result in a more elegant fashion.

First, we introduce a (countably infinite) set *SVar* of *state variables*. Each element of *SVar* is, by itself, a well-formed formula. Next, we introduce a new operator $\downarrow$ which behaves like a quantifier in first-order logic. We extend the language by stating that if φ is a well-formed formula and $x \in SVar$, then also $\downarrow x.\varphi$ is a well-formed formula.

To interpret these new formulas, we introduce *assignment functions*. If $\mathbb{M}$ is a model with domain W, then assignment functions are functions $g \colon SVar \to W$. If g is an assignment function, $w \in W$ is a state, and x is a state variable, then we define g_w^x to be the assignment function that is identical to g, except that it maps x onto w. Formally:

$$g_w^x \colon SVar \to W \colon y \mapsto \begin{cases} w & \text{if } y = x \\ g(y) & \text{otherwise} \end{cases}$$

Henceforth, all formulas will not be interpreted relative to a couple $(\mathbb{M}, w)$, but relative to a triple $(\mathbb{M}, g, w)$ —where $\mathbb{M}$ is a probabilistic Kripke model with domain W, $g \colon SVar \to W$ is an assignment function, and $w \in W$ is a state. The semantics for the 'old' formulas is as before (we just ignore the new component g), and the newly introduced formulas receive the following semantics:

$$\mathbb{M}, g, w \models x \qquad \text{iff} \quad g(x) = w \qquad \text{(for } x \in SVar)$$
$$\mathbb{M}, g, w \models \downarrow x.\varphi \quad \text{iff} \quad \mathbb{M}, g_w^x, w \models \varphi$$

Truth sets are now defined relative to both a model $\mathbb{M}$ *and* an assignment g:

$$[\![\varphi]\!]^{\mathbb{M},g} := \{w \in W \mid \mathbb{M}, g, w \models \varphi\}$$

It is straightforward to prove that if a formula φ does not contain the state variable x, then the value that the assignment g assigns to x is irrelevant for the interpretation of φ:

$$[\![\varphi]\!]^{\mathbb{M},g_w^x} = [\![\varphi]\!]^{\mathbb{M},g} \quad \text{(if } x \text{ does not occur in } \varphi) \tag{5.2}$$

Considering a probabilistic Kripke model $\mathbb{M} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V \rangle$, an assignment function $g \colon SVar \to W$ and a state $w \in W$, we now easily obtain an analogue of (5.1):

$$E_i[w] = [\![\hat{Q}_i x]\!]^{\mathbb{M},g_w^x} \tag{5.3}$$

Again, this means that we have found a formula that defines $E_i[w]$, as desired. Furthermore, the defining formula does not contain a nominal naming w, but just the 'neutral' state variable x. Of course, we are still 'smuggling in' reference to w via the special assignment function g_w^x, but this can be eliminated using the binding operator $\downarrow$. To illustrate this, we will show how we are now able to express an [EXP]-reduction axiom for $P_i(\varphi) \geq k$. Since φ is a finite formula, it contains only finitely many state variables. Since $SVar$ is infinite, we can thus always pick a state variable x that does not occur in φ (and thus also not in $[\text{EXP}]\,\varphi$). Now consider:

$$\mathbb{M}, g, w \models [\text{EXP}]\,P_i(\varphi) \geq k$$

$$\text{iff} \quad \mathbb{M}^e, g, w \models P_i(\varphi) \geq k$$

$$\text{iff} \quad \mu_i^e(w)([\![\varphi]\!]^{\mathbb{M}^e,g}) \geq k$$

$$\text{iff} \quad \frac{\mu_i(w)([\![\varphi]\!]^{\mathbb{M}^e,g} \cap E_i[w])}{\mu_i(w)(E_i[w])} \geq k$$

$$\text{iff} \quad \frac{\mu_i(w)([\![\text{[EXP]}\,\varphi]\!]^{\mathbb{M},g} \cap [\![\hat{Q}_i x]\!]^{\mathbb{M},g_w^x})}{\mu_i(w)([\![\hat{Q}_i x]\!]^{\mathbb{M},g_w^x})} \geq k \tag{5.3}$$

$$\text{iff} \quad \frac{\mu_i(w)([\![\text{[EXP]}\,\varphi]\!]^{\mathbb{M},g_w^x} \cap [\![\hat{Q}_i x]\!]^{\mathbb{M},g_w^x})}{\mu_i(w)([\![\hat{Q}_i x]\!]^{\mathbb{M},g_w^x})} \geq k \tag{5.2}$$

$$\text{iff} \quad \mathbb{M}, g_w^x, w \models P_i([\text{EXP}]\,\varphi \wedge \hat{Q}_i x) \geq k P_i(\hat{Q}_i x)$$

$$\text{iff} \quad \mathbb{M}, g, w \models \downarrow x. P_i([\text{EXP}]\,\varphi \wedge \hat{Q}_i x) \geq k P_i(\hat{Q}_i x)$$

5.1.2 Solution 2: binary experiments

The second solution involves only allowing *binary experiments*. The first, *syntactic* step of this strategy is to introduce two new, 'primitive' formulas α_1, α_2 into the languages $\mathcal{L}$, $\mathcal{L}([\text{EXP}])$ and $\mathcal{L}([\text{EXP}], [\text{DIAL}(\cdot)])$, thus obtaining respectively

$\mathcal{L}^\alpha$, $\mathcal{L}^\alpha([\text{EXP}])$ and $\mathcal{L}^\alpha([\text{EXP}], [\text{DIAL}(\cdot)])$. 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 exists (for $i = 1, 2$) a set $\mathcal{E}_i^\mathbb{F} \subseteq W$ such that

$$E_i = \left(\mathcal{E}_i^\mathbb{F} \times \mathcal{E}_i^\mathbb{F}\right) \cup \left((W - \mathcal{E}_i^\mathbb{F}) \times (W - \mathcal{E}_i^\mathbb{F})\right) \qquad (5.4)$$

(It is easy to check that under this assumption, E_i is still an equivalence relation.) In the third and final step, we link syntax and semantics by extending the valuations to the newly introduced α_i: for any valuation V on a probabilistic Kripke frame $\mathbb{F}$, we put $V(\alpha_i) := \mathcal{E}_i^\mathbb{F}$. Applying this definition to (5.4), we get:

$$E_i = \left(V(\alpha_i) \times V(\alpha_i)\right) \cup \left((W - V(\alpha_i)) \times (W - V(\alpha_i))\right) \qquad (5.5)$$

Compatibility with the dynamics. We need to check that (5.5) is 'compatible' with the two main types of dynamics discussed in this thesis, experimentation (Definition 7) and communication (Definition 25). First, note that if a probabilistic Kripke model $\mathbb{M} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V \rangle$ satisfies condition (5.5), then the experiment-updated model $\mathbb{M}^e$ will satisfy that condition as well:

$$
\begin{aligned}
E_i^e \quad &= E_i && (\text{Def. 7}) \\
&= \left(V(\alpha_i) \times V(\alpha_i)\right) \cup \left((W - V(\alpha_i)) \times (W - V(\alpha_i))\right) && (5.5) \\
&= \left(V^e(\alpha_i) \times V^e(\alpha_i)\right) \cup \left((W^e - V^e(\alpha_i)) \times (W^e - V^e(\alpha_i))\right) && (\text{Def. 7})
\end{aligned}
$$

To see that (5.5) is also compatible with the communication dynamics, first note that if a probabilistic Kripke model $\mathbb{M} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V \rangle$ satisfies condition (5.5), then the public announcement-updated model $\mathbb{M}^\varphi$ will satisfy that condition as well:

$$
\begin{aligned}
E_i^\varphi \quad &= E_i \cap (\llbracket\varphi\rrbracket^\mathbb{M} \times \llbracket\varphi\rrbracket^\mathbb{M}) && (\text{Def. 14}) \\
&= \left(\left(V(\alpha_i) \times V(\alpha_i)\right) \cup \left((W - V(\alpha_i)) \times (W - V(\alpha_i))\right)\right) \\
&\quad \cap (\llbracket\varphi\rrbracket^\mathbb{M} \times \llbracket\varphi\rrbracket^\mathbb{M}) && (5.5) \\
&= \left((V(\alpha_i) \cap \llbracket\varphi\rrbracket^\mathbb{M}) \times (V(\alpha_i) \cap \llbracket\varphi\rrbracket^\mathbb{M})\right) \cup \\
&\quad \left((\llbracket\varphi\rrbracket^\mathbb{M} - (V(\alpha) \cap \llbracket\varphi\rrbracket^\mathbb{M})) \times (\llbracket\varphi\rrbracket^\mathbb{M} - (V(\alpha) \cap \llbracket\varphi\rrbracket^\mathbb{M}))\right) \\
&= \left(V^\varphi(\alpha_i) \times V^\varphi(\alpha_i)\right) \cup \left((W^\varphi - V^\varphi(\alpha_i)) \times (W^\varphi - V^\varphi(\alpha_i))\right) && (\text{Def. 14})
\end{aligned}
$$

Since a dialogue about φ is a sequence of public announcements —which have just been shown to be compatible with (5.5)—, also the dialogue as a whole is compatible with (5.5).

Binary experiments. Informally, (5.5) 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'. We already discussed the analogy between carrying out an experiment and asking a question. Carrying out a binary experiment is analogous to asking a yes-no question: 'is α_i the case or not?'.

In this more restricted setup, it follows easily from condition (5.5) that

$$\begin{cases} \text{if } \mathbb{M}, w \models \alpha_i \text{ then } E_i[w] = [\![\alpha_i]\!]^{\mathbb{M}} \\ \text{if } \mathbb{M}, w \not\models \alpha_i \text{ then } E_i[w] = [\![\neg\alpha_i]\!]^{\mathbb{M}} \end{cases} \quad (5.6)$$

This means that we have indeed found a solution to the expressivity problem: $E_i[w]$ is now always definable by an $\mathcal{L}^\alpha([\mathrm{EXP}])$-formula: either by α_i or by $\neg\alpha_i$ (depending on whether $\mathbb{M}, w \models \alpha_i$). We illustrate this solution by showing how it allows us to express an [EXP]-reduction axiom for $P_i(\varphi) \geq k$:

$$\mathbb{M}, w \models [\mathrm{EXP}]\, P_i(\varphi) \geq k$$

$$\text{iff} \quad \mathbb{M}^e, w \models P_i(\varphi) \geq k$$

$$\text{iff} \quad \mu_i^e(w)([\![\varphi]\!]^{\mathbb{M}^e}) \geq k$$

$$\text{iff} \quad \frac{\mu_i(w)([\![\varphi]\!]^{\mathbb{M}^e} \cap E_i[w])}{\mu_i(w)(E_i[w])} \geq k$$

$$\text{iff} \quad \begin{cases} \mathbb{M}, w \models \alpha_i \text{ and } \frac{\mu_i(w)([\![[\mathrm{EXP}]\,\varphi]\!]^{\mathbb{M}} \cap [\![\alpha_i]\!]^{\mathbb{M}})}{\mu_i(w)([\![\alpha_i]\!]^{\mathbb{M}})} \geq k \\ \mathbb{M}, w \not\models \alpha_i \text{ and } \frac{\mu_i(w)([\![[\mathrm{EXP}]\,\varphi]\!]^{\mathbb{M}} \cap [\![\neg\alpha_i]\!]^{\mathbb{M}})}{\mu_i(w)([\![\neg\alpha_i]\!]^{\mathbb{M}})} \geq k \end{cases} \quad (5.6)$$

$$\text{iff} \quad \mathbb{M}, w \models \begin{aligned} &\Big(\alpha_i \wedge P_i([\mathrm{EXP}]\,\varphi \wedge \alpha_i) \geq k P_i(\alpha_i)\Big) \vee \\ &\Big(\neg\alpha_i \wedge P_i([\mathrm{EXP}]\,\varphi \wedge \neg\alpha_i) \geq k P_i(\neg\alpha_i)\Big) \end{aligned}$$

5.1.3 Evaluation

We will now evaluate both solutions to the syntactic expressivity issue that have been proposed in the previous two subsections. First, note that both solutions indeed allow us to explicitly define the sets $E_i[w]$ in the (extended) language, and thus also to express an [EXP]-reduction axiom for $P_i(\varphi) \geq k$:

1. $[\mathrm{EXP}]\, P_i(\varphi) \geq k \quad \leftrightarrow \quad \downarrow x. P_i([\mathrm{EXP}]\,\varphi \wedge \hat{Q}_i x) \geq k P_i(\hat{Q}_i x)$

2. $[\mathrm{EXP}]\, P_i(\varphi) \geq k \quad \leftrightarrow \quad \big(\alpha \wedge P_i([\mathrm{EXP}]\,\varphi \wedge \alpha_i) \geq k P_i(\alpha_i)\big) \vee$
   $\big(\neg\alpha_i \wedge P_i([\mathrm{EXP}]\,\varphi \wedge \neg\alpha_i) \geq k P_i(\neg\alpha_i)\big)$

It is clear that —at least on an intuitive level— there is a structural analogy between both reduction axioms. First of all, note that $\hat{Q}_i x$ and $(\neg)\alpha_i$ play similar roles, viz. defining the current experiment cell. That these axioms truly talk about the *current* experiment cell is ensured in the first solution by means of the binder $\downarrow$, whereas in the second solution it is ensured by making a case distinction (about which is the current experiment cell) which is explicitly represented in the reduction axiom.

Solution 1. The first solution involves introducing the Q_i-operator. This is not a major disadvantage, since it is merely an extension of the analogy —that was already discussed above— between our approach to experiments and the dynamic epistemic logic of questions (this operator is also introduced in [45]). The major advantage of the first solution is that it places no restrictions on the generality of the experiments: experiments continue to correspond to *any* kind of partition, no matter how fine/coarse (i.e. no restriction to binary partitions only).

However, the first solution also has some serious problems. *Technically* speaking, the move to hybrid logic (and in particular, the introduction of the binder ↓) negatively influences the metaproperties of the resulting system. For example, adding 'merely' the binder ↓ already leads to the undecidability of the satisfiability problem [9, Theorem 5.1].

Furthermore, the first solution has some broad *methodological* issues as well. Hybrid logic provides us with nice technical tools to ensure the definability of certain sets, but by themselves, these tools seem to lack any intuitive interpretation. So far, all operators of the object language had relatively straightforward intuitive meanings (knowledge before/after the experiment, subjective probability, etc.), but this is not the case for the binder ↓ and state variables x. Furthermore, note that hybrid logic can be seen as *yet another* conceptual toolbox, next to the ones of basic modal logic and probabilistic reasoning that are already being used. From this perspective, the system as a whole tends to become rather chaotic: it starts to look like a 'patchwork' that consists of several bits and pieces —each needed to solve a particular problem—, but that lacks overall coherence.

Solution 2. The second solution does not have any of these problems. *Technically* speaking, the metatheoretical properties are not hurt by introducing the special proposition letters α_i (more substance will be given to this claim in Section 5.4). *Methodologically* speaking, the second approach has a clear intuitive interpretation: agent i performs a binary experiment, i.e. she asks the yes-no question 'is α_i the case?' (cf. *supra*). As such, the second solution is based on a particular case of the general analogy between experiments and questions that was already discussed earlier, and that was already being used in the previous chapters.[4] The second solution is thus not a threat to the 'unity' of the system as a whole: it leaves the system as coherent as it was before.

Because of these remarks, the second solution seems to be preferable. One might object at this point that restricting to only binary experiments is too drastic. We have three replies to this objection.

First of all, we reiterate the remark with which we started Section 5.1: until now, our modeling of the experiments has been fully general. All the agreement

---

[4]Obviously, moving from arbitrary to binary experiments is a severe *restriction* of generality. Note, however, that this restriction does not hurt the *intuitive interpretation*. We restrict from arbitrary experiments (arbitrary questions) to binary experiments (yes/no questions), but the intuitive interpretation (the analogy experiment/question) remains intact.

theorems in Chapter 4 were proved in this general context. Restricting to binary experiments is only necessary when one decides to focus on syntactic issues.

Second, it would technically not be difficult to allow for ternary or quaternary experiments. All the conceptual issues, however, arise already at the level of binary experiments, and therefore we have chosen to stick to binary experiments.

Third, in the beginning of this thesis we noted that because the experiment relation can be arbitrarily fine-grained, it can also be used to model *sequences* of experiments (cf. Remark 13). This perspective can be reversed. The experiment relation is now only binary, but every finitary experiment can be represented as a finite sequence of binary experiments.[5] Of course, this requires a way of formally representing sequences of experiments. We will return to this in the final chapter.

Conclusion. We conclude this subsection by reiterating that of the two solutions to the syntactic expressivity issue that were proposed in Subsections 5.1.1 and 5.1.2, the second one is to be preferred on both technical and methodological grounds. Furthermore, the shortcomings of the second approach are not so bad as they might first look. Therefore we will henceforth fully and uniformly adopt the second solution (binary experiments).

5.2   Characterization results

In Chapter 4 we established various dynamic agreement theorems. These theorems required imposing two conditions on probabilistic Kripke models/frames. We will now show that (the global frame versions of) these conditions can be characterized by means of $\mathcal{L}$-formulas.

We first characterize the common prior assumption, i.e. condition 1 of Theorems 35 and 41. If φ is a 1-probability formula, then $\varphi[P_2/P_1]$ is the formula that is obtained by uniformly substituting P_2 for P_1 in φ. In particular, if φ is $\sum_{\ell=1}^n a_\ell P_1(\varphi_\ell) \geq k$, then $\varphi[P_2/P_1]$ is $\sum_{\ell=1}^n a_\ell P_2(\varphi_\ell) \geq k$. It is clear that if φ is a 1-probability formula, then $\varphi[P_2/P_1]$ is a 2-probability formula.

Lemma 42. 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 $\mu_1 = \mu_2$ iff for all atomic 1-probability formulas φ: $\mathbb{F} \models \varphi \leftrightarrow \varphi[P_2/P_1]$.[6]

Proof. We begin by noting that another, 'more obvious' attempt to characterize $\mu_1 = \mu_2$ does not work. This attempt tries to characterize $\mu_1 = \mu_2$ by means of the (single) formula $P_1(p) = P_2(p)$. However, this formula does not belong to the language: recall that different agent indices are not allowed in one and the same probability formula.

---

[5]For a k-ary experiment with $k \geq 2$, let n be the least natural number such that $k \leq 2^n$; then the k-ary experiment can be represented as a sequence of at most n binary experiments.
[6]Recall that an i-probability formula φ is said to be *atomic* iff it is of the form $\sum_{\ell=1}^n a_\ell P_i(p_\ell) \geq k$; cf. Definition 4.

We now prove the characterization result itself. The $\Rightarrow$-direction is trivial. We now prove the $\Leftarrow$-direction by contraposition. Suppose that $\mu_1 \neq \mu_2$. Hence there exist states $w, v \in W$ such that $\mu_1(w)(v) \neq \mu_2(w)(v)$. Assume that $\mu_1(w)(v) < \mu_2(w)(v)$ (the case that $\mu_1(w)(v) > \mu_2(w)(v)$ is analogous). Because $\mathbb{Q}$ 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 by putting $V(p) := \{v\}$. It now easily follows that $\langle \mathbb{F}, V \rangle, w \models P_2(p) \geq k$ and that $\langle \mathbb{F}, V \rangle, w \not\models P_1(p) \geq k$. Hence $\langle \mathbb{F}, V \rangle, w \not\models P_1(p) \geq k \leftrightarrow P_2(p) \geq k$. $\qquad\square$

We now characterize condition 2 of Theorems 35 and 41: if $(w, v) \in R^*$ then $\mu_i(w) = \mu_i(v)$.

Lemma 43. 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:

$$\text{for all } w, v \in W : \text{if } (w, v) \in R^* \text{ then } \mu_i(w) = \mu_i(v)$$
$$\text{iff}$$
$$\text{for all atomic } i\text{-probability formulas } \varphi : \mathbb{F} \models \big(\varphi \to C\varphi\big) \wedge \big(\neg\varphi \to C\neg\varphi\big)$$

Proof. The $\Rightarrow$-direction is trivial. We now prove the $\Leftarrow$-direction by contraposition. Suppose that there are states $w, v \in W$ such that $(w, v) \in R^*$ and yet $\mu_i(w) \neq \mu_i(v)$. Hence there is a state $x \in W$ such that $\mu_i(w)(x) \neq \mu_i(v)(x)$. Now define a valuation V such that $V(p) := \{x\}$. Since $\mu_i(w)(x) \neq \mu_i(v)(x)$, we know that one of the following two cases obtains.

First case: $\mu_i(w)(x) > \mu_i(v)(x)$. Because $\mathbb{Q}$ is dense in $\mathbb{R}$, there exists a $k \in \mathbb{Q}$ such that $\mu_i(w)(x) > k > \mu_i(v)(x)$. It now easily follows that $\langle \mathbb{F}, V \rangle, w \models P_i(p) \geq k$ and $\langle \mathbb{F}, V \rangle, v \not\models P_i(p) \geq k$, and thus $\langle \mathbb{F}, V \rangle, w \not\models C(P_i(p) \geq k)$. Hence $\langle \mathbb{F}, V \rangle, w \not\models P_1(p) \geq k \to C(P_1(p) \geq k)$.

Second case: analogous to the first case. $\qquad\square$

We have now established a characterization result for condition 2 of Theorems 35 and 41. This condition involves the reflexive transitive closure of R, and might therefore be called 'semi-global'. This aspect is reflected in the characterization result, 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 (recall that $E\varphi := K_1\varphi \wedge K_2\varphi$). Formally, this means that Lemma 43 can be improved as follows:

Lemma 44. 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:

$$\text{for all } w, v \in W : \text{if } (w, v) \in R^* \text{ then } \mu_i(w) = \mu_i(v)$$
$$\text{iff}$$
$$\text{for all atomic } i\text{-probability formulas } \varphi : \mathbb{F} \models \big(\varphi \to E\varphi\big) \wedge \big(\neg\varphi \to E\neg\varphi\big)$$

Proof. Again, the $\Rightarrow$-direction is trivial. The $\Leftarrow$-direction will this time be proved directly, so not by contraposition. Assume that $\mathbb{F} \models \big(\varphi \to E\varphi\big) \wedge \big(\neg\varphi \to$

$E\neg\varphi$) for all i-probability formulas φ, and call this assumption (†). We now prove for all $w, v \in W$ 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 $\mu_i(w) = \mu_i(v)$. We prove this by induction on n.

Base case: $n = 0$. For any $w, v \in W$ such that $(w, v) \in R^0$, we have that $w = v$, and thus trivially $\mu_i(w) = \mu_i(v)$. *Induction case:* $n \to n + 1$. Let $w, v \in W$ arbitrary and assume that $(w, v) \in R^{n+1}$. Hence there exists a $u \in W$ 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 $\mu_i(w) = \mu_i(u)$. We claim that also $\mu_i(u) = \mu_i(v)$, and hence it follows that $\mu_i(w) = \mu_i(v)$, as desired.

We now prove the claim that $\mu_i(u) = \mu_i(v)$. For a reductio, suppose that $\mu_i(u) \neq \mu_i(v)$. Therefore there exists a state $x \in W$ such that $\mu_i(u)(x) \neq \mu_i(v)(x)$. Now define a valuation V such that $V(p) := \{x\}$. Since $\mu_i(u)(x) \neq \mu_i(v)(x)$, we know that one of the following two cases obtains.

First case: $\mu_i(u)(x) > \mu_i(v)(x)$. Because $\mathbb{Q}$ is dense in $\mathbb{R}$, there exists a $k \in \mathbb{Q}$ such that $\mu_i(u)(x) > k > \mu_i(v)(x)$. It now easily follows that $\langle \mathbb{F}, V \rangle, u \models P_i(p) \geq k$ and $\langle \mathbb{F}, V \rangle, v \not\models P_i(p) \geq k$, and thus $\langle \mathbb{F}, V \rangle, u \not\models E(P_i(p) \geq k)$. Hence $\langle \mathbb{F}, V \rangle, u \not\models P_i(p) \geq k \to E(P_i(p) \geq k)$, which contradicts our assumption (†).

Second case: analogous to the first case. $\square$

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. In Lemma 43 we characterized it using the common knowledge operator C. In Lemma 44 we showed that this frame condition can also be characterized using the weaker general knowledge operator E. This result is still not fully satisfactory, however: the principle that $\varphi \to E\varphi$ (and also $\neg\varphi \to E\neg\varphi$) for (atomic) i-probability formulas φ requires the 'public availability' of agent i's subjective probabilistic setup. Since $E\varphi = K_1\varphi \wedge K_2\varphi$, this principle can be divided into two components: $\varphi \to K_i\varphi$ and $\varphi \to K_j\varphi$ (where j is the other agent; $j \in \{1, 2\} - \{i\}$). It is quite plausible to assume that each agent knows her *own* probabilistic setup; this is captured by the probabilistic-epistemic introspection principle $\varphi \to K_i\varphi$. The other component, $\varphi \to K_j\varphi$, however, says that agent i's probabilistic setup is automatically known by the other agent, j, which is much less plausible. We now show that in frames which satisfy the common prior assumption ($\mu_1 = \mu_2$), the condition that $\varphi \to E\varphi$ (for atomic i-probability formulas φ) can be weakened to its plausible component $\varphi \to K_i\varphi$, while discarding its implausible component $\varphi \to K_j\varphi$.[7]

---

[7] When examining the proof, it should be clear, however, that the implausible principle $\varphi \to K_j\varphi$ is not really discarded after all. The plausible probabilistic-epistemic introspection principle $\varphi \to K_i\varphi$ and the common prior assumption together *imply* the implausible principe $\varphi \to K_j\varphi$ (for φ an i-plausibility formula and $j \neq i$). From this perspective, Lemma 45 can be seen as an argument against the common prior assumption. Thanks to Eric Pacuit for pressing me on this point.

Lemma 45. 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 :
$$\mathbb{F} \models (\varphi \to K_i \varphi) \wedge (\neg \varphi \to K_i \neg \varphi)$$

Proof. The $\Rightarrow$-direction is again trivial; we prove the $\Leftarrow$-direction. Assume that

$$\mathbb{F} \models (\varphi \to K_i \varphi) \wedge (\neg \varphi \to K_i \neg \varphi) \tag{5.7}$$

for $i = 1, 2$ and for all i-probability formulas φ. Consider an arbitrary agent $i \in \{1, 2\}$ arbitrary; by Lemma 44 it suffices to show that $\mathbb{F} \models (\varphi \to E\varphi) \wedge (\neg \varphi \to E \neg \varphi)$ for all i-probability formulas φ. Let φ be an arbitrary i-probability formula, $V \colon Prop \to \wp(W)$ an arbitrary valuation and $w \in W$ an arbitrary state.

Assume that $\langle \mathbb{F}, V \rangle, w \models \varphi$. By (5.7) it follows that $\langle \mathbb{F}, V \rangle, w \models K_i \varphi$. Let $j \in \{1, 2\} - \{i\}$ be the other agent. Since $\mu_1 = \mu_2$ we have by Lemma 42 that

$$\langle \mathbb{F}, V \rangle \models \varphi \leftrightarrow \varphi[P_j / P_i] \tag{5.8}$$

Since φ is an i-probability formula, it is clear that $\varphi[P_j / P_i]$ is a j-probability formula, so by assumption (5.7) (applied to agent j) it follows that

$$\langle \mathbb{F}, V \rangle, w \models \varphi[P_j / P_i] \to K_j(\varphi[P_j / P_i]) \tag{5.9}$$

From (5.8) it follows that also

$$\langle \mathbb{F}, V \rangle \models K_j \varphi \leftrightarrow K_j(\varphi[P_j / P_i]) \tag{5.10}$$

Furthermore, combining (5.8) with $\langle \mathbb{F}, V \rangle, w \models \varphi$ we obtain $\langle \mathbb{F}, V \rangle, w \models \varphi[P_j / P_i]$. Combining this with (5.9) we get that $\langle \mathbb{F}, V \rangle, w \models K_j(\varphi[P_j / P_i])$. Combining this with (5.10) we get that $\langle \mathbb{F}, V \rangle, w \models K_j \varphi$. Together with $\langle \mathbb{F}, V \rangle, w \models K_i \varphi$ this implies that $\langle \mathbb{F}, V \rangle, w \models E\varphi$. We have now proved that $\langle \mathbb{F}, V \rangle, w \models \varphi \to E\varphi$; the proof for $\langle \mathbb{F}, V \rangle, w \models \neg \varphi \to E \neg \varphi$ is completely analogous. $\square$

Remark 46. We would like to emphasize once more how much we have been able to weaken the original characterization result of Lemma 43. In frames that satisfy the common prior property ($\mu_1 = \mu_2$) —and all the frames that we are interested in when proving agreement theorems indeed satisfy this property— we are able to capture a global frame property (which involves quantifying over R^*) by means of the plausible probabilistic-epistemic introspection principles $\varphi \to K_i \varphi$ and $\neg \varphi \to K_i \neg \varphi$, for i-probability formulas φ. Hence, the notion of common knowledge is *not* needed to characterize this property.

landmark

Remark 47. In Remarks 33, 36 and 39 we noted that the quantification over R^* in the second condition of our agreement theorems can actually be weakened to a quantification over $(R^e)^*$ (or over $\mathcal{R}^*$). In Lemmas 43, 44 and 45 we characterized the second frame property by means of common knowledge (C), general knowledge (E) and individual knowledge (K_i), respectively. It is easy to see that if we replace R^* with $(R^e)^*$, then we immediately obtain analogues of these three lemmas, which will involve post-experimental common knowledge (X), post-experimental general knowledge (F) and post-experimental individual knowledge (R_i).

Note, however, that these post-experimental analogues are conceptually not more plausible than the original lemmas. For concreteness, let's focus on individual knowledge and non-negated i-probability formulas φ. The original lemma involves the principle $\varphi \to K_i\varphi$: agent i knows the probabilities that she assigns to various propositions. The analogous post-experimental lemma involves the principle $\varphi \to R_i\varphi$: after carrying out the experiments, agent i will know the probabilities that she assigned (before the experiment) to various propositions.

Although the second principle ($\varphi \to R_i\varphi$) is technically speaking strictly weaker than the first one ($\varphi \to K_i\varphi$), it is conceptually speaking not more plausible. It would be very strange if an agent would first not know her own probabilistic setup, then carry out an experiment, and then afterwards miraculously come to know it after all. The agent's probabilistic setup and the experiment are conceptually independent; carrying out the experiment should not cause the agent to come to know her probabilistic setup. The only way in which an agent can know her probabilistic setup *after* the experiment is if she knew it already *before* the experiment (and thus has not learned it by means of the experiment). In other words: the principle $\varphi \to R_i\varphi$ is only justified to the extent that $\varphi \to K_i\varphi$ is already justified.

Hence, although replacing R^* with $(R^e)^*$ in the conditions of the agreement theorems would technically speaking be a weakening, it does not yield more plausible principles. This conceptual observation is the reason for working with R^* instead of $(R^e)^*$ in our main presentation of the agreement theorems.

5.3 The logics

In this section we will define various logics that can capture the reasoning behind the agreement theorems. For the sake of clarity, the axiomatizations will be presented in a modular fashion. We will emphasize how our search for axiomatizations is guided by the characterization results established in the previous section. Furthermore, we will discuss the intuitive plausibility of the individual axioms and derive some important theorems of the logics. Soundness and completeness theorems for these logics with respect to various classes of probabilistic Kripke frames will be established in Section 5.4.

The first logic that will be introduced is the basic *probabilistic epistemic logic* PEL, which captures the behavior of the epistemic and probabilistic operators. PEL also axiomatizes the relationship between pre- and post-experimental (com-

mon) knowledge and the fact that the (binary!) experiments always *succeed*, but it does not yet say anything about agreement theorems. We first present the axiomatization of PEL in terms of seven large-scale components.

Componentwise axiomatization of PEL

1. the propositional component
2. the individual knowledge component
3. the common knowledge component
4. the probabilistic component
5. the linear inequalities component
6. the pre-/post-experimental interaction component
7. the α_i-component

We will now define and discuss each of these components individually. The propositional component needs little comment:

---

1. The propositional component

1. all classical propositional tautologies
2. modus ponens: $\varphi \to \psi, \varphi/\psi$

We now turn to the individual knowledge component. We require agent i's (pre-experimental) knowledge operator K_i to satisfy all the usual S5-axioms. The notion of knowledge that is used is thus a very strong one: it has logical omniscience issues (cf. the necessitation rule and the distribution axiom) and satisfies positive and negative introspection. We have two reasons for using this strong notion of knowledge: (1) if one reads K_i as 'implicit' knowledge ('according to agent i's information, ...') —and some logicians claim that this is indeed the right interpretation—, then these issues become much less pressing; (2) Aumann's original result was formulated for an S5-type knowledge operator, and since our goal is to provide a logical formalization of that result, we should use an S5-type knowledge operator as well. Since R_i is merely the post-experimental analogue of K_i, the same remarks also apply to this operator.

---

2. The individual knowledge component

	pre-experimental	post-experimental
necessitation	$\varphi/K_i\varphi$	$\varphi/R_i\varphi$
distribution	$K_i(\varphi \to \psi) \to (K_i\varphi \to K_i\psi)$	$R_i(\varphi \to \psi) \to (R_i\varphi \to R_i\psi)$
factivity	$K_i\varphi \to \varphi$	$R_i\varphi \to \varphi$
pos. introsp.	$K_i\varphi \to K_iK_i\varphi$	$R_i\varphi \to R_iR_i\varphi$
neg. introsp.	$\neg K_i\varphi \to K_i\neg K_i\varphi$	$\neg R_i\varphi \to R_i\neg R_i\varphi$

The axioms for relativized common knowledge ($C^\varphi \psi$) are immediately adapted from [46], where this notion was introduced. The post-experimental version of relativized common knowledge ($X^\varphi \psi$) is governed by the immediate analogues of these axioms.

---

3. The common knowledge component

	pre-experimental	post-experimental
necessit.	$\psi/C^\varphi\psi$	$\psi/X^\varphi\psi$
distrib.	$C^\varphi(\psi \to \chi) \to (C^\varphi\psi \to C^\varphi\chi)$	$X^\varphi(\psi \to \chi) \to (X^\varphi\psi \to X^\varphi\chi)$
mix	$C^\varphi\psi \leftrightarrow E(\varphi \to (\psi \wedge C^\varphi\psi))$	$X^\varphi\psi \leftrightarrow F(\varphi \to (\psi \wedge X^\varphi\psi))$
induction	$C^\varphi(\psi \to E(\varphi \to \psi)) \to$	$X^\varphi(\psi \to F(\varphi \to \psi)) \to$
	$(E(\varphi \to \psi) \to C^\varphi\psi)$	$(F(\varphi \to \psi) \to X^\varphi\psi)$

The probabilistic component consists of two parts. The first part is a straightforward translation into the formal language $\mathcal{L}$ of the well-known Kolmogorov axioms of probability; it ensures that the formal symbol $P_i(\cdot)$ behaves like a real probability function. These axioms are adapted from [14]. The second part consists of the two formulas that characterize properties (i) and (ii) of probabilistic Kripke frames (cf. Definition 1). TIN abbreviates 'truth implies non-negative probability'; Lemma 5 shows that this corresponds to the frame property $\mu_i(w)(w) > 0$ for all $w \in W$. KIC abbreviats 'knowledge implies certainty (i.e. probability 1)'; Lemma 6 shows that this corresponds to the frame property $\mu_i(w)(v) = 0$ for $(w,v) \notin R_i$. This is the first illustration of how our search for logical axiomatizations is guided by the correspondence results established earlier; we will elaborate on this point later.

---

4. The probabilistic component

nonnegativity	$P_i(\varphi) \geq 0$
tautology	$P_i(\top) = 1$
finite additivity	$P_i(\varphi \wedge \psi) + P_i(\varphi \wedge \neg\psi) = P_i(\varphi)$
equivalence	$\varphi \leftrightarrow \psi/P_i(\varphi) = P_i(\psi)$
TIN	$\varphi \to P_i(\varphi) > 0$
KIC	$K_i\varphi \to P_i(\varphi) = 1$

The linear inequalities component axiomatizes (operations on) linear inequalities of probabilities. Finding a solution of a system of linear inequalities requires performing some operations on these inequalities. This component is a technical toolbox which ensures that the logic is strong enough to capture each of these operations, and thus also the solution process as a whole. This feature will be used when proving the completeness theorem in Section 5.4. Finally, note that also these axioms are adapted from [14].

---

5. The linear inequalities component

0-terms	$\sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) \geq k \leftrightarrow \sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) + 0a_{\ell+1} P_i(\varphi_{\ell+1}) \geq k$
permutation	$\sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) \geq k \leftrightarrow \sum_{\ell=1}^{n} a_{p_\ell} P_i(\varphi_\ell) \geq k$
	(for any permutation $p_1, \ldots, p_n$ of $1, \ldots, n$)
addition	$\sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) \geq k \ \wedge \ \sum_{\ell=1}^{n} a'_\ell P_i(\varphi_\ell) \geq k' \rightarrow$
	$\sum_{\ell=1}^{n} (a_\ell + a'_\ell) P_i(\varphi_\ell) \geq (k + k')$
multiplication	$\sum_{\ell=1}^{n} a_\ell P_i(\varphi_\ell) \geq k \leftrightarrow \sum_{\ell=1}^{n} da_\ell P_i(\varphi_\ell) \geq dk$
	(for any $d > 0$)
dichotomy	$\sum_\ell a_\ell P_i(\varphi_\ell) \geq k \vee \sum_\ell a_\ell P_i(\varphi_\ell) < k$
monotonicity	$\sum_\ell a_\ell P_i(\varphi_\ell) \geq k \rightarrow \sum_\ell a_\ell P_i(\varphi_\ell) > k'$ (for all $k' < k$)

The pre-/post-experimental interaction component describes the influence of the experiments on the agents' individual and common knowledge. Basically, it says that carrying out the experiments does not make the agents forget anything that they already knew before the experiments: if agent i knows that φ before the experiment ($K_i\varphi$), then after the experiment she will still know it ($R_i\varphi$). Similar remarks apply to the agents' (relativized) common knowledge. In line with the existing literature on dynamic epistemic logic, we name this property 'perfect recall'.

---

6. The pre-/post-experimental interaction component

perfect recall for individual knowledge	$K_i\varphi \rightarrow R_i\varphi$
perfect recall for (relativized) common knowledge	$C^\varphi \psi \rightarrow X^\varphi \psi$

The final component of PEL 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. 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; we remind the reader that these were only introduced in Section 5.1 to solve an expressivity issue. Finally, this component contains axioms which say that the agents' experiments are *successful*, in the sense that they always lead the agent to the correct answer. Recall the analogy between carrying out a (binary) experiment and asking a yes-no question: 'is α_i the case or not?'. If α_i is the case, then after carrying out her experiment (i.e. after asking the yes-no question to nature), agent i will know that α_i is the case (or actually: was the case before the experiment). In other words, agent i will have obtained the correct answer to her yes-no question. (Likewise if α_i is not the case.)

---

7. The α_i-component

definability of R_i	$R_i\varphi \leftrightarrow \Big(\big(\alpha_i \to K_i(\alpha_i \to \varphi)\big) \wedge \big(\neg\alpha_i \to K_i(\neg\alpha_i \to \varphi)\big) \Big)$
success 1	$\alpha_i \to R_i\alpha_i$
success 2	$\neg\alpha_i \to R_i\neg\alpha_i$

This concludes the presentation of the basic probabilistic epistemic logic PEL. We now introduce the second logic, viz. *probabilistic epistemic agreement logic* or PEAL. Again this logic is presented in a modular fashion: we construct it as a simple extension of PEL.

Componentwise axiomatization of PEAL

1–7.   the seven components of PEL
8.   the agreement component

The agreement component contains three axioms. All of these axioms are directly related to the characterization results of Section 5.2. We first display the axioms, and then discuss their intuitive meaning, and their relation with the agreement results.

---

8. The agreement component

1. $\varphi \leftrightarrow \varphi[P_2/P_1]$ (for 1-probability formulas φ)
2a. $\varphi \to K_i\varphi$ (for i-probability formulas φ)
2b. $\neg\varphi \to K_i\neg\varphi$ (for i-probability formulas φ)

According to Lemma 42, axiom 1 characterizes the common prior assumption, i.e. the first assumption that is made in the formulation of the agreement theorems. We will not discuss the plausibility of this axiom here, since that would merely be adding to the vast literature on this topic [32], whereas our main goal right now is merely to formalize Aumann's original agreement result *as it stands*. Lemma 45 says that axioms 2a and 2b jointly characterize the second condition of the agreement theorems. The intuitive plausibility of these probabilistic-epistemic introspection principles was already extensively discussed in Section 5.2; cf. in particular Remark 46. We conclude this paragraph by emphasizing again how much our search for the agreement (component of the) logic PEAL was guided directly by the characterization results of Section 5.2.

In our axiomatization, we have followed Lemma 45 to capture the second condition of the agreement theorems. Initially, however, this condition was captured by other (stronger) formulas; cf. Lemmas 43 and 44. We will now prove that these other characterizing formulas can be derived as *theorems* in the logical system PEAL.[8]

Lemma 48. Let φ be any i-probability formula. Then:

1. PEAL $\vdash \varphi \to E\varphi$ and PEAL $\vdash \varphi \to C\varphi$

2. PEAL $\vdash \neg\varphi \to E\neg\varphi$ and PEAL $\vdash \neg\varphi \to C\neg\varphi$

Proof. We will focus on item 1; item 2 is proved analogously. Let $j \neq i$ be the other agent. Consider the following (sketch of a) PEAL-derivation:

1.	$\varphi \to K_i\varphi$	(axiom 2a of the agreement component)
2.	$\varphi \leftrightarrow \varphi[P_j/P_i]$	(axiom 1 of the agreement component)
3.	$\varphi[P_j/P_i] \to K_j\varphi[P_j/P_i]$	(axiom 2a of the agreement component)
4.	$K_j\varphi \leftrightarrow K_j\varphi[P_j/P_i]$	(2, S5-properties of K_j)
5.	$\varphi \to K_j\varphi$	(2,3,4, propositional reasoning)
6.	$\varphi \to (K_i\varphi \wedge K_j\varphi)$	(1,5, propositional reasoning)
7.	$\varphi \to E\varphi$	(definition of E)
8.	$\varphi \to E(\top \to \varphi)$	(7, S5 for K_i and K_j, prop. reasoning)
9.	$C^\top(\varphi \to E(\top \to \varphi))$	(8, $C^\top$-necessitation)
10.	$C^\top(\varphi \to E(\top \to \varphi)) \to$ $(E(\top \to \varphi) \to C^\top\varphi)$	(induction axiom for $C^\top\varphi$)
11.	$E(\top \to \varphi) \to C^\top\varphi$	(9,10, propositional reasoning)
12.	$\varphi \to C^\top\varphi$	(8,11, propositional reasoning)
13.	$\varphi \to C\varphi$	(definition of $C\varphi$)

The derivation consisting of steps 1 to 7 shows that PEAL $\vdash \varphi \to E\varphi$; the entire derivation shows that PEAL $\vdash \varphi \to C\varphi$. $\square$

We now introduce the third and final logic, viz. *dynamic probabilistic epistemic agreement logic with explicit experimentation* or DPEALe. As is clear from its name, this logic allows us to explicitly represent the experiment dynamics, but not the communication dynamics. Again, we present it in a modular fashion, as an extension of PEAL:

Componentwise axiomatization of DPEALe

1–8.	the eight components of PEAL
9.	the reduction axioms for [EXP]

---

[8]Obviously, these syntactic derivations will to some extent resemble the semantic proofs of the 'improved' versions of Lemma 43, i.e. Lemmas 44 and 45.

We first display the reduction axioms for [EXP], and then make some remarks about their intuitive interpretation and about the technique of 'pre-encoding'.

---

9. The reduction axioms for [EXP]

1. $\quad\quad\quad\quad [\text{EXP}]\, p \quad\leftrightarrow\quad p \quad\quad\quad\quad\quad$ (for $p \in Prop \cup \{\alpha_1, \alpha_2\}$)
2. $\quad\quad\quad\quad [\text{EXP}]\, \neg\varphi \quad\leftrightarrow\quad \neg\, [\text{EXP}]\, \varphi$
3. $\quad\quad\quad [\text{EXP}](\varphi \wedge \psi) \quad\leftrightarrow\quad [\text{EXP}]\, \varphi \wedge [\text{EXP}]\, \psi$
4. $\quad\quad\quad\quad [\text{EXP}]\, K_i\varphi \quad\leftrightarrow\quad R_i\, [\text{EXP}]\, \varphi$
5. $\quad\quad\quad\quad [\text{EXP}]\, R_i\varphi \quad\leftrightarrow\quad R_i\, [\text{EXP}]\, \varphi$
6. $\quad\quad\quad\quad [\text{EXP}]\, C^\varphi\psi \quad\leftrightarrow\quad X^{[\text{EXP}]\,\varphi}\, [\text{EXP}]\, \psi$
7. $\quad\quad\quad\quad [\text{EXP}]\, X^\varphi\psi \quad\leftrightarrow\quad X^{[\text{EXP}]\,\varphi}\, [\text{EXP}]\, \psi$
8. $\quad\quad\quad [\text{EXP}]\, [\text{EXP}]\, \varphi \quad\leftrightarrow\quad [\text{EXP}]\, \varphi$
9. $\;[\text{EXP}] \sum_\ell a_\ell P_i(\varphi_\ell) \geq k \quad\leftrightarrow$

$$\begin{cases} \quad \alpha_i \to \sum_\ell a_\ell P_i([\text{EXP}]\, \varphi_\ell \wedge \alpha_i) \geq k P_i(\alpha_i) \\ \wedge \quad \neg\alpha_i \to \sum_\ell a_\ell P_i([\text{EXP}]\, \varphi_\ell \wedge \neg\alpha_i) \geq k P_i(\neg\alpha_i) \end{cases}$$

The first reduction axiom says that carrying out the experiments does not change the ontic facts (it only changes the agents' knowledge and their probabilistic setups). Reduction axioms 2 and 3 are as expected; in particular, axiom 2 says that carrying out the experiments is *functional*: it can be done in exactly one way. Axiom 4 says that agent i knows that φ *after* the experiments if and only if she knew *before* the experiments that φ was going to be the case *after* the experiments. This axiom illustrates how R_i allows us to pre-encode the dynamic effects of the [EXP]-operator on the agents' knowledge. Axiom 5 says that with R_i, we have reached an 'expressive equilibrium': to obtain a reduction action for R_i, we do not need to introduce yet another operator. Similar remarks apply to axioms 6 and 7, about (relativized) common knowledge. Axiom 8 says that repeating the experiments does not have any effect (in particular, the agents do not obtain any new information); cf. Lemma 11.

Axiom 9 describes the effects of carrying out the experiments on i-probability formulas. It makes a case distinction on whether or not α_i is the case. The axiom says that agent i adjusts her probabilities according to the Bayesian conditionalization rule: she conditionalizes on the outcome of her experiment, i.e. either on α_i or on $\neg\alpha_i$.

Note that DPEALe does *not* have a rule of [EXP]-necessitation (if DPEALe $\vdash$ φ then DPEALe $\vdash$ [EXP]φ). The reason for this will be discussed in the next section; cf. Remark 59.

5.4 Metatheoretical properties of the logics

In this section we will prove soundness and completeness of the logics defined in the previous section. To guarantee the natural build-up of the main text,

much technical material needed for the proofs has been placed in separate appendices: Appendix A for the material on PEL, Appendix B for the material on PEAL, Appendix C for the material on DPEALe, and finally, Appendix D for the reduction axioms for $[\text{DIAL}(\,\cdot\,)]$.

We begin this section by defining the two classes of frames with respect to which soundness and completeness results will be proved:

Definition 49. We write $\mathcal{PKB}$ for the class of all enriched probabilistic Kripke frames with binary experiments (i.e. satisfying condition (5.4)).

Definition 50. 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 the following two conditions:

1. $\mu_1 = \mu_2$

2. for all $w, v \in W :$ if $(w, v) \in R^*$ then $\mu_i(w) = \mu_i(v)$

We denote the class of all agreement frames with $\mathcal{AGR}$.

Note that the class of agreement frames $\mathcal{AGR}$ can be used to obtain immediate generalizations of the frame-based versions of the agreement theorems established in Chapter 4:

Theorem 51. The following hold:

1. $\mathcal{AGR} \models [\text{EXP}]\, C(P_1(\varphi) = a \wedge P_2(\varphi) = b) \rightarrow a = b$

2. $\mathcal{AGR} \models [\text{EXP}]\, [\text{DIAL}(\varphi)](P_1(\varphi) = a \wedge P_2(\varphi) = b) \rightarrow a = b$

Proof. These follow trivially from Theorems 35 and 41, respectively, since the frames described there are exactly the agreement frames. $\square$

The following two lemmas will be used later, in our discussion of the [EXP]-necessitation rule (cf. Remark 59).

Lemma 52. If $\mathcal{PKB} \models \varphi$ then also $\mathcal{PKB} \models [\text{EXP}]\, \varphi$.

Proof. Suppose that $\mathcal{PKB} \models \varphi$. Consider an arbitrary frame $\mathbb{F} \in \mathcal{PKB}$; we will show that $\mathbb{F} \models [\text{EXP}]\, \varphi$. Let V be an arbitrary valuation on $\mathbb{F}$. By Lemma 9 and the compatibility of binary experiments with the dynamics (cf. Subsection 5.1.2) it follows that also $\langle \mathbb{F}, V \rangle^e \in \mathcal{PKB}$, so $\langle \mathbb{F}, V \rangle^e \models \varphi$, and hence $\langle \mathbb{F}, V \rangle \models [\text{EXP}]\, \varphi$. Since this holds for all valuations V on $\mathbb{F}$, it follows that $\mathbb{F} \models [\text{EXP}]\, \varphi$. $\square$

Lemma 53. There exists a sentence φ such that $\mathcal{AGR} \models \varphi$ and yet $\mathcal{AGR} \not\models [\text{EXP}]\, \varphi$.

Proof. Let φ be the sentence $P_1(p) \geq 0.5 \leftrightarrow P_2(p) \geq 0.5$. We leave it as an easy exercise to check that $\mathcal{AGR} \models \varphi$, and show that $\mathcal{AGR} \not\models [\text{EXP}]\, \varphi$.

Consider the frame $\mathbb{F} := \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2 \rangle$, which is defined by $W = \{w, v\}$, $R_1 = R_2 = E_1 = W \times W$, $E_2 = \{(w, w), (v, v)\}$, and finally

$$\mu_1(w)(w) = \mu_1(w)(v) = \mu_1(v)(w) = \mu_1(v)(v) = 0.5$$
$$\mu_2(w)(w) = \mu_2(w)(v) = \mu_2(v)(w) = \mu_2(v)(v) = 0.5$$

We leave it to the reader to check that $\mathbb{F}$ is indeed an agreement frame. We will now show that $\mathbb{F} \not\models [\text{EXP}]\,\varphi$. Define a valuation V on $\mathbb{F}$ by putting $V(p) := \{v\}$, and define $\mathbb{M} := \langle \mathbb{F}, V \rangle$. Now we compute that

$$\mu_1^e(w)(\llbracket p \rrbracket^{\mathbb{M}^e}) = \frac{\mu_1(w)(\llbracket p \rrbracket^{\mathbb{M}} \cap E_1[w])}{\mu_1(w)(E_1[w])} = \frac{\mu_1(w)(v)}{\mu_1(w)(\{w,v\})} = \frac{0.5}{1} = 0.5$$

and that

$$\mu_2^e(w)(\llbracket p \rrbracket^{\mathbb{M}^e}) = \frac{\mu_2(w)(\llbracket p \rrbracket^{\mathbb{M}} \cap E_2[w])}{\mu_2(w)(E_2[w])} = \frac{\mu_2(w)(\emptyset)}{\mu_2(w)(w)} = \frac{0}{0.5} = 0$$

Hence it follows that

$$\mathbb{M}^e, w \models P_1(p) \geq 0.5 \wedge \neg(P_2(p) \geq 0.5)$$

By the definition of φ it follows that $\mathbb{M}^e, w \not\models \varphi$, so $\mathbb{M}, w \not\models [\text{EXP}]\,\varphi$, and thus $\mathbb{F} \not\models [\text{EXP}]\,\varphi$. $\qquad\square$

Remark 54. The proof above also shows that $\mathcal{AGR}$ is not closed under the experiment update (unlike $\mathcal{PKB}$, cf. Lemma 9). Although this might seem disappointing from a theoretical perspective, there is a very natural explanation for this from the modeling perspective. Recall that $\mu_1 = \mu_2$ is our representation of the agents having a common prior. The proof above shows that $\mu_1 = \mu_2$ does not imply $\mu_1^e \neq \mu_2^e$, i.e. having a *common prior* does not entail having a *common posterior*. Put this way, this is very plausible: to go from the prior to the posterior, each agent processes (by means of Bayesian conditionalization) the information she received by carrying out her experiment, and since the agents carry out different experiments, it is straightforward that they change their priors in different ways, thus obtaining different posteriors.

Remark 55. The proof above also shows that there exist sentences φ, ψ such that $\mathcal{AGR} \models \varphi \leftrightarrow \psi$ and yet $\mathcal{AGR} \not\models [\text{EXP}]\,\varphi \leftrightarrow [\text{EXP}]\,\psi$.

We are now ready to formulate and prove the key results of this section, viz. soundness and completeness for each of the three logical systems defined in the previous section.

Theorem 56. The logic PEL is sound and complete with respect to $\mathcal{PKB}$.

Proof. We refer the reader to Appendix A. $\qquad\square$

Theorem 57. The logic PEAL is sound and complete with respect to $\mathcal{AGR}$.

Proof. We refer the reader to Appendix B. $\qquad\square$

Theorem 58. The logic DPEALe is sound and complete with respect to $\mathcal{AGR}$.

Proof. We refer the reader to Appendix C. □

Remark 59. We have just proved that the logics PEAL and DPEALe are sound with respect to $\mathcal{AGR}$. Using the characterization results of Section 5.2, however, it is easy to see that these logics are *not* sound with respect to the broader class $\mathcal{PKB}$. (For example, on a frame where $\mu_1 \neq \mu_2$, the first axiom of the agreement component will fail; cf. Lemma 42.)

Lemmas 52 and 53 show that for the [EXP]-necessitation rule, the situation is exactly the other way around: this rule is sound with respect to $\mathcal{PKB}$, but not with respect to $\mathcal{AGR}$ (and for good reasons, cf. Remark 54). (Remark 55 implies that even the weaker rule $\varphi \leftrightarrow \psi / [\text{EXP}]\,\varphi \leftrightarrow [\text{EXP}]\,\psi$ is not sound with respect to $\mathcal{AGR}$.)

Hence the combination of DPEALe *and* the [EXP]-necessitation rule is not sound with respect to $\mathcal{AGR}$, nor with respect to $\mathcal{PKB}$. Therefore, we have not added the [EXP]-necessitation rule to our axiomatization of DPEALe in the previous section.

Completely analogous remarks can be made about the necessitation rule (and about the weaker 'equivalence rule') for $[\text{DIAL}(\,\cdot\,)]$.

We are now ready to establish the final technical result of this thesis. In Chapter 4 we established several semantic agreement theorems. We will now show that the first of these theorems (which only represents the experimentation dynamics) is syntactically derivable in the strongest logic defined in Section 5.3.

Theorem 60. $\text{DPEALe} \vdash [\text{EXP}]\,C(P_1(\varphi) = a \wedge P_2(\varphi) = b) \rightarrow a = b$

Proof. This follows from Theorem 51 by the completeness of DPEALe with respect to $\mathcal{AGR}$ (Theorem 58). □

Axiomatizing the experiments *and* the communication? Note that we have not provided a logic that allows us to reason explicitly about [EXP] and $[\text{DIAL}(\,\cdot\,)]$ simultaneously. It is straightforward to add reduction axioms for $[\text{DIAL}(\,\cdot\,)]$ to DPEALe (cf. Appendix D). However, at the moment we are not able to prove this system (call it DPEALec) to be complete with respect to $\mathcal{AGR}$.

The reason for this is the following.[9] Proving completeness of DPEALec would normally go by reducing it to the completeness theorem for PEAL by means of the reduction axioms (cf. the completeness proof in Appendix C). This requires that for any sentence $\varphi \in \mathcal{L}([\text{EXP}], [\text{DIAL}(\,\cdot\,)])$ there exists a sentence $\varphi' \in \mathcal{L}$ such that $\text{DPEALec} \vdash \varphi \leftrightarrow \varphi'$.

However, at the moment we are not able to prove the existence of such a sentence φ' in all cases. The sentence φ might contain a sequence of dynamic operators, e.g. φ is the sentence $[\text{EXP}]\,[\text{DIAL}(\psi_1)]\,[\text{EXP}]\,[\text{DIAL}(\psi_2)]\chi$ (for some $\chi \in \mathcal{L}$). In such cases, one normally works 'inside out' to obtain the sentence

---

[9]It was a remark by Davide Grossi, totally outside of the present context, that focused my attention on this issue.

φ', i.e. one looks at the dynamic operator that has no other dynamic operators in its scope; in our example: one looks at $[\mathrm{DIAL}(\psi_2)]\chi$. One reduces this to an $\mathcal{L}$-sentence θ, such that $\mathsf{DPEALec} \vdash [\mathrm{DIAL}(\psi_2)]\chi \leftrightarrow \theta$. Finally, one invokes the rules $\beta_1 \leftrightarrow \beta_2 / [\mathrm{EXP}]\,\beta_1 \leftrightarrow [\mathrm{EXP}]\,\beta_2$ and $\beta_1 \leftrightarrow \beta_2 / [\mathrm{DIAL}(\varphi)]\beta_1 \leftrightarrow [\mathrm{DIAL}(\varphi)]\beta_2$,[10] and concludes that

$$\mathsf{DPEALec} \vdash [\mathrm{EXP}]\,[\mathrm{DIAL}(\psi_1)]\,[\mathrm{EXP}]\,[\mathrm{DIAL}(\psi_2)]\chi \leftrightarrow [\mathrm{EXP}]\,[\mathrm{DIAL}(\psi_1)]\,[\mathrm{EXP}]\,\theta$$

In this way, one keeps 'peeling off' the dynamic modalities from inside out, and eventually obtains a sentence that contains no more nested dynamic modalities. One more application of the reduction axioms for $[\mathrm{EXP}]$ and $[\mathrm{DIAL}(\,\cdot\,)]$ then allows one to obtain the desired sentence $\varphi' \in \mathcal{L}$.

The problem with the methodology described above is that it relies essentially on the rules $\beta_1 \leftrightarrow \beta_2 / \square\beta_1 \leftrightarrow \square\beta_2$ or $\beta / \square\beta$ ($\square$ abbreviates one of the dynamic modalities, $[\mathrm{EXP}], [\mathrm{DIAL}(\,\cdot\,)]$). Remark 59 implies, however, that these rules are not sound with respect to $\mathcal{AGR}$, and thus they cannot be part of $\mathsf{DPEALec}$.

---

[10]These rules are either part of the axiomatization, or derived from the respective necessitation rules.

Chapter 6

Comparison with Dégremont and Roy's approach

In the previous chapters, we have formalized Aumann's agreement theorem in dynamic epistemic logic (using probabilistic Kripke models), and we have established some important metatheoretical properties of this approach. This is not the first time, however, that the agreement theorem is studied from the perspective of dynamic epistemic logic. The first such study was carried out recently by Dégremont and Roy [13]. They used epistemic plausibility models, rather than probabilistic Kripke models, as a general setting. This choice leads to several differences between both approaches.

In this chapter we provide a detailed comparison between the approach developed in this thesis and the one developed by Dégremont and Roy. Section 6.1 gives an overview of the most important aspects of Dégremont and Roy's approach. In Section 6.2 we discuss some obvious (mainly large-scale) similarities between their approach and ours. In Section 6.3, however, we turn to some important (technical and conceptual) differences between both approaches.

6.1   Overview of Dégremont and Roy's approach

Epistemic plausibility models. Dégremont and Roy (henceforth: D&R) use epistemic plausibility models $\langle W, R_1, R_2, \leq_1, \leq_2, V \rangle$.[1] Here, $\leq_i$ is agent i's plausibility ordering, which is a *qualitative* (non-probabilistic) representation of the relative strength of agent i's doxastic attitudes. In particular, $w \leq_i v$ means that agent i considers state w at least as plausible as v. We require that $\leq_i$ is

---

[1]We refer the reader to [6] for a general introduction to epistemic plausibility models and their dynamics.

a total pre-order on W. As usual, R_i is assumed to be an equivalence relation on W.

An epistemic plausibility model $\mathbb{M}$ is said to satisfy *well-foundedness* iff $\min_{\leq_i}(X)$ is nonempty for every nonempty set $X \subseteq W$ and agent i. It is said to satisfy *common prior* iff $\leq_1 = \leq_2$.

We define $\triangleright_i^\varphi := \{(w,v) \in W \times W \mid v \in \min_{\leq_i}(R_i[w] \cap [\![\varphi]\!]^{\mathbb{M}})\}$. The semantics for (common) knowledge and (common) belief can now be stated as follows:

$$
\begin{array}{lll}
\mathbb{M}, w \models K_i \varphi & \text{iff} & \forall v \in W : \text{if } (w,v) \in R_i \text{ then } \mathbb{M}, v \models \varphi \\
\mathbb{M}, w \models B_i^\varphi \psi & \text{iff} & \forall v \in W : \text{if } (w,v) \in \triangleright_i^\varphi \text{ then } \mathbb{M}, v \models \psi \\
\mathbb{M}, w \models C\varphi & \text{iff} & \forall v \in W : \text{if } (w,v) \in (R_1 \cup R_2)^* \text{ then } \mathbb{M}, v \models \varphi \\
\mathbb{M}, w \models CB\varphi & \text{iff} & \forall v \in W : \text{if } (w,v) \in (\triangleright_1^\top \cup \triangleright_2^\top)^* \text{ then } \mathbb{M}, v \models \varphi
\end{array}
$$

Static agreement theorem. D&R first prove a static agreement theorem. This can be formulated as follows:

> If an epistemic plausibility model $\mathbb{M}$ satisfies well-foundedness, and there exists a state $w \in W$ such that $\mathbb{M}, w \models CB(B_1 p \wedge \neg B_2 p)$, then $\leq_1 \neq \leq_2$.

By contraposition this is equivalent to:

> If an epistemic plausibility model $\mathbb{M}$ satisfies well-foundedness and common prior, then $\mathbb{M} \models \neg CB(B_1 p \wedge \neg B_2 p)$.

Since common knowledge implies common belief, we get as a corollary:

> If an epistemic plausibility model $\mathbb{M}$ satisfies well-foundedness and common prior, then $\mathbb{M} \models \neg C(B_1 p \wedge \neg B_2 p)$.

Neither well-foundedness nor common prior can be characterized in the basic epistemic language considered by D&R. Hence an immediate syntactic version of the static agreement theorem in the basic epistemic language is not possible. Therefore D&R consider a system of hybrid logic that includes the @-operator and the binder $\downarrow$ (cf. Subsection 5.1.1) (the exact details do not matter here). Well-foundedness of $\leq_i$ is captured by Löb's axiom for the (newly introduced) modality $[>_i]$:

$$[>_i]([>_i]p \to p) \to [>_i]p$$

and common prior is captured using the (newly introduced) modality $[\geq_i]$ (recall that $\langle >_i \rangle$ is just the dual of $[>_i]$, i.e. $\langle >_i \rangle := \neg[>_i]\neg$):

$$\langle \geq_1 \rangle p \leftrightarrow \langle \geq_2 \rangle p$$

Let's call the system $\mathcal{S}$. D&R show that

$$\mathcal{S} \vdash \neg C(B_1 p \wedge \neg B_2 p)$$

(but with C, B_1 and B_2 replaced with equivalent, more PDL-style operators).

Dynamic agreement theorem(s). D&R also establish two dynamic agreement results, which focus on the communication dynamics. In the first approach, the communication is modeled by means of conditionalizing (i.e. via conditional beliefs). In the second approach, the communication is modeled by means of public announcements. D&R themselves claim the second approach to be superior, because it better captures the intuitive idea that the dialogue between the agents happens *publicly*. Furthermore, the second approach is also closer to our own formalization of the communication dynamics. Therefore we will focus on this second approach.

D&R's notion of a *dialogue about* φ is very similar to ours (compare with Definition 25). A dialogue about φ is a series of public announcements: at each stage, each agent announces her doxastic attitude towards φ (believing/not believing) at that stage of the dialogue (the technical details do not matter here). It is shown that such sequences always reach a fixed point (compare with Lemma 23). Next, it is shown that the agents' attitudes about φ are common knowledge (compare with Lemma 37). Formally (letting $\mathbb{M}^*$ be this fixed point model):

$$\mathbb{M}^*, w \models \big(B_i\varphi \rightarrow C(B_i\varphi)\big) \wedge \big(\neg B_i\varphi \rightarrow C(\neg B_i\varphi)\big)$$

The dynamic agreement theorem says that after the dialogue about φ (i.e. at the fixed point model $\mathbb{M}^*$), the agents have the same doxastic attitude towards φ (compare with Theorem 38). Formally:

If $\mathbb{M}$ satisfies well-foundedness and common prior, then

$$\mathbb{M}^*, w \models \big(B_1\varphi \wedge B_2\varphi\big) \vee \big(\neg B_1\varphi \wedge \neg B_2\varphi\big)$$

6.2 Similarities

The similarities between D&R's approach and the one developed in this thesis are mostly broadly methodological in character.

(Dynamic) agreement theorems in DEL. Both approaches provide a formalization of Aumann's original agreeing to disagree theorem (and some extensions/refinements) in dynamic epistemic logic (broadly conceived). Furthermore, both approaches have a version of the agreement theorem in which the communication dynamics is made explicit. (Actually, D&R have *two* such versions; cf. the previous section). In both approaches, the communication is modeled as a sequence of public announcements reaching a fixed point.

Goals. The overall goals of both approaches are largely the same: to establish a new link between the game-theoretical and logical perspectives on (common) knowledge and related epistemic notions, and to use the logical formalization of the agreement theorem for conceptual clarification.

Semantic versus syntactic proofs. The results obtained by both approaches are comparable. D&R prove a static agreement theorem, both semantically and syntactically (although rather heavy machinery is needed for the syntactic proof). Furthermore, they also provide a dynamic agreement theorem, for which they give a semantical, but not a syntactic proof. Now compare this with the results obtained in this thesis. We proved only dynamic theorems, no static ones (cf. Section 7.3). Our first theorem (which only represents the experimentation dynamics) was proved semantically (Theorems 32 and 35) and syntactically (Theorem 60). Our second theorem (which represents both the experimentation and the communication dynamics) was proved only semantically (Theorems 38 and 41).

Common knowledge and communication. Both approaches offer the same perspective on the relationship between common knowledge and communication. First consider the approach developed in this thesis. If the communication is not explicitly represented in the theorem, then we need the notion of common knowledge to formulate the agreement theorem (Theorems 32 and 35). However, once the communication is explicitly represented, common knowledge is no longer needed in the formulation of the agreement theorem (Theorems 38 and 41). Common knowledge and communication thus seem to be two sides of the same coin: common knowledge is the result of communication. (Also compare with Remark 46, and with Section 7.2.) In D&R's approach, a very similar picture arises. In the static theorem, they use the notion of common knowledge; in the dynamic theorem, common knowledge is no longer needed in the formulation.

6.3 Differences

We now turn to the differences between both approaches. In Subsection 6.3.1 we discuss 'large-scale' conceptual differences; in Subsection 6.3.2 we highlight some differences regarding technical details.

6.3.1 Conceptual differences

Qualitative versus quantitative. The main conceptual difference between both approaches is, of course, how they represent the agents' *soft* information (beliefs/probabilities). Note that both approaches model the agents' *hard* information (knowledge) in exactly the same way: by means of equivalence relations in a Kripke model.[2] We represent the agents' soft information by means of prob-

---

[2] Note, by the way, that both a probabilistic Kripke model and an epistemic plausibility model can be seen as an ordinary Kripke model to which components (representing the agents' soft information) have been added: to obtain a probabilistic Kripke model, one adds μ_i, and to obtain a plausibility model, one adds $\leq_i$. (Of course, to obtain an (enriched) probabilistic Kripke model, one also needs to add E_i, but since Dégremont and Roy do not represent the experimentation dynamics at all, this is irrelevant.)

abilities, i.e. exact numerical values. D&R represent the agents' soft information by means of a qualitative plausibility ordering.

This qualitative/quantitative difference has many concrete manifestations. The most important one is perhaps the meaning of the notion 'common prior'. In our approach, this means: 'common prior *probability distribution'*, and is thus formalized as follows:

$$\mu_1(w) = \mu_2(w) \text{ (locally)} \qquad \text{or} \qquad \mu_1 = \mu_2 \text{ (globally)}$$

In D&R's approach, 'common prior' means: 'common prior *plausibility ordering'*, and is thus formalized as follows:

$$\leq_1 = \leq_2$$

To some extent, both representations are 'untranslatable' into each other. For example, D&R spend much attention to the assumption that the plausibility orders $\leq_i$ should be *well-founded* (and not merely *locally well-founded*). However, for our probability mass functions $\mu_i(w)$ this problem does not arise, simply because the predicate 'well-founded' is not applicable to these functions.

The probabilistic representation is much more *fine-grained* than the plausibility representation. This is especially clear in the way in which the agents can disagree about a formula φ. In our approach, the agents can disagree in infinitely many ways about φ, viz. by assigning different probabilities to it:

$$\bigvee_{a \neq b \in [0,1]} \big(P_1(\varphi) = a \wedge P_2(\varphi) = b \big)$$

In a plausibility model, however, the agents can disagree only in two different ways about φ:

$$(B_1\varphi \wedge \neg B_2\varphi) \vee (\neg B_1\varphi \wedge B_2\varphi)$$

The agreement theorem says that (if certain conditions are fulfilled) there can be no disagreement. D&R's qualitative theorem thus excludes *two* possibilities (two ways in which the agents can disagree), whereas our (and Aumann's) probabilistic theorem excludes *infinitely many* possibilities (infinitely many ways in which the agents can disagree). Hence the probabilistic version of the theorem seems to be stronger than the qualitative one.

Finally, we remark that Aumann's original result was a probabilistic result. Hence our framework is closer to his original result than the one by D&R. This is clear from the proofs: our proof of Theorem 32 is a straightforward adaptation of Aumann's original proof (cf. Remark 34), whereas D&R's proof (refuting well-foundedness by inductively constructing an infinite descending chain) is conceptually far removed from Aumann's original proof.

Common knowledge versus common belief. All agreement theorems developed in this thesis are based on common *knowledge*. D&R's static agreement theorem, however, is based on common *belief*. (Subsequently they derive a version with common knowledge as a corollary.) Since common belief is a strictly

weaker notion than common knowledge (common knowledge implies common belief, but not vice versa), this means that D&R's agreement theorem is stronger than ours (as they need weaker assumptions).

Furthermore, it does not seem likely that we will be able to establish a 'pure' (cf. infra) common belief-based agreement theorem in our framework. The reason for this is that we work with (probabilistic) Kripke models, and that Kripke models do a bad job at representing beliefs (assuming them to be KD45 modal operators), at least when one also wants to consider dynamic phenomena such as public announcements. In particular, if a formula φ is true but agent i (wrongly) believes it to be false, then after φ has been announced, agent i does not do some sensible form of belief revision, but rather goes crazy and starts believing everything [42]. Formally:

$$(\varphi \wedge B_i \neg \varphi) \rightarrow [!\varphi] B_i \bot \tag{6.1}$$

However, there is a possibility of getting a common belief-based agreement theorem in our framework after all. Remember that the main difference between D&R's approach and ours is the representation of the agents' soft information: D&R represent this qualitatively, whereas we represent it probabilistically. One can establish a link between these two, by *defining* the qualitative notion of belief in terms of probabilities, viz. believing that φ is defined as assigning a sufficiently large probability to φ:

$$B_i \varphi := P_i(\varphi) > \tau \quad \text{for some treshold } \tau \in [0.5; 1]$$

This solves the problem of representing belief and dynamics in Kripke models: (6.1) is now no longer universally true —of course, treating B_i no longer as a primitive KD45 modal operator, but rather as a defined notion. In other words, the following sentence is *not* true on all probabilistic Kripke models:

$$(\varphi \wedge P_i(\neg \varphi) > \tau) \rightarrow [!\varphi] P_i(\bot) > \tau$$

A similar proposal has been made in philosophy, to establish a link between (qualitative) traditional and (probabilistic) Bayesian epistemology. This proposal is sometimes called the *Lockean thesis* [16, 17, 18]. One of its main problems is that if the treshold $\tau \neq 1$, then the resulting belief operator is not closed under conjunction,[3] i.e. the following is not true on all probabilistic Kripke models:

$$\big(P_i(\varphi) > \tau \wedge P_i(\psi) > \tau\big) \rightarrow P_i(\varphi \wedge \psi) > \tau$$

One of the main reasons for still adopting the Lockean thesis is the dynamic behavior it gives rise to. Note that on epistemic plausibility models, the conditional belief operator has the following reduction axiom for public announcements [6, 42]:

$$[!\varphi] B_i^\psi \chi \leftrightarrow \big(\varphi \rightarrow B_i^{\varphi \wedge [!\varphi]\psi} [!\varphi]\chi\big) \tag{6.2}$$

---

[3] A particularly poignant way of formulating this is by means of the so-called *lottery paradoxes* [24].

On the other hand, on probabilistic Kripke models, the 'high conditional belief' operator has the following reduction axiom for public announcements:[4]

$$[!\varphi]P_i(\chi \,|\, \psi) > \tau \leftrightarrow \big(\varphi \rightarrow P_i([!\varphi]\chi \,|\, \varphi \wedge [!\varphi]\psi) > \tau\big) \qquad (6.3)$$

If we adopt the Lockean thesis, then (6.2) and (6.3) are the same sentence (despite being interpreted on two different types of models), i.e. (conditional) belief and (conditional) high probability have the same reduction axiom for public announcement. Following Alexandru Baltag's 'Erlangen program for epistemology' [4], we take this similarity in dynamic behavior to be strong evidence for the claim that belief and high probability are fundamentally the same operator, i.e. for the Lockean thesis.

Just like belief (as a KD45 modal operator) gives rise to the notion of common belief, also the notion of high probability gives rise to a notion of 'common high probability' [31]. If we assume the Lockean thesis (the plausibility of which has just been argued for), this notion of 'common high probability' even *coincides* with the 'classical' notion of common belief. In this way, we can start looking for an agreement theorem based on 'common belief' in our probabilistic Kripkean framework after all.

Static versus dynamic. D&R have one static agreement theorem, and two dynamic ones. Both dynamic agreement theorems focus on the communication dynamics, and leave the experimentation dynamics unexplored. In this thesis, however, we only proved dynamic agreement theorems: one which explicitly represents the experimentation dynamics, and one which explicitly represents both the experimentation and the communication dynamics. In other words, in our approach there is not any static agreement theorem. We will elaborate on this topic in Section 7.3.

6.3.2   Technical differences

Frame characterization results. D&R's static agreement theorem has two conditions: (i) common prior ($\leq_1 \,=\, \leq_2$) and (ii) well-foundedness of $\leq_i$. Neither of these properties is definable in the basic epistemic language (containing knowledge, conditional belief, common knowledge, and common belief operators). D&R therefore introduce new modalities $[>_i]$ and $[\geq_i]$ (for which no intuitive interpretation is given). The semantics for these modalities is as expected ($<_i$ is the strict (irreflexive) order based on $\leq_i$):

$$\mathbb{M}, w \models [>_i]\varphi \quad \text{iff} \quad \forall v \in W : \text{if } v <_i w \text{ then } \mathbb{M}, v \models \varphi$$
$$\mathbb{M}, w \models [\geq_i]\varphi \quad \text{iff} \quad \forall v \in W : \text{if } v \leq_i w \text{ then } \mathbb{M}, v \models \varphi$$

Using these new modalities, the two conditions can be expressed syntactically (cf. supra):

---

[4]We use $P_i(\theta_1 \,|\, \theta_2) > k$ to abbreviate $P_i(\theta_1 \wedge \theta_2) > kP_i(\theta_2)$. Furthermore, in the reduction axiom we assume that all conditional probabilities are well-defined, i.e. that $[!\varphi]P_i(\psi) > 0$ and that $P_i(\varphi \wedge [!\varphi]\psi) > 0$.

well-foundedness	$[>_i]([>_i]p \rightarrow p) \rightarrow [>_i]p$
common prior	$\langle \geq_i \rangle p \leftrightarrow \langle \geq_j \rangle p$

Thus, to define the two conditions of their static agreement theorem, D&R need to extend the language with modal operators for which they do not give an intuitive interpretation. In our approach, however, both conditions of the agreement theorems can be characterized by means of formulas of our basic language $\mathcal{L}$; cf. Section 5.2.

Global versus local conditions. D&R note that the conditions of well-foundedness and common prior are not definable in their basic epistemic language, because these conditions are *global* (to check whether they hold of a given model, one needs to take into account the entire model), whereas the basic epistemic language only contains *local* operators (their semantics is restricted to the agents' current epistemic equivalence class). In our approach, the notion of well-foundedness plays no role at all (cf. supra), so we focus on the common prior assumption.

In D&R's framework, common prior means that $\leq_1 = \leq_2$, which is indeed a global condition. In our approach, however, common prior means either that $\mu_1(w) = \mu_2(w)$, or that $\mu_1 = \mu_2$. The second of these conditions (used in Theorem 35) is also global; the first one (used in Theorems 32 and 38), however, is local.

One might object to this as follows: since $\mu_1(w)$ and $\mu_2(w)$ are (probability mass) functions which have W as their domain, the identity $\mu_1(w) = \mu_2(w)$ is still a global condition:

$$\forall v \in W : \mu_1(w)(v) = \mu_2(w)(v)$$

Note however, that all probabilistic Kripke models satisfy, by definition, the condition that $\mu_i(w)(v) = 0$ for $(w,v) \notin R_i$. Hence

$$\forall v \in W - (R_1[w] \cup R_2[w]) : \mu_1(w)(v) = 0 = \mu_2(w)(v)$$

Hence the condition that $\mu_1(w) = \mu_2(w)$ boils down to:

$$\forall v \in R_1[w] \cup R_2[w] : \mu_1(w)(v) = \mu_2(w)(v)$$

—and this is a genuinely local condition: the quantification of v is restricted to $R_1[w] \cup R_2[w]$.

Logics. To obtain a syntactic proof of their first (static) agreement theorem, D&R introduce a system of hybrid logic, containing the @-operator and the binder $\downarrow$. (It also contains the new modalities $[>_i]$ and $[\geq_i]$.) Our qualms with this move are twofold (cf. Subsection 5.1.1). First, the metatheoretical properties are hurt. Second, and more importantly, these newly added technicalities lack an intuitive interpretation, or an intuitive link with the 'informal' agreement theorem. In a slogan: 'Aumann did not need hybrid logic's binder

$\downarrow$ to formulate the agreement theorem, and therefore we should not use it to formalize this theorem'. Similarly, the modality $[>_i]$ does not have an intuitive interpretation (at least, D&R do not give one). This modality satisfies Löb's axiom; technically, it is an easy exercise to check that this guarantees the well-foundedness of agent i's plausibility order, but again, no intuitive reason is readily available as to why this axiom should hold for this modality.[5]

The logics developed in our framework do not seem to suffer from these problems. Consider the strongest logic developed in this thesis, DPEALe. This was built in three stages. We started from a basic probabilistic epistemic logic, which contains some technical components governing the behavior of probabilities and linear inequalities. The linear inequalities component, for example, is primarily a technical toolbox (to get a completeness theorem), but still, it has an obvious intuitive interpretation. Next, we add the 'agreement axioms', i.e. the sentences that we proved to characterize the conditions of the first agreement theorem. Finally, we dynamified this static base logic by adding a set of reduction axioms for the [EXP]-operator.

All components of this logic thus seem to have an intuitive interpretation, and all its notions (probabilities, (common) knowledge, experiments. . .) seem to be direct formalizations of elements present in Aumann's original agreement theorem, or rather, in the intuitive scenario behind this theorem (cf. Section 2.2). Furthermore, it enjoys some nice metatheorical properties, as was shown in Section 5.4. Still, this logic is strong enough to derive an interesting dynamic agreement theorem (Theorem 60).

---

[5]These remarks are context-dependent, of course. For example, we just claimed that Löb's axiom for the modality $[>_i]$ does not have an immediate intuitive interpretation. In its original context (provability logic), however, Löb's axiom *does* of course have such an immediate intuitive interpretation (viz. it is a modal representation of Löb's theorem for Peano Arithmetic) [12].

Chapter 7

Methodological comments

In this chapter we make some methodological comments about various topics that have arisen in previous chapters. Section 7.1 discusses the perspective on agreeing to disagree via axiomatizations (i.e. the perspective developed in this thesis) and contrasts it with the perspective on agreeing to disagree as theorems in weaker logics. Section 7.2 provides some comments on the role and importance of the notion of common knowledge in agreement results. Section 7.3 is the final and most important section of this chapter; it discusses the issue of static versus dynamic agreement theorems, and argues that the most natural agreement theorems are all dynamic in nature.

Section 7.1 focuses rather heavily on some logical technicalities, and will therefore mainly be of interest to logicians. Sections 7.2 and 7.3 are also based on observations about the logics, but the points they make are more general in nature and will thus also be interesting to game theorists.

7.1 Logics versus theorems

In this section[1] we will elaborate on our methodology of modeling agreement theorems by defining and axiomatizing new logical systems. Semantically speaking, we have proved theorems of the following form (cf. Theorems 32, 35, 38 and 41 in Chapter 4):

> 'if a structure $\mathbb{S}$ satisfies certain conditions $\mathcal{C}$,
> then some agreement result agr-thm is valid/true at $\mathbb{S}$'

The structure $\mathbb{S}$ can be a probabilistic Kripke frame $\mathbb{F}$ or a pointed probabilistic Kripke model $\mathbb{M}, w$. Examples of what the conditions $\mathcal{C}$ can look like can be found in the concrete agreement theorems proved in Chapter 4. Next, agr-thm is a sentence of some object language ($\mathcal{L}$, $\mathcal{L}([\text{EXP}])$, or $\mathcal{L}([\text{EXP}], [\text{DIAL}(\,\cdot\,)])$)

---

[1]Thanks to Johan van Benthem and Eric Pacuit for some very helpful discussions of the material presented in this section.

that formally expresses the impossibility of agreeing to disagree; again, concrete examples of agr-thm can be found in Chapter 4.

Our syntactic approach. In Chapter 5 we moved from the semantic to the syntactic perspective. We proved a theorem of the following form (cf. Theorem 60):

$$\text{AGR-L} \vdash \text{agr-thm}$$

Just like in the semantic version, agr-thm is a sentence of some object language that formally expresses the impossibility of agreeing to disagree; AGR-L is one of the formal logical systems introduced in Section 5.3 —in particular, it is either DPEALe or (given the reduction axioms for the dynamic [EXP]-operator) simply the static agreement logic PEAL.

When comparing the semantic and the syntactic approach, it should be clear that the model/frame conditions $\mathcal{C}$ of the semantic perspective are captured in the syntactic approach by the concrete axiomatization AGR-L. To elaborate on this claim, we reiterate that (the agreement component of) the agreement logics is directly inspired by the characterization results of the frame conditions $\mathcal{C}$ (cf. Sections 5.2 and 5.3).

Another syntactic approach. However, one might want to follow another route when translating the semantic results into a syntactic form. For the sake of concreteness, we will henceforth assume that the conditions $\mathcal{C}$ *coincide* with the common prior assumption $\mu_1 = \mu_2$ (although $\mathcal{C}$ actually comprises another frame condition as well). The other route one might want to take consists in proving theorems of the following form:

$$\text{(D)PEL} \vdash \text{commonprior} \rightarrow \text{agr-thm}$$

Here, (D)PEL is (a dynamic form of) the basic probabilistic epistemic logic introduced in Section 5.3. We emphasize that this logic does not contain any particular agreement component. Again, agr-thm is a sentence of some object language that formally expresses the impossibility of agreeing to disagree; similarly, commonprior is a sentence of some object language that formally expresses the frame conditions $\mathcal{C}$.

The second approach thus does not capture the semantic conditions $\mathcal{C}$ by means of a 'strong' axiomatic system; rather, it works with a 'weak' logic and captures the conditions $\mathcal{C}$ directly in the language (by means of commonprior), as the antecedent of a conditional statement.

The main motivation for adopting this second approach is *theoretical parsimony*. One does not want to develop a new axiom system for every new phenomenon that one is modeling. Rather, one wants to establish one very general logic (PEL, for example) and then derive many interesting probabilistic-epistemic phenomena (such as agreeing to disagree) as theorems of this logic.

Discussion and comparison. The main problem with this second approach, however, is that the sentence commonprior —which should formally express the

common prior assumption— does not exist. Lemma 42 provides a characterization of the common prior assumption ($\mu_1 = \mu_2$) which involves quantifying over all (atomic) 1-probability formulas. In Section 5.3 we add the axiom scheme $\varphi[P_2/P_1] \leftrightarrow \varphi$ (for all 1-probability formulas φ), and thus, we continue (implicitly) to quantify over all 1-probability formulas. The sentence commonprior would thus have to be (where $\mathcal{P}_1 = \{\varphi \mid \varphi$ is a 1-probability formula$\}$):

$$\text{commonprior} = \bigwedge_{\varphi \in \mathcal{P}_1} (\varphi[P_2/P_1] \leftrightarrow \varphi)$$

However, this is an infinite conjunction, and thus not expressible in any of the finitary languages that were introduced in this thesis. Therefore, the sentence commonprior does not exist.

At this point, one might propose to use (some weak version of) infinitary logic to solve this problem.[2,3] Our reply to this proposal is similar to our reply to the use of hybrid logic in Section 5.1: infinitary logic is a technical tool that solves this particular problem, but it hurts the naturality and coherence of the resulting system. (Again, in a slogan: 'one does not need infinite conjunctions to express the original agreement theorem, and therefore one should not use them in the logical formalization of this theorem'.)

Finally, we return to the motivation for the second approach, viz. theoretical parsimony (one does not want to develop a new logic for each phenomenon one wants to formalize). We would like to point out that agreeing to disagree is not a single result, but rather an entire family of theorems. In this thesis, we have formulated and proved basically two agreement theorems: one in which the experiments are represented explicitly, and one in which both the experiments and the communication are represented explicitly. However, many other agreement theorems are possible. The topic of repeated experiments can further be explored, one can look at versions with common belief rather than common knowledge, etc. (We return to these possible extensions in Chapter 8). Therefore, it does seem to make sense to develop an independent *agreement logic* in which all of these agreement-type theorems can be formalized —and thus, the theoretical parsimony argument for the second approach seems to lose some of its force.

7.2 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 [48]. Therefore, it is widely assumed that common knowledge (or at least common belief [13]) plays a central role in agreeing

---

[2]Note that because we only allow rational numbers into the formal languages, there are only countably many 1-probability formulas, and thus the conjunction in commonprior is only countably infinite.

[3]For more background on infinitary (modal) logic we refer the reader to [25, 37].

to disagree results. Throughout this thesis, however, we have made some remarks that imply that the importance of common knowledge is not so central as it is often thought to be.[4]

Common knowledge of the probability distribution. In Aumann's original setup, the (common) probability distribution itself is assumed to be common knowledge among the agents. This is reflected in our framework by the characterization results involving $\varphi \to C\varphi$ (and $\neg\varphi \to C\neg\varphi$) for i-probability formulas φ. However, we showed in Section 5.2 that this can be replaced with the much weaker individual probabilistic-epistemic introspection principle $\varphi \to K_i\varphi$ (for i-probability formulas φ), which no longer contains the common knowledge operator (cf. Remark 46). In other words, the assumption that the agents' probability distributions are common knowledge can be formally captured without making use of the common knowledge operator. This is the *first* way in which common knowledge is less important than is often thought.

Common knowledge and consensus. The *second* way involves the role of common knowledge to obtain consensus (i.e. identical posterior probabilities). Aumann's theorem says that if after carrying out the experiments, the agents have common knowledge of their posteriors, then these posteriors have to be identical (cf. Theorems 32, 35 and 60):

$$[\text{EXP}]\, C(P_1(\varphi) = a \wedge P_2(\varphi) = b) \to a = b$$

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 communication. Once we decide to make this communication dynamics an explicit part of the story (and thus, to explicitly represent it in the formal language), the notion of common knowledge again disappears (cf. Remark 40). The intuitive 'story' now becomes: after carrying out the experiments and carrying out a dialogue about φ, the agents will have identical posteriors for φ. The formal representation now becomes (cf. Theorems 38 and 41):

$$[\text{EXP}]\, [\text{DIAL}(\varphi)](P_1(\varphi) = a \wedge P_2(\varphi) = b) \to a = b$$

Hence, once we decide to represent both the experimentation and the communication dynamics, the agreement theorem can be formulated without making use of the common knowledge operator.

Formulating versus proving. We emphasize that by representing all the dynamics, the agreement theorem can be *formulated* without making use of common knowledge; however, we did not make any claim about *proving* the agreement theorem without making use of common knowledge. To prove the second agreement theorem (Theorem 38), we immediately made use of the fact

---

[4]Thanks to Eric Pacuit for suggesting me to explore this topic.

that after the dialogue about φ, the agents' probabilities for φ have become common knowledge (Lemma 37). Informally: if a person has never heard of the notion of common knowledge, then we can perfectly explain to her what the agreement theorem *says*, but we cannot explain our *proof* to her.

It might be thought that because common knowledge is still used in the proof of the agreement theorem, our methodological point about the importance of this notion is rather weak. However, this situation (a notion that is needed in a theorem's proof, but not in the theorem's formulation) is certainly not trivial. Consider, for example, Fermat's last theorem: the formulation of this theorem requires only very basic mathematical notions and can be understood by anyone with a high school knowledge of mathematics, but its proof involves several advanced mathematical notions and techniques, and is fully understood only by a few experts worldwide.

Finally, we remark that our comments on the relative unimportance of common knowledge for agreeing to disagree results are in line with results by Parikh and coauthors [34, 35]. 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 *consensus* (identical posteriors), but not common knowledge.

7.3 Static versus dynamic agreement theorems

Aumann's original agreeing to disagree theorem was a static result (cf. Chapter 2). Furthermore, the first formalization of this theorem in dynamic epistemic logic by Dégremont and Roy contains two dynamic agreement theorems, but still also one fully static agreement theorem (cf. Chapter 6). In this thesis, we have proved basically two agreement theorems: one in which the experiments are explicitly represented, and one in which both the experiments and the communication are explicitly represented. Hence, all of our agreement theorems are dynamic; we do not have any static agreement theorem at all.

However, the absence of a static (and thus 'classical') agreement theorem is not a disadvantage of our framework. Once one has taken the dynamic turn, it even seems that the only static agreement theorems are rather convoluted. The models that they talk about are chimæras: one such model seems to be composed of 'pieces' taken from many different 'normal' models.

Illustration. To illustrate this, we focus on the experimentation dynamics. In our approach, we have two clean, 'temporally uniform' models (cf. Chapter 3). First, the model $\mathbb{M} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V \rangle$ represents both the agents' knowledge and their probabilities *before* the experiments. Next, the model $\mathbb{M}^e = \langle W^e, R_1^e, R_2^e, E_1^e, E_2^e, \mu_1^e, \mu_2^e, V^e \rangle$ (which can be obtained out of $\mathbb{M}$ in a well-defined way) represents both the agents' knowledge and their probabilities *after* the experiments.

Now contrast this with Aumann's original agreement theorem [2]. This talks about 'temporally incoherent' models, which represent the agents' knowl-

edge *after* the experiments, but their probability distributions *before* the experiments. Formally speaking, such a chimæric model would be of the form $\langle W, R_1^e, R_2^e, E_1, E_2, \mu_1, \mu_2, V \rangle$; i.e. it is obtained by cutting the (temporally uniform) models $\mathbb{M}$ and $\mathbb{M}^e$ into pieces and then pasting these different pieces back together in a temporally incoherent way.

Analyis. This example can be analyzed as follows. The intuitive agreeing to disagree scenario is *intrinsically dynamic*: the diagram in Section 2.2 contains three consecutive times, each of which occurs after some dynamic process occurred. If one wants to prove a static agreement result (like Aumann), then one will need to 'smuggle' this dynamics in somehow —for example, by collapsing times 0, 1 and 2 of the diagram into one single model.

Dégremont and Roy provide dynamic agreement theorems which explicitly represent the communication dynamics, but not the experimentation dynamics. Hence they separate times 1 and 2 of the diagram, but they still collapse times 0 and 1.

In our approach, however, we represent all of the dynamics explicitly in the theorems, and thus we leave times 0, 1 and 2 conceptually separated (no collapse at all). From this perspective, if an epistemic plausibility model $\mathbb{M}$ corresponds to time 0, then $\mathbb{M}^e$ corresponds to time 1, and $(\mathbb{M}^e)^{dial_w(\varphi)}$ corresponds to time 2. Hence, there exists a complete structural analogy between the intuitive scenario on the one hand and the formal theorem on the other.

Conclusion. We have shown that static agreement theorems (such as Aumann's original one) are only possible at the cost of a convoluted notion of model that collapses some crucial distinctions of the intuitive scenario, and that our dynamic approach gives rise to formal theorems which are very close (by means of the structural analogy described above) to the intuitive scenario. The 'dynamification' of the original agreement theorem is thus not merely a manifestation of a particular research program such as dynamic epistemic logic; rather, it arises out of a general desire for conceptual clarity.

Chapter 8

Conclusion

This final chapter consists of two sections. In Section 8.1 we provide an overview of the results obtained in this thesis. The term 'result' is meant to include both formal-logical theorems and philosophical/methodological theses. On the basis of this overview we will also assess to what extent the four research goals formulated in Section 1.1 have been achieved. In Section 8.2 we collect and discuss some interesting open questions that arose throughout this thesis, and that will be the subject of future research.

8.1 Results of the thesis

A new link between epistemic logic and game theory. In this thesis we have formulated and proved various agreement theorems in probabilistic dynamic epistemic logic. In particular, we established model- and frame-based versions of an agreement theorem with experimentation (Theorems 32 and 35), and of an agreement theorem with experimentation and communication (Theorems 38 and 41). We developed a logical system within which the first agreement result is derivable syntactically (Theorem 60).

Throughout the thesis, we have emphasized our attempts to keep the models and the logics intuitively plausible, and directly connected with Aumann's original agreement result (see, in particular, Sections 5.1 and 7.1, and Subsection 6.3.2). The first goal of the thesis (establishing a new connection between epistemic logic and game theory) has thus certainly been achieved.

Further development of PDEL. To obtain the agreement theorems, we introduced (enriched) probabilistic Kripke models and various ways of updating them in Chapter 3. We showed that the assumption that $\mu_i(w)(w) > 0$ is rather modest, but that it is able to solve (in interaction with the precondition account of public announcements) basically all problems related to the case $\mu_i(w)(X) = 0$ in the various update rules. In particular, we claim that this solution is a leap forward in comparison with Kooi's [26] proposal to define

$\mu_i^{\varphi}(w) := \mu_i(w)$ if $\mu_i(w)(\llbracket\varphi\rrbracket^{\mathbb{M}}) = 0$.[1] By introducing the experiment relations E_i, we established a link between probabilistic epistemic logic and the dynamic epistemic logic of questions. We spent quite some time developing the notion of a *dialogue about* φ, as a particular sequence of public announcements that reaches a fixed point after finitely many steps.

Syntactically speaking, we provided frame characterization results for the conditions of the agreement theorems in Section 5.2. In Section 5.3 we used the characterization results as a guideline to develop agreement logics, and in Section 5.4 these logics were shown to be sound and complete.

Hence, the second goal (the further technical development of probabilistic dynamic epistemic logic) has certainly been achieved. Two main technical developments (the introduction of a 'common high probability' operator, and the development of a logic for experiments and communication) were left for future research, but the other ones (in particular, the notion of a dialogue about φ) have been explored fairly extensively in this thesis.

Conceptual clarification of the agreement theorem. On the basis of the technical results established in this thesis, we have also been able to make some methodological remarks about important concepts surrounding the agreement theorem.

In Section 7.2 we discussed the role of common knowledge in the agreement theorem, and claimed that it is not so important as is often thought. Furthermore, we emphasized that common knowledge and communication seem to be two sides of the same coin: common knowledge is the result of communication, so if the communication is explicitly represented in the agreement theorem, there is no need anymore to *assume* common knowledge (as this will now *follow* from the communication). (This also holds in Dégremont and Roy's framework; cf. Section 6.2.)

In Chapter 2 we discussed the original agreement result, and showed that this is a static theorem, while the intuitive motivation behind it contains a lot of dynamics. In Section 7.3 we then argued that representing this dynamics is essential to obtain a natural agreement result, and that static agreement theorems (including Aumann's original result) are only possible at the expense of a convoluted notion of model.

We claim that these methodological remarks offer a real clarification of the agreement theorem. Therefore, the third goal of this thesis (conceptual clarification of the agreement theorem) has also been achieved.

Comparison with Dégremont and Roy. In Chapter 6 we provided a detailed comparison between our approach and the one developed by Dégremont and Roy. We noted some large-scale similarities, but also emphasized several technical and conceptual differences between both approaches. On the basis of

---

[1]This gives rise to very counterintuitive results. On Kooi's account, it is for example possible that after the announcement of φ, agent i knows that φ, and simultaneously assigns probability 0 to φ being true...

this comparison, we believe that our approach is to be preferred to the one by Dégremont and Roy.

8.2   Further research

We now discuss some open questions that arose throughout the thesis, and that are left as subjects for further research.

Infinite probabilistic Kripke models.   Note that all models used in this thesis are *finite*. To obtain greater generality, we would like to extend the notion of probabilistic Kripke model to models with infinite domains. This will bring about several difficulties. For example, our mathematically naïve idea of assigning a probability mass function to each state becomes less plausible. Perhaps the move to infinite domains should be accompanied with a move to a more mathematically mature approach to probability (involving σ-algebras, etc., cf. [36]). Also Lemma 23 will become less trivial, because the proof of this lemma heavily relied on the finiteness of probabilistic Kripke models (cf. Remark 24).

A logic for experiments and communication.   We have not defined a logic which talks simultaneously about [EXP] and [DIAL$(\cdot)$]. The reason for this has been discussed extensively in Section 5.4. However, it is still desirable to have such a logic, to obtain a syntactic version of the second agreement theorem (just like Theorem 60 provides a syntactic version of the first agreement theorem).

Repeated experiments.   Recall that we first allowed arbitrary experiments, and since carrying out an experiment does not change the structure of the experiments (i.e. no Moore-type phenomena take place), it is possible to model an entire sequence of experiments as one big (fine-grained) experiment; cf. Remarks 12 and 13. Hence there is no reason to provide an explicit way of modeling repeated experiments: our formalization of single experiments already suffices.

However, in Section 5.1 we decided to restrict to binary experiments only. Hence the previous remarks no longer apply, and it does become interesting to look for explicit representations of repeated experiments. An additional motivation for doing this is that repeated experiments would allow us to circumvent the restrictions of binary experiments by 'reverse engineering': first we said that any sequence of experiments can be represented as one big experiment; now we say that any experiment (no matter how fine-grained) can be represented as a sequence of binary experiments.

Common high probability   We explained in Subsection 6.3.1 that a 'pure' common belief-based agreement theorem is not possible in our framework, because we make use of Kripke models. Assuming the Lockean thesis (the plausibility of which we have argued for on the basis of Alexandru Baltag's 'Erlangen program for epistemology'), however, we can simulate belief by means of 'high

probability', and common belief by means of 'common high probability'. It will be interesting to explore whether this probabilistically defined notion of common belief gives rise to a new agreement theorem.

Linking probability and plausibility. We argued in Subsection 6.3.1 that the main conceptual difference between Dégremont and Roy's approach and ours is how the agents' soft information is represented: either by means of a plausibility order (qualitatively), or by means of probability functions (quantitatively). However, these two perspectives do not necessarily exclude each other. Combined frameworks have already been developed [5], and it would be interesting to see what kind of agreement theorems can be obtained in such frameworks.

Bibliography

[1] Areces, C., and ten Cate, B. Hybrid logics. In *Handbook of Modal Logic*, P. Blackburn, J. van Benthem, and F. Wolter, Eds. Elsevier, 2007, pp. 821 – 868.

[2] Aumann, R. Agreeing to disagree. *Annals of Statistics 4* (1976), 1236 – 1239.

[3] Bacharach, M. The epistemic structure of a theory of a game. *Theory and Decision 37* (1994), 7 – 48.

[4] Baltag, A. Research profile. Available online at the author's homepage: `http://alexandru.tiddlyspot.com/`.

[5] Baltag, A., and Smets, S. Probabilistic dynamic belief revision. *Synthese 165* (2008), 179 – 202.

[6] Baltag, A., and Smets, S. A qualitative theory of dynamic interactive belief revision. In *Texts in Logic and Games*, G. Bonanno, W. van der Hoek, and M. Woolridge, Eds., vol. 3. Amsterdam University Press, 2008, pp. 9 – 58.

[7] Baltag, A., Smets, S., and Zvesper, J. Keep 'hoping' for rationality: a solution to the backward induction paradox. *Synthese 169* (2009), 301 – 333.

[8] Barwise, J. Three views of common knowledge. In *Proceedings of the 2nd conference on Theoretical Aspects of Reasoning about Knowledge* (1988), M. Vardi, Ed., Morgan Kaufmann, pp. 365 – 379.

[9] Blackburn, P., and Seligman, J. Hybrid languages. *Journal of Logic, Language and Information 4* (1995), 251 – 272.

[10] Bonanno, G. Modal logic and game theory: two alternative approaches. *Risk, Decision and Policy 7* (2002), 309 – 324.

[11] Bonanno, G., and Nehring, K. Agreeing to disagree: a survey. Available online at the first author's homepage: `http://www.econ.ucdavis.edu/faculty/bonanno/wpapers.htm`.

[12] BOOLOS, G. *The Logic of Provability.* Cambridge University Press, 1995.

[13] DÉGREMONT, C., AND ROY, O. Agreement theorems in dynamic-epistemic logic. In *Logic, Rationality, and Interaction. LORI 2009 Proceedings* (2009), X. He, J. Horty, and E. Pacuit, Eds., LNAI 5834, Springer, pp. 105 – 118.

[14] FAGIN, R., AND HALPERN, J. Reasoning about knowledge and probability. *Journal of the ACM 41* (1994), 340 – 367.

[15] FITTING, M. Many-valued modal logics. *Fundamenta Informaticae 15* (1991), 235 – 254.

[16] FOLEY, R. The epistemology of belief and the epistemology of degrees of belief. *American Philosophical Quarterly 29* (1992), 111 – 121.

[17] FOLEY, R. *Working Without a Net.* Oxford University Press, 1993.

[18] FOLEY, R. Beliefs, degrees of beliefs, and the Lockean thesis. In *Degrees of Beliefs.* Springer, 2008, pp. 37 – 47.

[19] GEANAKOPLOS, J. D., AND POLEMARCHAKIS, H. M. We can't disagree forever. *Journal of Economic Theory 28* (1982), 192 – 200.

[20] HALPERN, J. Characterizing the common prior assumption. In *Proceedings of the 7th conference on Theoretical Aspects of Rationality and Knowledge* (1998), I. Gilboa, Ed., Morgan Kaufmann, pp. 133 – 146.

[21] HALPERN, J. Substantive rationality and backward induction. *Games and Economic Behavior 37* (2001), 425 – 435.

[22] HALPERN, J., AND MOSES, Y. Knowledge and common knowledge in a distributed environment. *Journal of the ACM 37* (1990), 549 – 587.

[23] HALPERN, R. F. J., , AND MEGIDDO, N. A logic for reasoning about probabilities. *Information and Computation 87* (1990), 78 – 128.

[24] HENRY E. KYBURG, H. *Probability and the Logic of Rational Belief.* Wesleyan University Press., 1961.

[25] KARP, C. *Languages with Expressions of Infinite Length.* North-Holland Publishing Company, 1964.

[26] KOOI, B. Probabilistic dynamic epistemic logic. *Journal of Logic, Language and Information 12* (2003), 381 – 408.

[27] LEWIS, D. *Convention.* Harvard University Press, 1969.

[28] LISMONT, L., AND MONGIN, P. On the logic of common belief and common knowledge. *Theory and Decision 37* (1994), 75 – 106.

[29] MILGROM, P. An axiomatic characterization of common knowledge. *Econometrica 49* (1981), 219 – 222.

[30] MILGROM, P., AND STOKEY, N. Information, trade and common knowledge. *Journal of Economic Theory 26* (1982), 1327 – 1347.

[31] MONDERER, D., AND SAMET, D. Approximating common knowledge with common beliefs. *Games and Economic Behavior 1* (1989), 170 – 190.

[32] MORRIS, S. The common prior assumption in economic theory. *Economics and Philosophy 11* (1995), 227 – 253.

[33] OSBORNE, M. J., AND RUBINSTEIN, A. *A Course in Game Theory*. MIT Press, 1994.

[34] PARIKH, R., AND KRASUCKI, P. Communication, consensus and knowledge. *Journal of Economic Theory 52* (1990), 178 – 189.

[35] PARIKH, R., AND PACUIT, E. Reasoning about communication graphs. In *Interactive Logic*, J. van Benthem, B. Loewe, and D. Gabbay, Eds. Amsterdam University Press, 2007.

[36] SACK, J. Extending probabilistic dynamic epistemic logic. *Synthese 169* (2009), 241 – 257.

[37] SEGERBERG, K. A model existence theorem in infinitary propositional modal logic. *Journal of Philosophical Logic 23* (1994), 337 – 367.

[38] STALNAKER, R. Belief revision in games: forward and backward induction. *Mathematical Social Sciences 36* (1998), 31 – 56.

[39] VAN BENTHEM, J. *Exploring Logical Dynamics*. CSLI Publications, 1996.

[40] VAN BENTHEM, J. Games in dynamic epistemic logic. *Bulletin of Economic Research 53* (2001), 219 – 248.

[41] VAN BENTHEM, J. Conditional probability meets update logic. *Journal of Logic, Language and Information 12* (2003), 409 – 421.

[42] VAN BENTHEM, J. Dynamic logic for belief revision. *Journal for Applied Non-Classical Logics 17* (2007), 129 – 155.

[43] VAN BENTHEM, J. Rational dynamics and epistemic logic in games. *International Game Theory Review 9* (2007), 13 – 45.

[44] VAN BENTHEM, J., GERBRANDY, J., AND KOOI, B. Dynamic update with probabilities. *Studia Logica 93* (2009), 67 – 96.

[45] VAN BENTHEM, J., AND MINICĂ, S. Toward a dynamic logic of questions. In *Logic, Rationality, and Interaction. LORI 2009 Proceedings* (2009), X. He, J. Horty, and E. Pacuit, Eds., LNAI 5834, Springer, pp. 27 – 41.

[46] van Benthem, J., van Eijck, J., and Kooi, B. Logics of communication and change. *Information and Computation 204* (2006), 1620 – 1662.

[47] van Ditmarsch, H., van der Hoek, W., and Kooi, B. *Dynamic Epistemic Logic*. Springer, 2008.

[48] Vanderschraaf, P., and Sillari, G. Common knowledge (entry in the Stanford Encyclopedia of Philosophy), 2007.

Appendix A

Metatheory of PEL

In this appendix we prove the soundness and completeness of PEL with respect to $\mathcal{PKB}$.

A.1 Soundness

Theorem 57, part 1. PEL is sound with respect to $\mathcal{PKB}$.

Proof. The soundness of the axioms in components 1–5 with respect to (probabilistic) Kripke models is standard; we leave it as an exercise to the reader. (Note that the soundness of the TIN and KIC axioms of the probabilistic component follows from Lemmas 5 and 6, respectively.) We will now establish the soundness of the axioms in components 6 and 7. Consider an arbitrary enriched probabilistic Kripke model (with binary experiments) $\mathbb{M} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V \rangle$ and an arbitrary state $w \in W$. Since $R_i^e = R_i \cap E_i \subseteq R_i$, we get that:

$$
\begin{aligned}
\mathbb{M}, w \models K_i\varphi \quad &\Leftrightarrow \quad \forall v \in W : \text{if } (w,v) \in R_i \text{ then } \mathbb{M}, v \models \varphi \\
&\Rightarrow \quad \forall v \in W : \text{if } (w,v) \in R_i \cap E_i \text{ then } \mathbb{M}, v \models \varphi \\
&\Leftrightarrow \quad \mathbb{M}, w \models R_i\varphi
\end{aligned}
$$

and thus $\mathbb{M}, w \models K_i\varphi \to R_i\varphi$. Completely analogously, it follows that $\mathbb{M}, w \models C^\varphi\psi \to X^\varphi\psi$. These were the axioms of component 6, the pre-/post-experimental interaction component.

We now turn to component 7, the α_i-component. Since $\mathbb{M}$ has binary experiments, we can use (5.6) and get that:

$$
\begin{aligned}
\mathbb{M}, w \models R_i\varphi \quad &\Leftrightarrow \quad \forall v \in W : \text{if } (w,v) \in R_i \cap E_i \text{ then } \mathbb{M}, v \models \varphi \\
&\Leftrightarrow \quad \begin{cases} \mathbb{M}, w \models \alpha_i \Rightarrow \forall v \in W : \big((w,v) \in R_i \Rightarrow (\mathbb{M}, v \models \alpha_i \Rightarrow \mathbb{M}, v \models \psi)\big) \\ \mathbb{M}, w \not\models \alpha_i \Rightarrow \forall v \in W : \big((w,v) \in R_i \Rightarrow (\mathbb{M}, v \not\models \alpha_i \Rightarrow \mathbb{M}, v \models \psi)\big) \end{cases} \\
&\Leftrightarrow \quad \mathbb{M}, w \models \big(\alpha_i \to K_i(\alpha_i \to \varphi)\big) \wedge \big(\neg\alpha_i \to K_i(\neg\alpha_i \to \varphi)\big)
\end{aligned}
$$

and thus $\mathbb{M}, w \models R_i \varphi \leftrightarrow \Big(\big(\alpha_i \rightarrow K_i(\alpha_i \rightarrow \varphi) \big) \wedge \big(\neg\alpha_i \rightarrow K_i(\neg\alpha_i \rightarrow \varphi) \big) \Big)$. It also follows from (5.6) that:

$$\mathbb{M}, w \models \alpha_i \quad \Rightarrow \quad \forall v \in W : \text{if } (w, v) \in R_i \cap E_i \text{ then } \mathbb{M}, v \models \alpha_i$$
$$\Leftrightarrow \quad \mathbb{M}, w \models R_i \alpha_i$$

and thus $\mathbb{M}, w \models \alpha_i \rightarrow R_i \alpha_i$. Completely analogously, it also follows that $\mathbb{M}, w \models \neg\alpha_i \rightarrow R_i \neg\alpha_i$. $\qquad \square$

A.2 Completeness

We will now prove the completeness of PEL with respect to $\mathcal{PKB}$. First we introduce several additional definitions and lemmas, which will be used in the proof of the actual completeness theorem at the end of this section.

Definition 61. Let $\varphi \in \mathcal{L}^\alpha$ be a formula. We define the *closure of φ*, written $Sub^+(\varphi)$, as the smallest set S such that:

- S contains the formulas $\varphi, \alpha_i, R_i \alpha_i, P_i(\alpha_i) > 0$ (for $i = 1, 2$)

- S is closed under subformulas

- if $\psi \in S$ and ψ is not a negation, then $\neg\psi \in S$

- if $K_i \psi \in S$, then $P_i(K_i \psi) = 1 \in S$ (for $i = 1, 2$)

- if $R_i \psi \in S$, then $K_i(\alpha_i \rightarrow \psi), K_i(\neg\alpha_i \rightarrow \psi) \in S$ (for $i = 1, 2$)

- if $C^\psi \chi \in S$, then $K_i(\psi \rightarrow (\chi \wedge C^\psi \chi)) \in S$ (for $i = 1, 2$)

- if $X^\psi \chi \in S$, then $R_i(\psi \rightarrow (\chi \wedge X^\psi \chi)) \in S$ (for $i = 1, 2$)

Remark 62. It follows from Definition 61 that for all $\varphi \in \mathcal{L}^\alpha$, the set $Sub^+(\varphi)$ contains the following formulas: $\alpha_1, \alpha_2, R_1 \alpha_1, R_2 \alpha_2, \neg\alpha_1, \neg\alpha_2, \neg R_1 \alpha_1, \neg R_2 \alpha_2$.

Definition 63. Let $\varphi \in \mathcal{L}^\alpha$ arbitrary. We define $\sim\varphi$ as follows: (i) if φ is a negation, say $\varphi = \neg\psi$, then $\sim\varphi := \psi$, and (ii) if φ is not a negation, then $\sim\varphi := \neg\varphi$.

Lemma 64. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$ and a set Γ, which is a maximal PEL-consistent subset of $Sub^+(\varphi)$. Let $\beta \in Sub^+(\varphi)$ arbitrary. Then $\beta \in \Gamma$ or $\sim\beta \in \Gamma$.

Proof. This is proved by following the case distinction made in Definition 63 of $\sim$; we leave the details to the reader. $\qquad \square$

Lemma 65. For all $\varphi \in \mathcal{L}^\alpha$, the set $Sub^+(\varphi)$ is finite.

Proof. This is proved by induction on the complexity of φ. $\qquad \square$

Definition 66. Let $\varphi \in \mathcal{L}^\alpha$ arbitrary and consider a set $\Gamma \subseteq Sub^+(\varphi)$. (Note that since $Sub^+(\varphi)$ is finite, Γ is finite as well.) We say that Γ is a *maximal* PEL-*consistent subset of* $Sub^+(\varphi)$ iff

- Γ is PEL-consistent, i.e. PEL $\nvdash \neg \bigwedge \Gamma$

- Γ is maximal in $Sub^+(\varphi)$, i.e. there exists no $\Gamma' \subseteq Sub^+(\varphi)$ such that Γ' is PEL-consistent and $\Gamma \subset \Gamma'$

Lemma 67. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$, and let S be the set of maximal PEL-consistent subsets of $Sub^+(\varphi)$. Then S is finite.

Proof. Consider an arbitrary $\Gamma \in S$. Hence $\Gamma \subseteq Sub^+(\varphi)$, so by Lemma 65 it follows that Γ is finite: $|\Gamma| \leq |Sub^+(\varphi)|$. By Lemma 64 we know that for each $\psi \in Sub^+(\varphi)$, either $\psi \in \Gamma$ or $\sim\psi \in \Gamma$ (but not both, because Γ is consistent), and thus $|\Gamma| \geq |Sub^+(\varphi)|$. We conclude that $|\Gamma| = |Sub^+(\varphi)|$, and that Γ consists *exactly* of $(\sim)\psi$ for $\psi \in Sub^+(\varphi)$. Hence there are exactly $2^{|Sub^+(\varphi)|}$ possible sets Γ; i.e. $|S| = 2^{|Sub^+(\varphi)|}$. $\qquad\square$

Definition 68. For any finite $\Gamma \subseteq \mathcal{L}^\alpha$, we abbreviate $\varphi_\Gamma := \bigwedge \Gamma$.

Lemma 69. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$ and a set Γ, which is a maximal PEL-consistent subset of $Sub^+(\varphi)$. Let $\beta \in Sub^+(\varphi)$ arbitrary. If PEL $\vdash \varphi_\Gamma \to \beta$, then $\beta \in \Gamma$.

Proof. Suppose that PEL $\vdash \varphi_\Gamma \to \beta$ and, for a reductio, that $\beta \notin \Gamma$. Then $\Gamma \subset \Gamma \cup \{\beta\}$. Furthermore, since $\Gamma \subseteq Sub^+(\varphi)$ and $\beta \in Sub^+(\varphi)$, also $\Gamma \cup \{\beta\} \subseteq Sub^+(\varphi)$. By the maximality of Γ in $Sub^+(\varphi)$, it follows that $\Gamma \cup \{\beta\}$ is not PEL-consistent, and thus PEL $\vdash \varphi_\Gamma \to \neg\beta$. Together with PEL $\vdash \varphi_\Gamma \to \beta$, this implies that PEL $\vdash \neg\varphi_\Gamma$, which contradicts the PEL-consistency of Γ. $\qquad\square$

Lemma 70. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$ and a PEL-consistent set $\Gamma \subseteq Sub^+(\varphi)$. Then Γ can be extended to a maximal PEL-consistent subset of $Sub^+(\varphi)$.

Proof. This Lindenbaum-type lemma is proved in the standard way. $\qquad\square$

Lemma 71. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$, and consider maximal PEL-consistent sets $\Delta, \Delta' \subseteq Sub^+(\varphi)$. If PEL $\nvdash \varphi_\Delta \to K_i\neg\varphi_{\Delta'}$, then $\{K_i\chi \,|\, K_i\chi \in \Delta\} = \{K_i\chi \,|\, K_i\chi \in \Delta'\}$.

Proof. We prove the contrapositive claim. Suppose that $\{K_i\chi \,|\, K_i\chi \in \Delta\} \neq \{K_i\chi \,|\, K_i\chi \in \Delta'\}$; hence one of the following cases obtains:

- there is a sentence $K_i\chi$ such that $K_i\chi \in \Delta$ and $K_i\chi \notin \Delta'$

  Since $K_i\chi \notin \Delta'$, we get by Lemma 64 that $\neg K_i\chi = \sim K_i\chi \in \Delta'$, and hence PEL $\vdash K_i\chi \to \neg\varphi_{\Delta'}$. By K_i-necessitation and -distribution, it follows that PEL $\vdash K_iK_i\chi \to K_i\neg\varphi_{\Delta'}$. By positive introspection for K_i we get PEL $\vdash K_i\chi \to K_i\neg\varphi_{\Delta'}$. Since $K_i\chi \in \Delta$ we have PEL $\vdash \varphi_\Delta \to K_i\chi$, and thus PEL $\vdash \varphi_\Delta \to K_i\neg\varphi_{\Delta'}$.

- there is a sentence $K_i\chi$ such that $K_i\chi \notin \Delta$ and $K_i\chi \in \Delta'$

  Since $K_i\chi \notin \Delta$, we get by Lemma 64 that $\neg K_i\chi = \sim K_i\chi \in \Delta$, and hence $\mathsf{PEL} \vdash \varphi_\Delta \rightarrow \neg K_i\chi$. By negative introspection for K_i we get $\mathsf{PEL} \vdash \varphi_\Delta \rightarrow K_i\neg K_i\chi$ (†). Since $K_i\chi \in \Delta'$, we have $\mathsf{PEL} \vdash \neg K_i\chi \rightarrow \neg\varphi_{\Delta'}$. By K_i-necessitation and -distribution it follows that $\mathsf{PEL} \vdash K_i\neg K_i\chi \rightarrow K_i\neg\varphi_{\Delta'}$. Combining this with (†), we get that $\mathsf{PEL} \vdash \varphi_\Delta \rightarrow K_i\neg\varphi_{\Delta'}$.

  $\square$

Lemma 72. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$, and let S be the set of maximal PEL-consistent subsets of $Sub^+(\varphi)$. Then $\mathsf{PEL} \vdash \bigvee_{\Gamma \in S} \varphi_\Gamma$.

Proof. This is a standard result. See, for example, [47, Exercise 7.16]. Furthermore, the proof very much resembles the proof of the next lemma, which is given in full detail. $\square$

Lemma 73. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$, an arbitrary formula $\psi \in Sub^+(\varphi)$, and define $S^\psi := \{\Gamma \in S \mid \psi \in \Gamma\}$. (Note that $S^\psi \subseteq S$, and thus, by Lemma 67, S^ψ is finite.) Then $\mathsf{PEL} \vdash \psi \leftrightarrow \bigvee_{\Gamma \in S^\psi} \varphi_\Gamma$.

Proof. First, we show that $\mathsf{PEL} \vdash \bigvee_{\Gamma \in S^\psi} \varphi_\Gamma \rightarrow \psi$. Consider an arbitrary $\Gamma \in S^\psi$. Then by the definitions of S^ψ and φ_Γ it follows that $\mathsf{PEL} \vdash \varphi_\Gamma \rightarrow \psi$. Since this holds for all $\Gamma \in S^\psi$, we get that $\mathsf{PEL} \vdash \bigvee_{\Gamma \in S^\psi} \varphi_\Gamma \rightarrow \psi$.

Now, we prove that $\mathsf{PEL} \vdash \psi \rightarrow \bigvee_{\Gamma \in S^\psi} \varphi_\Gamma$. For a reductio, suppose that $\mathsf{PEL} \nvdash \psi \rightarrow \bigvee_{\Gamma \in S^\psi} \varphi_\Gamma$. Since S^ψ is finite, we can write $S^\psi = \{\Gamma_1, \ldots, \Gamma_n\}$ for some $n \in \mathbb{N}$. Note that each Γ_i is finite (actually $|\Gamma_i| = |Sub^+(\varphi)| =: m$), so we can write $\Gamma_i = \{\gamma_{i,1}, \ldots, \gamma_{i,m}\}$. Assume the following:

$$\mathsf{PEL} \vdash \psi \rightarrow (\gamma_{1,1} \vee \gamma_{2,1} \vee \cdots \vee \gamma_{n,1})$$
$$\mathsf{PEL} \vdash \psi \rightarrow (\gamma_{1,2} \vee \gamma_{2,1} \vee \cdots \vee \gamma_{n,1})$$

$$\vdots$$

$$\mathsf{PEL} \vdash \psi \rightarrow (\gamma_{1,b_1} \vee \gamma_{2,1} \vee \cdots \vee \gamma_{n,1})$$
$$\mathsf{PEL} \vdash \psi \rightarrow (\gamma_{1,1} \vee \gamma_{2,2} \vee \cdots \vee \gamma_{n,1})$$

$$\vdots$$

$$\mathsf{PEL} \vdash \psi \rightarrow (\gamma_{1,b_1} \vee \gamma_{2,b_2} \vee \cdots \vee \gamma_{n,m})$$

(note that this are m^n assumptions). From these assumptions, it follows by propositional reasoning that

$$\mathsf{PEL} \vdash \psi \rightarrow \bigwedge_{\substack{1 \leq j_1 \leq m \\ \vdots \\ 1 \leq j_n \leq m}} (\gamma_{1,j_1} \vee \cdots \vee \gamma_{n,j_n})$$

94

By distributivity this implies that

$$\mathsf{PEL} \vdash \psi \to \bigwedge_{1 \le j_1 \le m} \gamma_{1,j_1} \vee \cdots \vee \bigwedge_{1 \le j_n \le m} \gamma_{n,j_n}$$

Or equivalently: $\mathsf{PEL} \vdash \psi \to \bigvee_{1 \le i \le n} \bigwedge_{1 \le j \le m} \gamma_{i,j}$. By the definitions introduced above, this can be rewritten as $\mathsf{PEL} \vdash \psi \to \bigvee_{\Gamma \in S^\psi} \varphi_\Gamma$. But this contradicts our assumption that $\mathsf{PEL} \nvdash \psi \to \bigvee_{\Gamma \in S^\psi} \varphi_\Gamma$.

Hence one of our m^n assumptions has to be dropped, i.e. there exist $x_1 \in \{1, \ldots, m\}, \ldots, x_n \in \{1, \ldots, m\}$ such that $\mathsf{PEL} \nvdash \psi \to (\gamma_{1,x_1} \vee \cdots \vee \gamma_{n,x_n})$, and thus $X := \{\psi, \neg\gamma_{1,x_1}, \ldots, \neg\gamma_{n,x_n}\}$ is PEL-consistent. Trivially $X \subseteq Sub^+(\varphi)$; and hence, by Lemma 70 X can be extended to a maximal PEL-consistent subset $X^+ \in S$. Note that $\psi \in X \subseteq X^+$, and hence $X^+ \in S^\psi = \{\Gamma_1, \ldots, \Gamma_n\}$.

However, $\gamma_{1,x_1} \in \Gamma_1$, so $\neg\gamma_{1,x_1} \notin \Gamma_1$, and yet $\neg\gamma_{1,x_1} \in X \subseteq X^+$, so $X^+ \ne \Gamma_1$. Similarly, $\gamma_{2,x_2} \in \Gamma_2$, so $\neg\gamma_{2,x_2} \notin \Gamma_2$, and yet $\neg\gamma_{2,x_2} \in X \subseteq X^+$, so $X^+ \ne \Gamma_2$. Etc. Hence we conclude that $X^+ \notin \{\Gamma_1, \ldots, \Gamma_n\}$. Contradiction. $\qquad\square$

Lemma 74. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$, and let S be the set of maximal PEL-consistent subsets of $Sub^+(\varphi)$. If $\Gamma \ne \Delta \in S$, then $\mathsf{PEL} \vdash \varphi_\Gamma \to \neg\varphi_\Delta$.

Proof. Since $\Gamma \ne \Delta$ there exists a sentence ψ such that $\psi \in \Gamma$ and $\psi \notin \Delta$, or $\psi \notin \Gamma$ and $\psi \in \Delta$. We focus on the first case (the second one is analogous). Since $\psi \in \Gamma$ we have $\psi \in Sub^+(\varphi)$ and $\mathsf{PEL} \vdash \varphi_\Gamma \to \psi$ (†). Since $\psi \notin \Delta$, it follows by Lemma 64 that $\sim\psi \in \Delta$, so $\mathsf{PEL} \vdash \varphi_\Delta \to \sim\psi$, and thus $\mathsf{PEL} \vdash \psi \to \neg\varphi_\Delta$. Combining this with (†), we find that $\mathsf{PEL} \vdash \varphi_\Gamma \to \neg\varphi_\Delta$. $\qquad\square$

Lemma 75. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$, and an arbitrary formula $\psi \in Sub^+(\varphi)$. Recall that $S^\psi = \{\Gamma \in S \mid \psi \in \Gamma\}$. It holds that $\mathsf{PEL} \vdash P_i(\psi) = \sum_{\Gamma \in S^\psi} P_i(\varphi_\Gamma)$.

Proof. By Lemma 73 we have that $\mathsf{PEL} \vdash \psi \leftrightarrow \bigvee_{\Gamma \in S^\psi} \varphi_\Gamma$. By the equivalence rule it follows that $\mathsf{PEL} \vdash P_i(\psi) = P_i(\bigvee_{\Gamma \in S^\psi} \varphi_\Gamma)$. By Lemma 74 all the φ_Γ are mutually inconsistent, and hence it follows by finite additivity that $\mathsf{PEL} \vdash P_i(\psi) = \sum_{\Gamma \in S^\psi} P_i(\varphi_\Gamma)$. $\qquad\square$

Lemma 76. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$, and an arbitrary i-probability formula $\psi \in Sub^+(\varphi)$. Recall that S is the set of maximal PEL-consistent subsets of $Sub^+(\varphi)$. Then there exist rational numbers $\{c_\Gamma^\psi\}_{\Gamma \in S}$ and b^ψ such that $\mathsf{PEL} \vdash \psi \leftrightarrow \sum_{\Gamma \in S} c_\Gamma^\psi P_i(\varphi_\Gamma) \ge b^\psi$.

Proof. Suppose that ψ is the sentence $a_1 P_i(\chi_1) + \cdots a_n P_i(\chi_n) \ge b$. Hence χ_k is a subformula of ψ, and thus $\chi_k \in Sub^+(\varphi)$, so by Lemma 75 we get that $\mathsf{PEL} \vdash P_i(\chi_k) = \sum_{\Gamma \in S^{\chi_k}} P_i(\varphi_\Gamma)$, and hence also

$$\mathsf{PEL} \vdash a_k P_i(\chi_k) = a_k \sum_{\Gamma \in S^{\chi_k}} P_i(\varphi_\Gamma)$$

(this holds for $1 \leq k \leq n$). It easily follows that

$$\mathsf{PEL} \vdash \psi \leftrightarrow a_1 \sum_{\Gamma \in S^{\chi_1}} P_i(\varphi_\Gamma) + \cdots + a_n \sum_{\Gamma \in S^{\chi_n}} P_i(\varphi_\Gamma) \geq b$$

Note that for any $\Gamma \in S$, the term $P_i(\varphi_\Gamma)$ occurs (in general) several times in the above formula, namely for each k such that $\Gamma \in S^{\chi_k}$ (with corresponding coefficient a_k). By Lemma 4.1 of Halpern and Fagin [14], these multiple occurrences can be grouped together, with the new coefficient being the sum of all the old corresponding coefficients. Hence, if we define $c_\Gamma^\psi := \sum_{\{\ell \mid \Gamma \in S^{\chi_\ell}\}} a_\ell$ for each $\Gamma \in S$, and $b^\psi := b$, then we get:[1]

$$\mathsf{PEL} \vdash \psi \leftrightarrow \sum_{\Gamma \in S} c_\Gamma^\psi P_i(\varphi_\Gamma) \geq b^\psi$$

$\square$

Definition 77. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$, an agent $i \in \{1, 2\}$, and a maximal PEL-consistent subset $\Gamma \in S$. Note that by Lemma 67, S is finite. The *characteristic system* of i, Γ (abbreviated $\sigma_{i,\Gamma}$) is a system in $|S|$ variables $x_{i,\Gamma,\Delta}$ (for $\Delta \in S$). This system is defined as follows:

- First of all, $\sigma_{i,\Gamma}$ has the (in)equalities displayed below.[2] These (in)equalities will sometimes be called the 'fixed' (in)equalities of $\sigma_{i,\Gamma}$, because they do not depend on any i-probability formula ψ.

$$\sum_{\Delta \in S} x_{i,\Gamma,\Delta} = 1 \tag{A.1}$$

$$x_{i,\Gamma,\Delta} \geq 0 \qquad \text{(for each } \Delta \in S) \tag{A.2}$$

$$x_{i,\Gamma,\Gamma} > 0 \tag{A.3}$$

$$x_{i,\Gamma,\Delta} = 0 \qquad \text{(for each } \Delta \in \overline{R_1[\Gamma]} \cup \overline{R_2[\Gamma]}) \tag{A.4}$$

- By Lemma 65, $Sub^+(\varphi)$ is finite, and thus contains only finitely many i-probability formulas ψ. For each such ψ there exist, by Lemma 76, rational numbers $\{c_\Delta^\psi\}_{\Delta \in S}$ and b^ψ such that $\mathsf{PEL} \vdash \psi \leftrightarrow \sum_{\Delta \in S} c_\Delta^\psi P_i(\varphi_\Delta) \geq b^\psi$. For each of these finitely many probability formulas $\psi \in Sub^+(\varphi)$, note that by Lemma 64 either $\psi \in \Gamma$ or $\neg\psi = \sim\psi \in \Gamma$. For any i-probability formula $\psi \in Sub^+(\varphi)$, if $\psi \in \Gamma$, add the inequality

$$\sum_{\Delta \in S} c_\Delta^\psi x_{i,\Gamma,\Delta} \geq b^\psi$$

---

[1] We agree on the convention that $\sum_\emptyset a_i = 0$.

[2] We define $\overline{R_i[\Gamma]} := \{\Delta \in S \mid \{K_i\chi \mid K_i\chi \in \Gamma\} \neq \{K_i\chi \mid K_i\chi \in \Delta\}\}$. For now, these can be considered as strictly formally defined sets; later, however, it will be clear that these sets are related with agent i's epistemic accessibility relation in the PEL-canonical model, cf. Definition 81.

to the system $\sigma_{i,\Gamma}$, and if $\neg\psi \in \Gamma$, add the inequality

$$\sum_{\Delta \in S} c_\Delta^\psi x_{i,\Gamma,\Delta} < b^\psi$$

to the system $\sigma_{i,\Gamma}$. This finishes the definition of $\sigma_{i,\Gamma}$.

Remark 78. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$, an agent $i \in \{1,2\}$, and a set $\Gamma \in S$. Now consider the characteristic system $\sigma_{i,\Gamma}$. Note that for any inequality χ in the system $\sigma_{i,\Gamma}$ it holds that $\chi[P_i(\varphi_\Delta)/x_{i,\Gamma,\Delta}]$ is a sentence of $\mathcal{L}^\alpha$ (where $\chi[P_i(\varphi_\Delta)/x_{i,\Gamma,\Delta}]$ is the result of uniformly substituting the formula $P_i(\varphi_\Delta)$ for the variable $x_{i,\Gamma,\Delta}$ in χ, for each $\Delta \in S$). By $\bigwedge \sigma_{i,\Gamma}[P_i(\varphi_\Delta)/x_{i,\Gamma,\Delta}]$ we represent the conjunction of all the $\mathcal{L}^\alpha$-sentences that are obtained by carrying out this uniform substitution on inequalities of $\sigma_{i,\Gamma}$.

Lemma 79. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$, an agent $i \in \{1,2\}$ and a set $\Gamma \in S$. Then $\mathsf{PEL} \vdash \varphi_\Gamma \to \bigwedge \sigma_{i,\Gamma}[P_i(\varphi_\Delta)/x_{i,\Gamma,\Delta}]$.

Proof. It suffices to show that $\mathsf{PEL} \vdash \varphi_\Gamma \to \chi[P_i(\varphi_\Delta)/x_{i,\Gamma,\Delta}]$ for each (in)equality χ of the system $\sigma_{i,\Gamma}$. We prove this for all of these inequalities χ separately:

- χ is the 'fixed' equality $\sum_{\Delta \in S} x_{i,\Gamma,\Delta} = 1$

  By Lemma 72 we have $\mathsf{PEL} \vdash \bigvee_{\Delta \in S} \varphi_\Delta$, and thus $\mathsf{PEL} \vdash \bigvee_{\Delta \in S} \varphi_\Delta \leftrightarrow (p \vee \neg p)$ (for any $p \in Prop$). By the tautology axiom and the equivalence rule for probabilities, it follows that $\mathsf{PEL} \vdash P_i(\bigvee_{\Delta \in S} \varphi_\Delta) = 1$. Since S is finite, and all the φ_Δ are mutually inconsistent (cf. Lemma 74), we get by finite additivity that $\mathsf{PEL} \vdash \sum_{\Delta \in S} P_i(\varphi_\Delta) = 1$, and thus certainly $\mathsf{PEL} \vdash \varphi_\Gamma \to \sum_{\Delta \in S} P_i(\varphi_\Delta) = 1$.

- χ is the 'fixed' inequality $x_{i,\Gamma,\Delta} \geq 0$ (for some $\Delta \in S$)

  By the non-negativity axiom for probabilities, we have $\mathsf{PEL} \vdash P_i(\varphi_\Delta) \geq 0$, and thus certainly $\mathsf{PEL} \vdash \varphi_\Gamma \to P_i(\varphi_\Delta) \geq 0$.

- χ is the 'fixed' inequality $x_{i,\Gamma,\Gamma} > 0$

  By the TIN axiom we have $\mathsf{PEL} \vdash \varphi_\Gamma \to P_i(\varphi_\Gamma) > 0$.

- χ is the 'fixed' inequality $x_{i,\Gamma,\Delta} = 0$ (for some $i \in \{1,2\}$ and $\Delta \in \overline{R_i[\Gamma]}$)

  Since $\Delta \notin \overline{R_i[\Gamma]}$ we get that $\{K_i\chi \,|\, K_i\chi \in \Gamma\} \neq \{K_i\chi \,|\, K_i\chi \in \Delta\}$. By Lemma 71 it follows that $\mathsf{PEL} \vdash \varphi_\Gamma \to K_i\neg\varphi_\Delta$. Combining this with the KIC axiom, we get that $\mathsf{PEL} \vdash \varphi_\Gamma \to P_i(\neg\varphi_\Delta) = 1$, so by finite additivity it follows that $\mathsf{PEL} \vdash \varphi_\Gamma \to P_i(\varphi_\Delta) = 0$.

- χ is $\sum_{\Delta \in S} c_\Delta^\psi x_{i,\Gamma,\Delta} \geq b^\psi$ (for some i-probability formula $\psi \in Sub^+(\varphi)$ such that $\psi \in \Gamma$)

  Since $\psi \in \Gamma$ we have $\mathsf{PEL} \vdash \varphi_\Gamma \to \psi$. By Lemma 76 we have $\mathsf{PEL} \vdash \psi \leftrightarrow \sum_{\Delta \in S} c_\Delta^\psi P_i(\varphi_\Delta) \geq b^\psi$, and hence $\mathsf{PEL} \vdash \varphi_\Gamma \to \sum_{\Delta \in S} c_\Delta^\psi P_i(\varphi_\Delta) \geq b^\psi$.

- χ is $\sum_{\Delta \in S} c_\Delta^\psi x_{i,\Gamma,\Delta} < b^\psi$ (for some i-probability formula $\psi \in Sub^+(\varphi)$ such that $\neg\psi \in \Gamma$)

  Since $\neg\psi \in \Gamma$ we have $\vdash \varphi_\Gamma \to \neg\psi$. By Lemma 76 we have $\mathsf{PEL} \vdash \psi \leftrightarrow \sum_{\Delta \in S} c_\Delta^\psi P_i(\varphi_\Delta) \geq b^\psi$, and thus $\mathsf{PEL} \vdash \varphi_\Gamma \to \sum_{\Delta \in S} c_\Delta^\psi P_i(\varphi_\Delta) < b^\psi$.

  $\square$

Lemma 80. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$, an agent $i \in \{1,2\}$ and a set $\Gamma \in S$. Then the characteristic system $\sigma_{i,\Gamma}$ has a solution $\{x^*_{i,\Gamma,\Delta}\}_{\Delta \in S}$.

Proof. For a reductio, suppose that $\sigma_{i,\Gamma}$ has no solution. Hence for every assignment $a\colon \{x_{i,\Gamma,\Delta}\}_{\Delta \in S} \to \mathbb{R}$ it holds that $a \models \neg \bigwedge \sigma_{i,\Gamma}$. Hence $\neg \bigwedge \sigma_{i,\Gamma}$ is a 'valid linear inequality' (in the terminology of [23]). Therefore, $\neg \bigwedge \sigma_{i,\Gamma}[P_i(\varphi_\Delta)/x_{i,\Gamma,\Delta}]$ is an *instance* of a valid linear inequality, and thus $\mathsf{PEL} \vdash \neg \bigwedge \sigma_{i,\Gamma}[P_i(\varphi_\Delta)/x_{i,\Gamma,\Delta}]$ (since the linear inequalities component is complete for instances of valid linear inequalities; cf. Section 4 of [23]). Together with Lemma 79 this implies that $\mathsf{PEL} \vdash \neg\varphi_\Gamma$, which contradicts the PEL-consistency of Γ. $\square$

Definition 81. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$. The PEL-*canonical model for* $Sub^+(\varphi)$ is defined as $\mathbb{M} := \langle S, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V \rangle$, where

- $S := \{\Gamma \subseteq \mathcal{L}^\alpha \,|\, \Gamma$ is a maximal PEL-consistent subset of $Sub^+(\varphi)\}$

- $R_i := \{(\Gamma, \Delta) \in S \times S \,|\, \{K_i\psi \,|\, K_i\psi \in \Gamma\} = \{K_i\psi \,|\, K_i\psi \in \Delta\}\}$

- for any $p \in Prop \cup \{\alpha_1, \alpha_2\}$, put $V(p) := \{\Gamma \in S \,|\, p \in \Gamma\}$

- $E_i := \big(V(\alpha_i) \times V(\alpha_i)\big) \cup \big((S - V(\alpha_i)) \times (S - V(\alpha_i))\big)$

- for any $\Gamma, \Delta \in S$, put $\mu_i(\Gamma)(\Delta) := x^*_{i,\Gamma,\Delta}$ (where $\{x^*_{i,\Gamma,\Delta}\}_{\Delta \in S}$ is a solution to the system $\sigma_{i,\Gamma}$; this solution is guaranteed to exist by Lemma 80)

Lemma 82. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$, and assume that φ is PEL-consistent. Then the PEL-canonical model for $Sub^+(\varphi)$, $\mathbb{M}$, is a probabilistic Kripke model with binary experiments.

Proof. Note the following:

- S is finite (by Lemma 67), and non-empty: since φ is PEL-consistent, Lemma 70 says that the set $\{\varphi\} \subseteq Sub^+(\varphi)$ can be extended to an element $\Gamma \in S$ —and hence $S \neq \emptyset$

- R_1, R_2 are equivalence relations (this is easy to check)

- E_i satisfies condition (5.5) on binary experiments (by definition of E_i)

- for each agent $i \in \{1,2\}$ and state $\Gamma \in S$, it holds that $\mu_i(\Gamma)$ is a probability mass function on S: because $\{\mu_i(\Gamma)(\Delta)\}_{\Delta \in S}$ is (by definition) a solution of the characteristic system $\sigma_{i,\Gamma}$, we have that $\sum_{\Delta \in S} \mu_i(\Gamma)(\Delta) = 1$ and that $\mu_i(\Gamma)(\Delta) \geq 0$ for all $\Delta \in S$

- condition (i) of Definition 1 is fulfilled: for all $\Gamma \in S$, we have that $\mu_i(\Gamma)(\Gamma) > 0$ (since $\mu_i(\Gamma)(\Gamma)$ is a solution to the inequality $x_{i,\Gamma} > 0$ of the system $\sigma_{i,\Gamma}$

- condition (ii) of Definition 1 is fulfilled: for all $\Gamma, \Delta \in S$, if $(\Gamma, \Delta) \notin R_i$ then $\mu_i(\Gamma)(\Delta) = 0$ (since $\mu_i(\Gamma)(\Delta)$ is a solution to the equality $x_{i,\Gamma,\Delta} = 0$ for all $\Delta \notin R_i[\Gamma]$ of the sytem $\sigma_{i,\Gamma}$)

$\square$

Lemma 83. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$, and consider maximal PEL-consistent sets $\Delta, \Delta' \subseteq Sub^+(\varphi)$. Note that by Remark 62 and Lemma 64 it holds that $\alpha_i \in \Delta$ or $\neg\alpha_i \in \Delta$, and similarly $\alpha_i \in \Delta'$ or $\neg\alpha_i \in \Delta'$. If $\mathsf{PEL} \nvdash \varphi_\Delta \to R_i\neg\varphi_{\Delta'}$, then Δ and Δ' agree about α_i (i.e. either Δ and Δ' both contain α_i, or they both contain $\neg\alpha_i$).

Proof. We prove the contrapositive claim. Suppose that Δ and Δ' disagree about α_i; hence one of the following cases obtains:

- $\alpha_i \in \Delta, \neg\alpha_i \in \Delta'$

  Then $\mathsf{PEL} \vdash \varphi_\Delta \to \alpha_i$ and $\mathsf{PEL} \vdash \alpha_i \to \neg\varphi_{\Delta'}$. By R_i-necessitation and -distribution, we get $\mathsf{PEL} \vdash R_i\alpha \to R_i\neg\varphi_{\Delta'}$ (†). By the success 1 axiom we get from $\mathsf{PEL} \vdash \varphi_\Delta \to \alpha_i$ that also $\mathsf{PEL} \vdash \varphi_\Delta \to R_i\alpha_i$. Combining this with (†), we get that $\mathsf{PEL} \vdash \varphi_\Delta \to R_i\neg\varphi_{\Delta'}$.

- $\neg\alpha_i \in \Delta, \alpha_i \in \Delta'$

  Then $\mathsf{PEL} \vdash \varphi_\Delta \to \neg\alpha_i$ and $\mathsf{PEL} \vdash \neg\alpha_i \to \neg\varphi_{\Delta'}$. By R_i-necessitation and -distribution, we get $\mathsf{PEL} \vdash R_i\neg\alpha_i \to R_i\neg\varphi_{\Delta'}$ (‡). By the success 2 axiom we get from $\mathsf{PEL} \vdash \varphi_\Delta \to \neg\alpha_i$ that also $\mathsf{PEL} \vdash \varphi_\Delta \to R_i\neg\alpha_i$. Combining this with (‡), we get that $\mathsf{PEL} \vdash \varphi_\Delta \to R_i\neg\varphi_{\Delta'}$.

$\square$

Lemma 84. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$, and arbitrary maximal PEL-consistent subsets $\Gamma, \Delta \subseteq Sub^+(\varphi)$. If $\{K_i\chi \,|\, K_i\chi \in \Gamma\} \cup \{\neg K_i\chi \,|\, \neg K_i\chi \in \Gamma\} \subseteq \Delta$, then $\{K_i\chi \,|\, K_i\chi \in \Gamma\} = \{K_i\chi \,|\, K_i\chi \in \Delta\}$.

Proof. Assume that $\{K_i\chi \,|\, K_i\chi \in \Gamma\} \cup \{\neg K_i\chi \,|\, \neg K_i\chi \in \Gamma\} \subseteq \Delta$. Consider an arbitrary formula $K_i\chi \in Sub^+(\varphi)$; we will show that $K_i\chi \in \Gamma$ iff $K_i\chi \in \Delta$.

From right to left. Suppose that $K_i\chi \in \Gamma$. Hence $K_i\chi \in \{K_i\chi \,|\, K_i\chi \in \Gamma\} \cup \{\neg K_i\chi \,|\, \neg K_i\chi \in \Gamma\} \subseteq \Delta$.

From left to right. We prove this by contraposition. Suppose that $K_i\chi \notin \Gamma$. Since $K_i\chi \in Sub^+(\varphi)$, it follows by Lemma 64 that $\neg K_i\chi = {\sim}K_i\chi \in \Gamma$, and hence $\neg K_i\chi \in \{K_i\chi \,|\, K_i\chi \in \Gamma\} \cup \{\neg K_i\chi \,|\, \neg K_i\chi \in \Gamma\} \subseteq \Delta$. By the consistency of Δ it follows that $K_i\chi \notin \Delta$. $\square$

Lemma 85. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$, and an arbitrary maximal PEL-consistent subset $\Gamma \subseteq Sub^+(\varphi)$. If

$$\mathsf{PEL} \vdash (\bigwedge_{K_i\chi \in \Gamma} K_i\chi \wedge \bigwedge_{\neg K_i\chi \in \Gamma} \neg K_i\chi \wedge (\neg)\alpha_i) \to \psi$$

then also

$$\mathsf{PEL} \vdash (\bigwedge_{K_i\chi \in \Gamma} K_i\chi \wedge \bigwedge_{\neg K_i\chi \in \Gamma} \neg K_i\chi \wedge (\neg)\alpha_i) \to R_i\psi$$

Proof. By applying R_i-necessitation and -distribution to the assumption, we get that

$$\mathsf{PEL} \vdash (\bigwedge_{K_i\chi \in \Gamma} R_i K_i\chi \wedge \bigwedge_{\neg K_i\chi \in \Gamma} R_i\neg K_i\chi \wedge R_i(\neg)\alpha_i) \to R_i\psi \qquad (\text{A.5})$$

By positive introspection for K_i and perfect recall for individual knowledge, we have $\mathsf{PEL} \vdash K_i\chi \to R_i K_i\chi$ for any χ. By negative introspection for K_i and perfect recall for individual knowledge, we have $\mathsf{PEL} \vdash \neg K_i\chi \to R_i\neg K_i\chi$ for any χ. Finally, we also have the success axioms $\mathsf{PEL} \vdash \alpha_i \to R_i\alpha_i$ and $\mathsf{PEL} \vdash \neg\alpha_i \to R_i\neg\alpha_i$. Hence for every conjunct θ in the antecedent of the assumption, we have $\mathsf{PEL} \vdash \theta \to R_i\theta$. Combining this with (A.5) we find that

$$\mathsf{PEL} \vdash (\bigwedge_{K_i\chi \in \Gamma} K_i\chi \wedge \bigwedge_{\neg K_i\chi \in \Gamma} \neg K_i\chi \wedge (\neg)\alpha_i) \to R_i\psi$$

$\square$

Lemma 86. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$. Recall that S is the set of all maximal PEL-consistent subsets of $Sub^+(\varphi)$ (and also the domain of the canonical model for $Sub^+(\varphi)$). Let $\Gamma \in S$ arbitrary. Then:

$$X^\psi\chi \in \Gamma \text{ iff } \forall\Delta \in S : (\Gamma, \Delta) \in \left(R^e \cap (S \times S^\psi)\right)^+ \Rightarrow \chi \in \Delta$$

Proof. From left to right. Suppose that $X^\psi\chi \in \Gamma$. Since $\mathcal{R}^+ = \bigcup_{n\geq 1} \mathcal{R}^n$ for any binary relation $\mathcal{R}$, we have to show that

$$\forall n \geq 1 : \forall\Delta \in S : (\Gamma, \Delta) \in \left(R^e \cap (S \times S^\psi)\right)^n \Rightarrow \chi \in \Delta$$

We will prove a slightly stronger claim, viz.

$$\forall n \geq 1 : \forall\Delta \in S : (\Gamma, \Delta) \in \left(R^e \cap (S \times S^\psi)\right)^n \Rightarrow X^\psi\chi \in \Delta \text{ and } \chi \in \Delta$$

We prove this by induction on n.

Base case: $n = 1$. Suppose $(\Gamma, \Delta) \in R^e \cap (S \times S^\psi)$. Hence there is an agent $i \in \{1, 2\}$ such that $(\Gamma, \Delta) \in R_i^e \cap (S \times S^\psi)$. Hence $(\Gamma, \Delta) \in R_i^e = R_i \cap E_i$, and $\psi \in \Delta$. Recall Remark 62, which says that $\alpha_i \in Sub^+(\varphi)$. It follows by Lemma 64 that $\alpha_i \in \Gamma$ or $\neg\alpha_i = {\sim}\alpha_i \in \Gamma$. Assume that $\alpha_i \in \Gamma$ (the case $\neg\alpha_i \in \Gamma$ is analogous). Since $(\Gamma, \Delta) \in E_i$ it follows that also $\alpha_i \in \Delta$. Since $X^\psi\chi \in \Gamma \subseteq Sub^+(\varphi)$, we also have $R_i(\psi \to (\chi \wedge X^\psi\chi)) \in Sub^+(\varphi)$, and thus

also $\alpha_i \to K_i(\alpha_i \to (\psi \to (\chi \land X^\psi \chi))) \in Sub^+(\varphi)$. Since $\alpha_i \in \Gamma$ it follows by Lemma 69 that $K_i(\alpha_i \to (\psi \to (\chi \land X^\psi \chi))) \in \Gamma$. Since $(\Gamma, \Delta) \in R_i$, it follows that also $K_i(\alpha_i \to (\psi \to (\chi \land X^\psi \chi))) \in \Delta$. Since $\alpha_i, \psi \in \Delta$, several applications of Lemma 69 give us that $X^\psi \chi \in \Delta$ and $\chi \in \Delta$.

Induction case: from n to $n+1$. Suppose that $(\Gamma, \Delta) \in \left(R^e \cap (S \times S^\psi)\right)^{n+1}$. Hence there exists a $\Delta' \in S$ such that $(\Gamma, \Delta') \in \left(R^e \cap (S \times S^\psi)\right)^n$ and $(\Delta', \Delta) \in R^e \cap (S \times S^\psi)$. Hence there exists an agent $i \in \{1, 2\}$ such that $(\Delta', \Delta) \in R_i \cap (S \times S^\psi)$. Hence $(\Delta, \Delta') \in R_i \cap E_i$, and $\psi \in \Delta$. Recall Remark 62, which says that $\alpha_i \in Sub^+(\varphi)$. It follows by Lemma 64 that $\alpha_i \in \Delta'$ or $\neg \alpha_i = \sim\alpha_i \in \Delta'$. Assume that $\alpha_i \in \Gamma$ (the case $\neg\alpha_i \in \Gamma$ is analogous). By the induction hypothesis it follows that $X^\psi \chi \in \Delta'$. We now reason as in the base case, and conclude that $X^\psi \chi \in \Delta$ and $\psi \in \Delta$.

From right to left. Define

$$Z := \{\Delta \in S \mid \forall \Delta' \in S : (\Delta, \Delta') \in \left(R^e \cap (S \times S^\psi)\right) \Rightarrow \chi \in \Delta'\}$$

and define $\gamma := \bigvee_{\Delta \in Z} \varphi_\Delta$. We make three claims (the proofs of these three claims are standard; see, for example, [47, Lemma 7.60]):

1. $\mathsf{PEL} \vdash \varphi_\Gamma \to F\gamma$

2. $\mathsf{PEL} \vdash \gamma \to (\psi \to \chi)$

3. $\mathsf{PEL} \vdash \gamma \to F(\psi \to \gamma)$

By X^ψ-necessitation on 3, we get: $\mathsf{PEL} \vdash X^\psi(\gamma \to F(\psi \to \gamma))$. By the induction axiom for X^ψ it follows that $\mathsf{PEL} \vdash F(\psi \to \gamma) \to X^\psi \gamma$. Combining this with 3, we find that $\mathsf{PEL} \vdash \gamma \to X^\psi \gamma$ (†). By X^ψ-necessitation and -distribution on 2, we find that $\mathsf{PEL} \vdash X^\psi \gamma \to (X^\psi \psi \to X^\psi \chi)$. Combining this with (†), we get that $\mathsf{PEL} \vdash \gamma \to (X^\psi \psi \to X^\psi \chi)$. It follows from the induction axiom for X^ψ that $\mathsf{PEL} \vdash X^\psi \psi$, and hence $\mathsf{PEL} \vdash \gamma \to X^\psi \chi$. By F-necessitation and -distribution, it follows that $\mathsf{PEL} \vdash F\gamma \to FX^\psi \chi$, and hence also $\mathsf{PEL} \vdash F\gamma \to X^\psi \chi$. Combining this with 1, we find that $\mathsf{PEL} \vdash \varphi_\Gamma \to X^\psi \chi$. By Lemma 69 it follows that $X^\psi \chi \in \Gamma$. $\qquad\square$

Lemma 87. (Truth lemma.) Consider an arbitary formula $\varphi \in \mathcal{L}^\alpha$. Let $\mathbb{M}$ be the PEL-canonical model for $Sub^+(\varphi)$. It holds for all $\psi \in Sub^+(\varphi)$ and all $\Gamma \in S$ that $\mathbb{M}, \Gamma \models \psi$ iff $\psi \in \Gamma$.

Proof. This is proved by induction on the complexity of ψ. The base case and the induction cases for $\neg\psi, \psi_1 \land \psi_2, K_i\psi$ and $C^{\psi_1}\psi_2$ are standard; we refer the reader to [47, Lemma 7.61]. We will now prove the induction cases for $R_i\psi, X^{\psi_1}\psi_2$ and $\sum_\ell a_\ell P_i(\varphi_\ell) \geq k$.

Induction case for $R_i\psi$. Suppose that $R_i\psi$ is in the set $Sub^+(\varphi)$. Let $\Gamma \in S$ arbitrary; we will show that $\mathbb{M}, \Gamma \models R_i\psi$ iff $R_i\psi \in \Gamma$.

From left to right. Assume that $\mathbb{M}, \Gamma \models R_i\psi$. Recall Remark 62, which says that $\alpha_i \in Sub^+(\varphi)$. It follows by Lemma 64 that $\alpha_i \in \Gamma$ or $\neg\alpha_i = {\sim}\alpha_i \in \Gamma$. Now define

$$Z := \begin{cases} \{{\sim}\psi\} \cup \{K_i\chi \mid K_i\chi \in \Gamma\} \cup \{\neg K_i\chi \mid \neg K_i\chi \in \Gamma\} \cup \{\alpha_i\} & \text{if } \alpha_i \in \Gamma \\ \{{\sim}\psi\} \cup \{K_i\chi \mid K_i\chi \in \Gamma\} \cup \{\neg K_i\chi \mid \neg K_i\chi \in \Gamma\} \cup \{\neg\alpha_i\} & \text{if } \neg\alpha_i \in \Gamma \end{cases}$$

For a reductio, suppose that Z is consistent. It is easy to check that $Z \subseteq Sub^+(\varphi)$. By Lemma 70, Z can be extended to a maximal PEL-consistent $Z^+ \subseteq Sub^+(\varphi)$. Since $\{K_i\chi \mid K_i\chi \in \Gamma\} \cup \{\neg K_i\chi \mid \neg K_i\chi \in \Gamma\} \subseteq Z \subseteq Z^+$, it follows by Lemma 84 that $(\Gamma, Z^+) \in R_i$. By definition of Z, we know that Γ and Z^+ agree about α_i, and thus $(\Gamma, Z^+) \in E_i$. Hence $(\Gamma, Z^+) \in R_i \cap E_i$. Since $\mathbb{M}, \Gamma \models R_i\psi$, it follows that $\mathbb{M}, Z^+ \models \psi$, and thus, by the induction hypothesis, $\psi \in Z^+$. But by definition of Z, also ${\sim}\psi \in Z^+$, which contradicts the PEAL-consistency of Z^+.

Hence we conclude that Z is not consistent, so

$$\mathsf{PEL} \vdash \big(\bigwedge_{K_i\chi \in \Gamma} K_i\chi \wedge \bigwedge_{\neg K_i\chi \in \Gamma} \neg K_i\chi \wedge (\neg)\alpha_i \big) \to \psi$$

By Lemma 85 it follows that

$$\mathsf{PEL} \vdash \big(\bigwedge_{K_i\chi \in \Gamma} K_i\chi \wedge \bigwedge_{\neg K_i\chi \in \Gamma} \neg K_i\chi \wedge (\neg)\alpha_i \big) \to R_i\psi$$

It is easy to see that for each conjunct θ of the above antecedent, we have $\mathsf{PEL} \vdash \varphi_\Gamma \to \theta$. Hence it follows that $\mathsf{PEL} \vdash \varphi_\Gamma \to R_i\psi$. Since $R_i\psi \in Sub^+(\varphi)$ it follows by Lemma 69 that $R_i\psi \in \Gamma$.

From left to right. Assume that $R_i\psi \in \Gamma$. Consider an arbitrary $\Delta \in S$ and suppose that $(\Gamma, \Delta) \in R_i \cap E_i$; we will show that $\mathbb{M}, \Delta \models \psi$. Recall Remark 62, which says that $\alpha_i \in Sub^+(\varphi)$. It follows by Lemma 64 that $\alpha_i \in \Gamma$ or $\neg\alpha_i = {\sim}\alpha_i \in \Gamma$. Assume that $\alpha_i \in \Gamma$ (the case $\neg\alpha_i \in \Gamma$ is analogous). By the axiom which defines R_i in terms of K_i, we have that $\mathsf{PEL} \vdash (R_i\psi \wedge \alpha_i) \to K_i(\alpha_i \to \psi)$. Since $R_i\psi, \alpha_i \in \Gamma$, we get that $\mathsf{PEL} \vdash \varphi_\Gamma \to K_i(\alpha_i \to \psi)$. Since $R_i\psi \in Sub^+(\varphi)$, we also have $K_i(\alpha_i \to \psi) \in Sub^+(\varphi)$, so by Lemma 69 it follows that $K_i(\alpha_1 \to \psi) \in \Gamma$. Since $(\Gamma, \Delta) \in R_i$, this means that $K_i(\alpha_i \to \psi) \in \Delta$, and hence also $\alpha_i \to \psi \in \Delta$ (by another application of Lemma 69). Since $\alpha_i \in \Gamma$ and $(\Gamma, \Delta) \in E_i$, we get by the definition of E_i that $\alpha_i \in \Delta$, and hence $\psi \in \Delta$ (by yet another application of Lemma 69). By the induction hypothesis it follows that $\mathbb{M}, \Delta \models \psi$.

Induction case for $X^{\psi_1}\psi_2$. Suppose that $X^{\psi_1}\psi_2$ is in the set $Sub^+(\varphi)$. Let $\Gamma \in S$ arbitrary; then we have:

$$\begin{aligned} \mathbb{M}, \Gamma \models X^{\psi_1}\psi_2 \quad &\Leftrightarrow \quad \forall \Delta \in S : (\Gamma, \Delta) \in \big(R^e \cap (S \times [\![\psi_1]\!]^{\mathbb{M}}) \big)^+ \Rightarrow \mathbb{M}, \Delta \models \psi_2 \\ &\overset{\text{IH}}{\Leftrightarrow} \quad \forall \Delta \in S : (\Gamma, \Delta) \in \big(R^e \cap (S \times S^{\psi_1}) \big)^+ \Rightarrow \psi_2 \in \Delta \\ &\Leftrightarrow \quad X^{\psi_1}\psi_2 \in \Gamma \qquad \text{(by Lemma 86)} \end{aligned}$$

Induction case for $\sum_\ell a_\ell P_i(\psi_\ell) \geq k$.[3] Suppose that $\psi := \sum_\ell a_\ell P_i(\psi_\ell) \geq k$ is in the set $Sub^+(\varphi)$. Let $\Gamma \in S$ arbitrary; we will show that $\mathbb{M}, \Gamma \models \psi$ iff $\psi \in \Gamma$.

From right to left. Suppose that $\psi \in \Gamma$; we will show that $\mathbb{M}, \Gamma \models \psi$. This means that we have to show that

$$\sum_\ell a_\ell \mu_i(\Gamma)([\![\psi_\ell]\!]^{\mathbb{M}}) \geq b$$

For any ℓ, we have by the induction hypothesis that $\mathbb{M}, \Delta \models \psi_\ell$ iff $\psi_\ell \in \Delta$, for any $\Delta \in S$. Hence for all ℓ we have that

$$\mu_i(\Gamma)([\![\psi_\ell]\!]^{\mathbb{M}}) = \sum_{\Delta \in [\![\psi_\ell]\!]^{\mathbb{M}}} \mu_i(\Gamma)(\Delta) = \sum_{\{\Delta \in S \,|\, \psi_\ell \in \Delta\}} \mu_i(\Gamma)(\Delta) = \sum_{\Delta \in S^{\psi_\ell}} \mu_i(\Gamma)(\Delta)$$

Hence it suffices to show that

$$\sum_\ell \Big(\sum_{\Delta \in S^{\psi_\ell}} \mu_i(\Gamma)(\Delta) \Big) \geq b$$

i.e.:

$$\sum_\ell \sum_{\Delta \in S^{\psi_\ell}} a_\ell \mu_i(\Gamma)(\Delta) \geq b \qquad \text{(A.6)}$$

Since $\psi \in \Gamma$, the characteristic system $\sigma_{i,\Gamma}$ contains the inequality $\sum_{\Delta \in S} c_\Delta^\psi x_{i,\Gamma,\Delta} \geq b^\psi$. Since $\{\mu_i(\Gamma)(\Delta)\}_{\Delta \in S}$ is (by definition) a solution of this system, it follows that $\sum_{\Delta \in S} c_\Delta^\psi \mu_i(\Gamma)(\Delta) \geq b^\psi$. Recall from the proof of Lemma 76 that $c_\Delta^\psi = \sum_{\{\ell \,|\, \Delta \in S^\psi\}} a_\ell = \sum_{\{\ell \,|\, \psi_\ell \in \Delta\}} a_\ell$ (for $\Delta \in S$) and that $b^\psi = b$. Hence we get:

$$\sum_{\Delta \in S} \Big(\sum_{\{\ell \,|\, \psi_\ell \in \Delta\}} a_\ell \Big) \mu_i(\Gamma)(\Delta) \geq b$$

i.e.:

$$\sum_{\Delta \in S} \sum_{\{\ell \,|\, \psi_\ell \in \Delta\}} a_\ell \mu_i(\Gamma)(\Delta) \geq b$$

But this is just another way of expressing (A.6), which was what we needed to prove.

From left to right. We will prove the contrapositive claim, viz. if $\psi \notin \Gamma$ then $\mathbb{M}, \Gamma \not\models \psi$. Suppose that $\psi \notin \Gamma$. Then by Lemma 64 we have $\neg\psi = {\sim}\psi \in \Gamma$, and thus the characteristic system $\sigma_{i,\Gamma}$ contains the inequality $\sum_{\Delta \in S} c_\Delta^\psi x_{i,\Gamma,\Delta} < b^\psi$. From this point onwards, the proof is completely analogous to the proof for the right to left direction. $\square$

---

[3]Many thanks to Dick de Jongh for helping me sort out the details of this case.

Theorem 57, part 2. PEL is complete with respect to $\mathcal{PKB}$.

Proof. We have to show that $\mathcal{PKB} \models \varphi \Rightarrow \mathsf{PEL} \vdash \varphi$ for all $\varphi \in \mathcal{L}^\alpha$. Let $\varphi \in \mathcal{L}^\alpha$ arbitrary; we proceed by contraposition. Suppose that $\mathsf{PEL} \nvdash \varphi$. Hence $\neg\varphi$ is PEL-consistent. Now consider $\mathbb{M}$, the PEL-canonical model for $Sub^+(\neg\varphi)$. Since $\neg\varphi$ is PEL-consistent, Lemma 82 tells us that $\mathbb{M}$ is a probabilistic Kripke model with binary experiments, and Lemma 70 tells us that $\{\neg\varphi\}$ can be extended to a state Γ of $\mathbb{M}$. Since $\neg\varphi \in \Gamma$, it follows by Lemma 87 that $\mathbb{M}, \Gamma \models \neg\varphi$, so $\mathbb{M}, \Gamma \nmodels \varphi$, and thus $\mathcal{PKB} \nmodels \varphi$. $\qquad\square$

Appendix B

Metatheory of PEAL

In this appendix we prove the soundness and completeness of PEAL with respect to $\mathcal{AGR}$.

B.1 Soundness

Theorem 58, part 1. PEAL is sound with respect to $\mathcal{AGR}$.

Proof. We already showed in Appendix A that the axioms in components 1–7 are sound with respect to $\mathcal{PKB}$. Hence they are certainly sound with respect to $\mathcal{AGR}$, which is a subclass of $\mathcal{PKB}$. We now turn to component 8, the agreement component. Note that every agreement frame satisfies, by definition, the frame conditions being characterized by (i.e. the 'left-hand-sides' of) Lemmas 42 and 45. It follows from Lemma 42 that axiom 1 is sound with respect to $\mathcal{AGR}$. It follows from Lemma 45 that axioms 2a and 2b are sound with respect to $\mathcal{AGR}$. □

B.2 Completeness

We will now prove the completeness of PEAL with respect to $\mathcal{AGR}$. First we introduce several additional definitions and lemmas, which will be used in the proof of the actual completeness theorem at the end of this section.

In Appendix A we introduced several definitions and lemmas about PEL. These have direct analogues for PEAL (which are often obtained by simply substituting PEAL for PEL everywhere in the formulations and proofs). Rather than introducing all these definitions, lemmas and remarks all over again in this appendix, we will just say 'Lemma X for PEAL', where Lemma X is one of these lemmas about PEL proved in Appendix A. In particular, this is the case for Lemma 64, Definition 66, and Lemmas 67, 69, 70, 72, 73, 74, 75, and 76.

We now introduce a new notion of characteristic system.

Definition 88. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$ and a $\Gamma \in S$ (we now use S to denote the set of maximal PEAL-consistent subsets $\Gamma \subseteq Sub^+(\varphi)$). Note that by Lemma 67 for PEAL, S is finite. The *characteristic system* of Γ (abbreviated σ_Γ) is a system in $|S|$ variables x_Δ (for $\Delta \in S$). This system is defined as follows:

- First of all, σ_Γ has the (in)equalities displayed below.[1] These (in)equalities will sometimes be called the 'fixed' (in)equalities of $\sigma_{i,\Gamma}$, because they do not depend on any i-probability formula ψ.

$$\sum_{\Delta \in S} x_\Delta = 1 \tag{B.1}$$

$$x_\Delta \geq 0 \qquad \text{(for each } \Delta \in S\text{)} \tag{B.2}$$

$$x_\Gamma > 0 \tag{B.3}$$

$$x_\Delta = 0 \qquad \text{(for each } \Delta \in \overline{R_1[\Gamma]} \cup \overline{R_2[\Gamma]}\text{)} \tag{B.4}$$

- By Lemma 65, $Sub^+(\varphi)$ is finite, and thus contains only finitely many i-probability formulas ψ. For each such ψ there exist, by Lemma 76 for PEAL, rational numbers $\{c_\Delta^\psi\}_{\Delta \in S}$ and b^ψ such that PEAL $\vdash \psi \leftrightarrow \sum_{\Delta \in S} c_\Delta^\psi P_i(\varphi_\Delta) \geq b^\psi$. For each of these finitely many probability formulas $\psi \in Sub^+(\varphi)$, note that by Lemma 64 for PEAL either $\psi \in \Gamma$ or $\neg \psi = {\sim}\psi \in \Gamma$. For any i-probability formula $\psi \in Sub^+(\varphi)$, if $\psi \in \Gamma$, add the inequality

$$\sum_{\Delta \in S} c_\Delta^\psi x_{i,\Gamma,\Delta} \geq b^\psi$$

to the system σ_Γ, and if $\neg \psi \in \Gamma$, add the inequality

$$\sum_{\Delta \in S} c_\Delta^\psi x_{i,\Gamma,\Delta} < b^\psi$$

to the system σ_Γ. This finishes the definition of σ_Γ.

Lemma 89. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$. Recall that S is the set of all maximal PEAL-consistent subsets of $Sub^+(\varphi)$. For $i = 1, 2$, define a relation $R_i \subseteq S \times S$ by putting:

$$(\Gamma, \Delta) \in R_i :\Leftrightarrow \begin{cases} \{K_i\chi \mid K_i\chi \in \Gamma\} = \{K_i\chi \mid K_i\chi \in \Delta\} \\ \text{for all 1- and 2-probability formulas } \psi \in Sub^+(\varphi) : \psi \in \Gamma \Leftrightarrow \psi \in \Delta \end{cases}$$

and also put $R = R_1 \cup R_2$. Let $\Delta \in S$ arbitrary. Then for all $\Delta' \in S$: if $(\Delta, \Delta') \in R^*$, then Δ and Δ' agree on all 1- and 2-probability formulas.

---

[1] We define $\overline{R_i[\Gamma]} := \{\Delta \in S \mid \text{either } \{K_i\chi \mid K_i\chi \in \Gamma\} \neq \{K_i\chi \mid K_i\chi \in \Delta\}$ or Γ and Δ disagree about a probability formula $\psi \in Sub^+(\varphi)\}$. For now, these can be considered as strictly formally defined sets; later, however, it will be clear that these sets are related with agent i's epistemic accessibility relation in the PEAL-canonical model, cf. Definition 95.

Proof. Since $R^* = \bigcup_{n \geq 0} R^n$, we have to show that $\forall n \geq 0 : \forall \Delta' \in S :$ if $(\Delta, \Delta') \in R^n$ then Δ and Δ' agree on all 1- and 2-probability formulas. We prove this by induction on n.

Base case: $n = 0$. For any $\Delta' \in S$, if $(\Delta, \Delta') \in R^0$ then $\Delta' = \Delta$, and so Δ and Δ' trivially agree on all 1- and 2-probability formulas.

Induction case. Consider an arbitrary $\Delta' \in S$ and suppose that $(\Delta, \Delta') \in R^{n+1}$. Then there exists a $\Delta'' \in S$ such that $(\Delta, \Delta'') \in R^n$ and $(\Delta'', \Delta') \in R = R_1 \cup R_2$. Suppose that $(\Delta'', \Delta') \in R_1$; the case $(\Delta'', \Delta') \in R_2$ is analogous. By the definition of R_1 it follows that Δ'' and Δ' agree on all 1- and 2-probability formulas. Since $(\Delta, \Delta'') \in (R^e)^n$ we get by the induction hypothesis that Δ and Δ'' agree on all 1- and 2-probability formulas. Hence also Δ and Δ' agree on all 1- and 2-probability formulas. $\square$

Remark 90. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$ and a set $\Gamma \in S$. Now consider the characteristic system σ_Γ. Note that for any inequality χ in the system σ_Γ it holds that $\chi[P_1(\varphi_\Delta)/x_\Delta]$ is a sentence of $\mathcal{L}^\alpha$ (where $\chi[P_1(\varphi_\Delta)/x_\Delta]$ is the result of uniformly substituting $P_1(\varphi_\Delta)$ for x_Δ in χ, for each $\Delta \in S$). Also note that we everywhere use P_1 —we could as well have used P_2 (cf. infra), but it is important that we always use the *same* agent subscript in these substitutions. By $\bigwedge \sigma_\Gamma[P_1(\varphi_\Delta)/x_\Delta]$ we represent the conjunction of all the $\mathcal{L}_i^\alpha$-sentences that are obtained by carrying out this uniform substitution on inequalities of σ_Γ.

Lemma 91. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$. Recall that S is the set of all maximal PEAL-consistent subsets of $Sub^+(\varphi)$. Consider arbitrary $\Gamma, \Gamma' \in S$ and suppose that $(\Gamma, \Gamma') \in R^*$ (where R is the relation defined in the formulation of Lemma 89). Then $\vdash \varphi_\Gamma \to \bigwedge \sigma_{\Gamma'}[P_1(\varphi_\Delta)/x_\Delta]$.

Proof. It suffices to show that PEAL $\vdash \varphi_\Gamma \to \chi[P_1(\varphi_\Delta)/x_\Delta]$ for each (in)equality χ of the system $\sigma_{\Gamma'}$. We prove this for all of these inequalities χ separately:

- χ is the 'fixed' equality $\sum_{\Delta \in S} x_\Delta = 1$

  One can show that PEAL $\vdash \chi[P_1(\varphi_\Delta)/x_\Delta]$ (cf. the proof of Lemma 79), from which it follows that PEAL $\vdash \varphi_\Gamma \to \chi[P_1(\varphi_\Delta)/x_\Delta]$.

- χ is the 'fixed' inequality $x_\Delta \geq 0$ (for some $\Delta \in S$)

  One can show that PEAL $\vdash \chi[P_1(\varphi_\Delta)/x_\Delta]$ (cf. the proof of Lemma 79), from which it follows that PEAL $\vdash \varphi_\Gamma \to \chi[P_1(\varphi_\Delta)/x_\Delta]$.

- χ is the 'fixed' equality $x_\Delta = 0$ (for some $i \in \{1, 2\}$ and $\Delta \in \overline{R_i[\Gamma']}$)

  Since $\Delta \in \overline{R_i[\Gamma']}$, we know that either $\{K_i\psi \mid K_i\psi \in \Gamma'\} \neq \{K_i\psi \mid K_i\psi \in \Delta\}$ or Γ' and Δ disagree about a 1- or a 2-probability formula.

  Suppose that $\{K_i\psi \mid K_i\psi \in \Gamma'\} \neq \{K_i\psi \mid K_i\psi \in \Delta\}$; we will show that Γ' and Δ disagree about an i-probability formula. From our assumption, it follows that there is a $K_i\chi \in Sub^+(\varphi)$ such that $K_i\chi \in \Gamma'$ and $K_i\chi \notin \Delta$ (the case $K_i\chi \notin \Gamma', K_i\chi \in \Delta$ is analogous). Since $K_i\chi \in Sub^+(\varphi)$, we also have $P_i(K_i\chi) = 1, \neg(P_i(K_i\chi) = 1) \in Sub^+(\varphi)$. Since $K_i\chi \notin \Delta$ we get by Lemma 64 for PEAL that $\neg K_i\chi = \sim K_i\chi \in \Delta$. By positive introspection

for K_i and the axiom that $K_i\varphi \rightarrow P_i(\varphi) = 1$, we have that $\mathsf{PEAL} \vdash K_i\chi \rightarrow P_i(K_\chi) = 1$. Since $K_i\chi \in \Gamma$, it follows by Lemma 69 for PEAL that $P_i(K_i\chi) = 1 \in \Gamma'$. Furthermore, by negative introspection for K_i and the axiom that $K_i\varphi \rightarrow P_i(\varphi) = 1$, we have that $\mathsf{PEAL} \vdash \neg K_i\chi \rightarrow P_i(\neg K_\chi) = 1$. By finite additivity we get that $\mathsf{PEAL} \vdash \neg K_i\chi \rightarrow P_i(K_i\chi) = 0$, and thus $\mathsf{PEAL} \vdash \neg K_i\chi \rightarrow \neg(P_i(K_i\chi) = 1)$. Since $\neg K_i\chi \in \Delta$, it follows by Lemma 69 for PEAL that $\neg(P_i(K_i\chi) = 1) \in \Delta$. Hence Γ' and Δ disagree about an i-probability formula, viz. $P_i(K_i\chi) = 1$.

We know already that either $\{K_i\psi \,|\, K_i\psi \in \Gamma'\} \neq \{K_i\psi \,|\, K_i\psi \in \Delta\}$, or Γ' and Δ disagree about a 1- or a 2-probability formula. We have just proved that if $\{K_i\psi \,|\, K_i\psi \in \Gamma'\} \neq \{K_i\psi \,|\, K_i\psi \in \Delta\}$, then Γ' and Δ disagree about an i-probability formula. So in all cases, Γ' and Δ disagree about a 1- or a 2-probability formula.

Since $(\Gamma, \Gamma') \in R^*$, we know by Lemma 89 that Γ and Γ' agree about all 1- and 2-probability formulas. Since Γ' and Δ disagree about a 1- or a 2-probability formula, it follows that also Γ and Δ disagree about a 1- or a 2-probability formula —call it ψ. We now make the following case distinction:

- $\psi \in \Gamma, \psi \notin \Delta$

  By Lemma 64 for PEAL we have that $\neg\psi = {\sim}\psi \in \Delta$. Hence $\mathsf{PEAL} \vdash \varphi_\Gamma \rightarrow \psi$ and $\mathsf{PEAL} \vdash \psi \rightarrow \neg\varphi_\Delta$. Note that ψ is an i-probability formula (for either $i = 1$ or $i = 2$). Therefore we have $\mathsf{PEAL} \vdash \psi \rightarrow K_i\psi$ (by the second axiom scheme of the agreement component), and thus $\mathsf{PEAL} \vdash \varphi_\Gamma \rightarrow K_i\psi$ (†). By K_i-necessitation and -distribution on $\mathsf{PEAL} \vdash \psi \rightarrow \neg\varphi_\Delta$ we get that $\mathsf{PEAL} \vdash K_i\psi \rightarrow K_i\neg\varphi_\Delta$. By the axiom $K_i\varphi \rightarrow P_i(\varphi) = 1$ and finite additivity, we have that $\mathsf{PEAL} \vdash K_i\neg\varphi_\Delta \rightarrow P_i(\varphi_\Delta) = 0$. Hence $\mathsf{PEAL} \vdash K_i\psi \rightarrow P_i(\varphi_\Delta) = 0$. Combining this with (†), we get that $\mathsf{PEAL} \vdash \varphi_\Gamma \rightarrow P_i(\varphi_\Delta) = 0$.

  If $i = 1$, then this means that $\mathsf{PEAL} \vdash \varphi_\Gamma \rightarrow P_1(\varphi_\Delta) = 0$, as desired. If $i = 2$, however, this means that $\mathsf{PEAL} \vdash \varphi_\Gamma \rightarrow P_2(\varphi_\Delta) = 0$. Since $\mathsf{PEAL} \vdash P_1(\varphi_\Delta) = 0 \leftrightarrow P_2(\varphi_\Delta) = 0$ (by the first axiom scheme of the agreement component), we still obtain $\mathsf{PEAL} \vdash \varphi_\Gamma \rightarrow P_1(\varphi_\Delta) = 0$, as desired.

- $\psi \notin \Gamma, \psi \in \Delta$

  In an analogous fashion, we derive that $\mathsf{PEAL} \vdash \varphi_\Gamma \rightarrow P_1(\varphi_\Delta) = 0$.

- χ is $\sum_{\Delta \in S} c_\Delta^\psi x_\Delta \geq b^\psi$ (for some probability formula $\psi \in Sub^+(\varphi)$ such that $\psi \in \Gamma'$)

  Since $(\Gamma, \Gamma') \in R^*$, we know by Lemma 89 that Γ and Γ' agree on all probability formulas. Since Γ' contains the probability formula ψ, we thus also have $\psi \in \Gamma$. Hence $\mathsf{PEAL} \vdash \varphi_\Gamma \rightarrow \psi$. Note that ψ is either a 1-probability formula or a 2-probability formula:

  - ψ is a 1-probability formula

By Lemma 76 for PEAL it holds that $\vdash \psi \leftrightarrow \sum_{\Delta \in S} c_\Delta^\psi P_1(\varphi_\Delta) \geq b^\psi$, and hence $\mathsf{PEAL} \vdash \varphi_\Gamma \to \sum_{\Delta \in S} c_\Delta^\psi P_1(\varphi_\Delta) \geq b^\psi$.

- ψ is a 2-probability formula

  By Lemma 76 for PEAL it holds that $\mathsf{PEAL} \vdash \psi \leftrightarrow \sum_{\Delta \in S} c_\Delta^\psi P_2(\varphi_\Delta) \geq b^\psi$, and hence $\mathsf{PEAL} \vdash \varphi_\Gamma \to \sum_{\Delta \in S} c_\Delta^\psi P_2(\varphi_\Delta) \geq b^\psi$. By axiom 1 of the agreement component we have $\mathsf{PEAL} \vdash \sum_{\Delta \in S} c_\Delta^\psi P_1(\varphi_\Delta) \geq b^\psi \leftrightarrow \sum_{\Delta \in S} c_\Delta^\psi P_2(\varphi_\Delta) \geq b^\psi$, and thus $\mathsf{PEAL} \vdash \varphi_\Gamma \to \sum_{\Delta \in S} c_\Delta^\psi P_1(\varphi_\Delta) \geq b^\psi$.

- χ is $\sum_{\Delta \in S} c_\Delta^\psi x_{i,\Gamma,\Delta} < b^\psi$ (for some probability formula $\psi \in Sub^+(\varphi)$ such that $\neg\psi \in \Gamma'$)

  Analogous to the previous case.

$\square$

Remark 92. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$. Recall that S is the set of all maximal PEAL-consistent subsets of $Sub^+(\varphi)$, and recall the relation $R \subseteq S \times S$ defined in the formulation of Lemma 89. Note that R^* is an equivalence relation on S. Hence it gives rise to a partition $\mathcal{P}$ of S. Since S is finite, also this partition will have finitely many cells, say $\mathcal{P} = \{C_1, \ldots, C_n\}$. For every such cell C_i, we define a big system of (in)equalities σ_{C_i} by putting together all the systems σ_Γ for $\Gamma \in C_i$. Note that

$$\bigwedge \sigma_{C_i}[P_1(\varphi_\Delta)/x_\Delta] = \bigwedge_{\Gamma \in C_i} \bigwedge \sigma_\Gamma[P_1(\varphi_\Delta)/x_\Delta]$$

Lemma 93. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$ and a set $\Gamma \in S$. Then $\mathsf{PEAL} \vdash \varphi_\Gamma \to \bigwedge \sigma_{R^*[\Gamma]}[P_1(\varphi_\Delta)/x_\Delta]$.

Proof. For any $\Gamma' \in R^*[\Gamma]$ we have by Lemma 91 that $\mathsf{PEAL} \vdash \varphi_\Gamma \to \bigwedge \sigma_{\Gamma'}[P_1(\varphi_\Delta)/x_\Delta]$. Hence also

$$\mathsf{PEAL} \vdash \varphi_\Gamma \to \bigwedge_{\Gamma' \in R^*[\Gamma]} \bigwedge \sigma_{\Gamma'}[P_1(\varphi_\Delta)/x_\Delta]$$

i.e. $\mathsf{PEAL} \vdash \varphi_\Gamma \to \bigwedge \sigma_{R^*[\Gamma]}[P_1(\varphi_\Delta)/x_\Delta]$. $\square$

Lemma 94. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$ and a set $\Gamma \in S$. Then the characteristic system $\sigma_{R^*[\Gamma]}$ has a solution $\{x_\Delta^*\}_{\Delta \in S}$.

Proof. For a reductio, suppose that $\sigma_{R^*[\Gamma]}$ has no solution. Hence for every assignment $a\colon \{x_\Delta\}_{\Delta \in S} \to \mathbb{R}$ it holds that $a \models \neg\bigwedge \sigma_{R^*[\Gamma]}$. Hence $\neg\bigwedge \sigma_{R^*[\Gamma]}$ is a 'valid linear inequality' (in the terminology of [23]). Therefore, $\neg\bigwedge \sigma_{R^*[\Gamma]}[P_1(\varphi_\Delta)/x_\Delta]$ is an *instance* of a valid linear inequality, and thus $\mathsf{PEAL} \vdash \neg\bigwedge \sigma_{R^*[\Gamma]}[P_1(\varphi_\Delta)/x_\Delta]$ (since the linear inequalities component is complete for instances of valid linear inequalities; cf. Section 4 of [23]). Together with Lemma 93 this implies that $\mathsf{PEAL} \vdash \neg\varphi_\Gamma$, which contradicts the PEAL-consistency of Γ. $\square$

Definition 95. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$. The PEAL-*canonical model for* $Sub^+(\varphi)$ is defined as $\mathbb{M} := \langle S, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V \rangle$, where

- $S := \{\Gamma \subseteq \mathcal{L}^\alpha \,|\, \Gamma \text{ is a maximal PEAL-consistent subset of } Sub^+(\varphi)\}$

- $R_i \subseteq S \times S$ is defined by:

  $(\Gamma, \Delta) \in R_i :\Leftrightarrow \begin{cases} \{K_i\chi \,|\, K_i\chi \in \Gamma\} = \{K_i\chi \,|\, K_i\chi \in \Delta\} \\ \text{for all 1- and 2-probability formulas } \psi \in Sub^+(\varphi) : \psi \in \Gamma \Leftrightarrow \psi \in \Delta \end{cases}$

- for any $p \in Prop \cup \{\alpha_1, \alpha_2\}$, put $V(p) := \{\Gamma \in S \,|\, p \in \Gamma\}$

- $E_i := \big(V(\alpha_i) \times V(\alpha_i)\big) \cup \big((S - V(\alpha_i)) \times (S - V(\alpha_i))\big)$

- for any $\Gamma, \Delta \in S$, put $\mu_i(\Gamma)(\Delta) := x_\Delta^*$ (where $\{x_\Delta^*\}_{\Delta \in S}$ is a solution to the characteristic system $\sigma_{R^*[\Gamma]}$; this solution is guaranteed to exist by Lemma 94)

Lemma 96. Consider an arbitrary formula $\varphi \in \mathcal{L}^\alpha$, and assume that φ is PEAL-consistent. Then the PEAL-canonical model for $Sub^+(\varphi)$, $\mathbb{M}$, is an agreement model.

Proof. Note the following:

- S is finite (by Lemma 67 for PEAL), and non-empty: since φ is PEAL-consistent, Lemma 70 for PEAL says that the set $\{\varphi\} \subseteq Sub^+(\varphi)$ can be extended to an element $\Gamma \in S$ —and hence $S \neq \emptyset$

- R_1, R_2 are equivalence relations (this is easy to check)

- E_i satisfies condition (5.5) on binary experiments (by definition of E_i)

- for each agent $i \in \{1, 2\}$ and state $\Gamma \in S$, it holds that $\mu_i(\Gamma)$ is a probability mass function on S: because $\{\mu_i(\Gamma)(\Delta)\}_{\Delta \in S}$ is (by definition) a solution of the characteristic system $\sigma_{R^*[\Gamma]}$, which contains σ_Γ as a 'subsystem', we have that $\sum_{\Delta \in S} \mu_i(\Gamma)(\Delta) = 1$ and that $\mu_i(\Gamma)(\Delta) \geq 0$ for all $\Delta \in S$

- condition (i) of Definition 1 is fulfilled: for all $\Gamma \in S$, we have that $\mu_i(\Gamma)(\Gamma) > 0$ (since $\mu_i(\Gamma)(\Gamma)$ is a solution to the inequality $x_{i,\Gamma} > 0$ of the system $\sigma_{i,\Gamma}$

- condition (ii) of Definition 1 is fulfilled: for all $\Gamma, \Delta \in S$, if $(\Gamma, \Delta) \notin R_i$ then $\mu_i(\Gamma)(\Delta) = 0$ (since $\mu_i(\Gamma)(\Delta)$ is a solution to the equality $x_{i,\Gamma,\Delta} = 0$ for all $\Delta \notin R_i[\Gamma]$ of the sytem $\sigma_{i,\Gamma}$)

- $\mu_1 = \mu_2$: this follows from the definition, since $\mu_i(\Gamma)(\Delta)$ is defined as x_Δ^* independent of agent i (and hence $\mu_1(\Gamma)(\Delta) = x_\Delta^* = \mu_2(\Gamma)(\Delta)$)

- for all $\Gamma, \Gamma' \in S$: if $(\Gamma, \Gamma') \in R^*$ then $\mu_i(\Gamma) = \mu_i(\Gamma')$

  If $(\Gamma, \Gamma') \in R^*$ then the system $\sigma_{R^*[\Gamma]}$ is identical to the system $\sigma_{R^*[\Gamma']}$. Hence they will have the same solution, and thus $\mu_1(\Gamma)(\Delta) = \mu_1(\Gamma')(\Delta)$ for all $\Delta \in S$, and thus $\mu_1(\Gamma) = \mu_1(\Gamma')$. Analogously for agent 2.

$\square$

Lemma 97. (Truth lemma.) Consider an arbitary formula $\varphi \in \mathcal{L}^\alpha$. Let $\mathbb{M}$ be the PEAL-canonical model for $Sub^+(\varphi)$. It holds for all $\psi \in Sub^+(\varphi)$ and all $\Gamma \in S$ that $\mathbb{M}, \Gamma \models \psi$ iff $\psi \in \Gamma$.

Proof. The proof of this lemma proceeds largely the same as the proof of the truth lemma for PEL. We need to take into account that the definitions of R_i and μ_i have changed, which leads to some subtle changes in the induction cases for the epistemic and probabilistic formulas. Filling in the details is a tedious but straightforward exercise. $\square$

Theorem 58, part 2. PEAL is complete with respect to $\mathcal{AGR}$.

Proof. We have to show that $\mathcal{AGR} \models \varphi \Rightarrow$ PEAL $\vdash \varphi$ for all $\varphi \in \mathcal{L}^\alpha$. Let $\varphi \in \mathcal{L}^\alpha$ arbitrary; we proceed by contraposition. Suppose that PEAL $\nvdash \varphi$. Hence $\neg\varphi$ is PEAL-consistent. Now consider $\mathbb{M}$, the PEAL-canonical model for $Sub^+(\neg\varphi)$. Since $\neg\varphi$ is PEAL-consistent, Lemma 96 tells us that $\mathbb{M}$ is a probabilistic Kripke model with binary experiments, and Lemma 70 for PEAL tells us that $\{\neg\varphi\}$ can be extended to a state Γ of $\mathbb{M}$. Since $\neg\varphi \in \Gamma$, it follows by Lemma 97 that $\mathbb{M}, \Gamma \models \neg\varphi$, so $\mathbb{M}, \Gamma \nvDash \varphi$, and thus $\mathcal{AGR} \nvDash \varphi$. $\square$

Appendix C

Metatheory of DPEALe

In this appendix we prove the soundness and completeness of DPEALe with respect to $\mathcal{AGR}$.

C.1 Soundness

Theorem 59, part 1. DPEALe is sound with respect to $\mathcal{AGR}$.

Proof. We already showed in Appendix B that the axioms in components 1–8 are sound with respect to $\mathcal{AGR}$. We now turn to component 9, the reduction axioms for [EXP]. We will explicitly prove the soundness of the reduction axioms for $K_i\varphi, X^\varphi\psi, [\text{EXP}]\,\varphi$, and $\sum_\ell a_\ell P_i(\varphi_\ell) \geq k$, and leave the other ones as exercises.

Consider an arbitrary enriched probabilistic Kripke model (with binary experiments) $\mathbb{M} = \langle W, R_1, R_2, E_1, E_2, \mu_1, \mu_2, V \rangle$ and an arbitrary state $w \in W$. For $K_i\varphi$, we have:

$$
\begin{aligned}
\mathbb{M}, w \models [\text{EXP}]\,K_i\varphi \quad &\Leftrightarrow \quad \mathbb{M}^e, w \models K_i\varphi \\
&\Leftrightarrow \quad \forall v \in W^e : \text{if } (w,v) \in R_i^e \text{ then } \mathbb{M}^e, v \models \varphi \\
&\Leftrightarrow \quad \forall v \in W : \text{if } (w,v) \in R_i \cap E_i \text{ then } \mathbb{M}, v \models [\text{EXP}]\,\varphi \\
&\Leftrightarrow \quad \mathbb{M}, w \models R_i\,[\text{EXP}]\,\varphi
\end{aligned}
$$

Next, for $X^\varphi\psi$, note that $(R_i^e)^e = R_i^e \cap E_i^e = R_i \cap E_i \cap E_i = R_i \cap E_i = R_i^e$, and hence also $(R^e)^e = R^e$. We thus have

$$
\begin{aligned}
\mathbb{M}, w \models [\text{EXP}]\,X^\varphi\psi \quad &\Leftrightarrow \quad \mathbb{M}^e, w \models X^\varphi\psi \\
&\Leftrightarrow \quad \forall v \in W^e : (w,v) \in \big((R^e)^e \cap (W^e \times [\![\varphi]\!]^{\mathbb{M}^e})\big)^+ \Rightarrow \mathbb{M}^e, v \models \psi \\
&\Leftrightarrow \quad \forall v \in W : (w,v) \in \big(R^e \cap (W \times [\![[\text{EXP}]\,\varphi]\!]^{\mathbb{M}})\big)^+ \Rightarrow \mathbb{M}, v \models [\text{EXP}]\,\psi \\
&\Leftrightarrow \quad \mathbb{M}, w \models X^{[\text{EXP}]\,\varphi}\,[\text{EXP}]\,\psi
\end{aligned}
$$

For $[\text{EXP}]\,\varphi$, recall Lemma 11, which says that $(\mathbb{M}^e)^e = \mathbb{M}^e$; we thus get:

$$
\begin{aligned}
\mathbb{M}, w \models [\text{EXP}][\text{EXP}]\,\varphi \quad &\Leftrightarrow \quad \mathbb{M}^e, w \models [\text{EXP}]\,\varphi \\
&\Leftrightarrow \quad (\mathbb{M}^e)^e, w \models [\text{EXP}]\,\varphi \\
&\Leftrightarrow \quad \mathbb{M}^e, w \models \varphi \\
&\Leftrightarrow \quad \mathbb{M}, w \models [\text{EXP}]\,\varphi
\end{aligned}
$$

Finally, for $\sum_\ell a_\ell P_i(\varphi_\ell) \geq k$, we recall (5.6) and get:

$$\mathbb{M}, w \models [\text{EXP}]\sum_\ell a_\ell P_i(\varphi_\ell) \geq k$$

$$\Leftrightarrow \quad \mathbb{M}^e, w \models \sum_\ell a_\ell P_i(\varphi_\ell) \geq k$$

$$\Leftrightarrow \quad \sum_\ell a_\ell \mu_i^e(w)(\llbracket \varphi_\ell \rrbracket^{\mathbb{M}^e}) \geq k$$

$$\Leftrightarrow \quad \sum_\ell a_\ell \mu_i^e(w)(\llbracket [\text{EXP}]\,\varphi_\ell \rrbracket^{\mathbb{M}}) \geq k$$

$$\Leftrightarrow \quad \sum_\ell a_\ell \frac{\mu_i(w)(\llbracket [\text{EXP}]\,\varphi_\ell \rrbracket^{\mathbb{M}} \cap E_i[w])}{\mu_i(w)(E_i[w])} \geq k$$

$$\Leftrightarrow \quad \begin{cases} \text{if } \mathbb{M}, w \models \alpha_i \text{ then } \sum_\ell a_\ell \dfrac{\mu_i(w)(\llbracket [\text{EXP}]\,\varphi_\ell \rrbracket^{\mathbb{M}} \cap \llbracket \alpha_i \rrbracket^{\mathbb{M}})}{\mu_i(w)(\llbracket \alpha_i \rrbracket^{\mathbb{M}})} \geq k \\[3mm] \text{if } \mathbb{M}, w \models \neg\alpha_i \text{ then } \sum_\ell a_\ell \dfrac{\mu_i(w)(\llbracket [\text{EXP}]\,\varphi_\ell \rrbracket^{\mathbb{M}} \cap \llbracket \neg\alpha_i \rrbracket^{\mathbb{M}})}{\mu_i(w)(\llbracket \neg\alpha_i \rrbracket^{\mathbb{M}})} \geq k \end{cases}$$

$$\Leftrightarrow \quad \begin{cases} \text{if } \mathbb{M}, w \models \alpha_i \text{ then } \sum_\ell a_\ell \mu_i(w)(\llbracket [\text{EXP}]\,\varphi_\ell \rrbracket^{\mathbb{M}} \cap \llbracket \alpha_i \rrbracket^{\mathbb{M}}) \geq k\mu_i(w)(\llbracket \alpha_i \rrbracket^{\mathbb{M}}) \\[3mm] \text{if } \mathbb{M}, w \models \neg\alpha_i \text{ then } \sum_\ell a_\ell \mu_i(w)(\llbracket [\text{EXP}]\,\varphi_\ell \rrbracket^{\mathbb{M}} \cap \llbracket \neg\alpha_i \rrbracket^{\mathbb{M}}) \geq k\mu_i(w)(\llbracket \neg\alpha_i \rrbracket^{\mathbb{M}}) \end{cases}$$

$$\Leftrightarrow \quad \begin{aligned} \mathbb{M}, w \models \quad & \alpha_i \rightarrow \sum_\ell a_\ell P_i([\text{EXP}]\,\varphi_\ell \wedge \alpha_i) \geq kP_i(\alpha_i) \\ \wedge \quad & \neg\alpha_i \rightarrow \sum_\ell a_\ell P_i([\text{EXP}]\,\varphi_\ell \wedge \neg\alpha_i) \geq kP_i(\neg\alpha_i) \end{aligned}$$

$\square$

C.2 Completeness

Theorem 59, part 2. DPEALe is complete with respect to $\mathcal{AGR}$.

Proof. We have to show that $\mathcal{AGR} \models \varphi \Rightarrow \text{DPEALe} \vdash \varphi$ for all $\varphi \in \mathcal{L}([\text{EXP}])$. Let $\varphi \in \mathcal{L}([\text{EXP}])$ arbitrary; we proceed by contraposition. Suppose that DPEALe $\nvdash \varphi$; we will show that $\mathcal{AGR} \not\models \varphi$. Note that because we have reduction axioms for [EXP], there exists a sentence $\varphi' \in \mathcal{L}$ such that[1]

$$\text{DPEALe} \vdash \varphi \leftrightarrow \varphi' \tag{C.1}$$

Since DPEALe $\nvdash \varphi$, it follows that also DPEALe $\nvdash \varphi'$. Since PEAL is a subsystem of DPEALe, it follows that also PEAL $\nvdash \varphi'$. We showed in Appendix B that PEAL is complete with respect to $\mathcal{AGR}$, and hence we have that

$$\mathcal{AGR} \not\models \varphi' \tag{C.2}$$

---

[1]Proving this exactly involves introducing a new complexity measure $c(\varphi)$, which has the property that for any [EXP]-reduction axiom, the right-hand-side of the axiom has a strictly lower complexity than the left-hand-side; and then, proceeding by induction on this new complexity measure. For details, see, for example, [47, Section 7.4].

We proved in the first section of this appendix that DPEALe is sound with respect to $\mathcal{AGR}$ and hence it follows from (C.1) that $\mathcal{AGR} \models \varphi \leftrightarrow \varphi'$. Combining this with (C.2), we find that

$$\mathcal{AGR} \not\models \varphi$$

as desired. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Appendix D

Reduction axioms for $[\mathrm{DIAL}(\,\cdot\,)]$

Lemma 98. The following sentences are sound with respect to $\mathcal{AGR}$:

$$
\begin{aligned}
[\mathrm{DIAL}(\varphi)]p &\leftrightarrow p \quad (\text{for } p \in Prop) \\
[\mathrm{DIAL}(\varphi)]\neg\psi &\leftrightarrow \neg[\mathrm{DIAL}(\varphi)]\psi \\
[\mathrm{DIAL}(\varphi)](\psi \wedge \chi) &\leftrightarrow [\mathrm{DIAL}(\varphi)]\psi \wedge [\mathrm{DIAL}(\varphi)]\chi \\
[\mathrm{DIAL}(\varphi)]K_i\psi &\leftrightarrow K_i[\mathrm{DIAL}(\varphi)]\psi \\
[\mathrm{DIAL}(\varphi)]R_i\psi &\leftrightarrow R_i[\mathrm{DIAL}(\varphi)]\psi \\
[\mathrm{DIAL}(\varphi)]C^\psi\chi &\leftrightarrow C^{[\mathrm{DIAL}(\varphi)]\psi}[\mathrm{DIAL}(\varphi)]\chi \\
[\mathrm{DIAL}(\varphi)]X^\psi\chi &\leftrightarrow X^{[\mathrm{DIAL}(\varphi)]\psi}[\mathrm{DIAL}(\varphi)]\chi \\
[\mathrm{DIAL}(\varphi)]\textstyle\sum_\ell a_\ell P_i(\varphi_\ell) \geq k &\leftrightarrow \textstyle\sum_\ell a_\ell P_i([\mathrm{DIAL}(\varphi)]\varphi_\ell) \geq k
\end{aligned}
$$

Proof. We will explicitly prove the soundness of the reduction axioms for $K_i\psi$, $X^\psi\chi$ and $\sum_\ell a_\ell P_i(\varphi_\ell) \geq k$, and leave the other ones as exercises.

Consider an arbitrary agreement frame $\mathbb{F}$, let V be a valuation onto $\mathbb{F}$, and let w be a state of $\mathbb{F}$. Define $\mathbb{M} := \langle \mathbb{F}, V \rangle$. Let $\Phi(\mathbb{M}, w, \varphi)$ be the sentence guaranteed to exist by Lemma 29. This $\Phi(\mathbb{M}, w, \varphi)$ is a conjunction of probability formulas and probability formulas embedded in (sequences of) public announcements. By the reduction axioms for public announcement, it follows that $\Phi(\mathbb{M}, w, \varphi)$ is equivalent to a conjunction of probability formulas. Recall that $\mathsf{PEAL} \vdash \varphi \to C\varphi$ for i-probability formulas, by Lemma 48. Since PEAL is sound with respect to $\mathcal{AGR}$ (cf. Appendix B) it follows that $\mathbb{M}, w \models C\Phi(\mathbb{M}, w, \varphi)$, so

$$
R^*[w] \subseteq [\![\Phi(\mathbb{M}, w, \varphi)]\!]^{\mathbb{M}} \tag{D.1}
$$

Hence also

$$
R_i[w] \subseteq [\![\Phi(\mathbb{M}, w, \varphi)]\!]^{\mathbb{M}} \tag{D.2}
$$

and thus

$$
R_i[w] \cap [\![\Phi(\mathbb{M}, w, \varphi)]\!]^{\mathbb{M}} = R_i[w] \tag{D.3}
$$

Finally, it is easy to check componentwise that

$$\mathbb{M}^{dial_w(\varphi)} = \mathbb{M}^{\Phi(\mathbb{M},w,\varphi)} \qquad\qquad\qquad (\text{D.4})$$

Now for $K_i\psi$ we have that:

$$
\begin{aligned}
&\quad\ \mathbb{M}, w \models [\text{DIAL}(\varphi)]K_i\psi \\
\Leftrightarrow &\quad\ \mathbb{M}^{dial_w(\varphi)}, w \models K_i\psi \\
\Leftrightarrow &\quad\ \forall v \in W^{dial_w(\varphi)} : (w,v) \in R_i^{dial_w(\varphi)} \Rightarrow \mathbb{M}^{dial_w(\varphi)}, v \models \psi \\
\Leftrightarrow &\quad\ \forall v \in W^{\Phi(\mathbb{M},w,\varphi)} : (w,v) \in R_i^{\Phi(\mathbb{M},w,\varphi)} \Rightarrow \mathbb{M}^{dial_w(\varphi)}, v \models \psi &(\text{D.4}) \\
\Leftrightarrow &\quad\ \forall v \in W : (w,v) \in R_i \Rightarrow \big(w, v \in [\![\Phi(\mathbb{M},w,\varphi)]\!]^{\mathbb{M}} \Rightarrow \mathbb{M}, v \models [\text{DIAL}(\varphi)]\psi\big) &(\text{Def. 14}) \\
\Leftrightarrow &\quad\ \forall v \in W : (w,v) \in R_i \Rightarrow \mathbb{M}, v \models [\text{DIAL}(\varphi)]\psi &(\text{D.2}) \\
\Leftrightarrow &\quad\ \mathbb{M}, w \models K_i[\text{DIAL}(\varphi)]\psi
\end{aligned}
$$

For $X^\psi\chi$ we have that:

$$
\begin{aligned}
&\quad\ \mathbb{M}, w \models [\text{DIAL}(\varphi)]X^\psi\chi \\
\Leftrightarrow &\quad\ \mathbb{M}^{dial_w(\varphi)}, w \models X^\psi\chi \\
\Leftrightarrow &\quad\ \forall v \in W^{dial_w(\varphi)} : (w,v) \in \big((R_i^{dial_w(\varphi)})^e \cap (W^{dial_w(\varphi)} \cap [\![\psi]\!]^{\mathbb{M}^{dial_w(\varphi)}})\big)^+ \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \Rightarrow \mathbb{M}^{dial_w(\varphi)}, v \models \psi \\
\Leftrightarrow &\quad\ \forall v \in W^{\Phi(\mathbb{M},w,\varphi)} : (w,v) \in \big((R_i^{\Phi(\mathbb{M},w,\varphi)})^e \cap (W^{\Phi(\mathbb{M},w,\varphi)} \times [\![[\text{DIAL}(\varphi)]\psi]\!]^{\mathbb{M}})\big)^+ \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \Rightarrow \mathbb{M}^{dial_w(\varphi)}, v \models \chi &(\text{D.4}) \\
\Leftrightarrow &\quad\ \forall v \in W : (w,v) \in \big(R_i^{\Phi(\mathbb{M},w,\varphi)} \cap E_i^{\Phi(\mathbb{M},w,\varphi)} \cap ([\![\Phi(\mathbb{M},w,\varphi)]\!]^{\mathbb{M}} \times [\![[\text{DIAL}(\varphi)]\psi]\!]^{\mathbb{M}})\big)^+ \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \Rightarrow \mathbb{M}, v \models [\text{DIAL}(\varphi)]\chi &(\text{Def. 14}) \\
\Leftrightarrow &\quad\ \forall v \in W : (w,v) \in \big(R_i \cap E_i \cap (W \times [\![[\text{DIAL}(\varphi)]\psi]\!]^{\mathbb{M}})\big)^+ \Rightarrow \mathbb{M}, v \models [\text{DIAL}(\varphi)]\chi &(\text{D.1}) \\
\Leftrightarrow &\quad\ \mathbb{M}, w \models X^{[\text{DIAL}(\varphi)]\psi}[\text{DIAL}(\varphi)]\chi
\end{aligned}
$$

Finally, for $\sum_\ell a_\ell P_i(\varphi_\ell) \geq k$ we have that:

$$
\begin{aligned}
&\quad\ \mathbb{M}, w \models [\text{DIAL}(\varphi)]\textstyle\sum_\ell a_\ell P_i(\varphi_\ell) \geq k \\
\Leftrightarrow &\quad\ \mathbb{M}^{dial_w(\varphi)}, w \models \textstyle\sum_\ell a_\ell P_i(\varphi_\ell) \geq k \\
\Leftrightarrow &\quad\ \textstyle\sum_\ell a_\ell \mu_i^{dial_w(\varphi)}(w)([\![\varphi_\ell]\!]^{\mathbb{M}^{dial_w(\varphi)}}) \geq k \\
\Leftrightarrow &\quad\ \textstyle\sum_\ell a_\ell \mu_i^{\Phi(\mathbb{M},w,\varphi)}(w)([\![[\text{DIAL}(\varphi)]\varphi_\ell]\!]^{\mathbb{M}}) \geq k &(\text{D.4}) \\
\Leftrightarrow &\quad\ \textstyle\sum_\ell a_\ell \frac{\mu_i(w)([\![[\text{DIAL}(\varphi)]\varphi_\ell]\!]^{\mathbb{M}} \cap [\![\Phi(\mathbb{M},w,\varphi)]\!]^{\mathbb{M}})}{\mu_i(w)([\![\Phi(\mathbb{M},w,\varphi)]\!]^{\mathbb{M}})} \geq k &(\text{Def. 14}) \\
\Leftrightarrow &\quad\ \textstyle\sum_\ell a_\ell \frac{\mu_i(w)([\![[\text{DIAL}(\varphi)]\varphi_\ell]\!]^{\mathbb{M}} \cap [\![\Phi(\mathbb{M},w,\varphi)]\!]^{\mathbb{M}} \cap R_i[w])}{\mu_i(w)([\![\Phi(\mathbb{M},w,\varphi)]\!]^{\mathbb{M}} \cap R_i[w])} \geq k &(\text{Lemma 3}) \\
\Leftrightarrow &\quad\ \textstyle\sum_\ell a_\ell \frac{\mu_i(w)([\![[\text{DIAL}(\varphi)]\varphi_\ell]\!]^{\mathbb{M}} \cap R_i[w])}{\mu_i(w)(R_i[w])} \geq k &(\text{D.3}) \\
\Leftrightarrow &\quad\ \textstyle\sum_\ell a_\ell \frac{\mu_i(w)([\![[\text{DIAL}(\varphi)]\varphi_\ell]\!]^{\mathbb{M}})}{\mu_i(w)(W)} \geq k &(\text{Lemma 3}) \\
\Leftrightarrow &\quad\ \mathbb{M}, w \models \textstyle\sum_\ell a_\ell P_i([\text{DIAL}(\varphi)]\varphi_\ell) \geq k
\end{aligned}
$$

$\square$