CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Bern Open Repository and Information System (BORIS)

Justification Logics with Common Knowledge

Inauguraldissertation
der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von
Samuel Bucheli
von Buttisholz und Malters LU

Leiter der Arbeit:
Prof. Dr. G. Jäger
Institut für Informatik und angewandte Mathematik

Justification Logics with Common Knowledge

Inauguraldissertation
der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

Samuel Bucheli

von Buttisholz und Malters LU

Leiter der Arbeit:
Prof. Dr. G. Jäger
Institut für Informatik und angewandte Mathematik

Von der Philosophisch-naturwissenschaftlichen Fakultät angenommen.

Der Dekan:
Prof. Dr. S. Decurtins

Bern, 21. Mai 2012

Acknowledgements

First of all, I would like to thank Prof. Dr. Gerhard Jäger, who has made this thesis possible and whose guidance and advice I value very much.

Many thanks also to the members of the Logic and Theory group of the Institut für angewandte Mathematik und Informatik and the Logic Group of the Mathematisches Institut of the University of Bern, who have provided an inspiring environment. In particular, I would like to thank PD Dr. Thomas Studer and Dr. Roman Kuznets, who not only have proven to be excellent co-authors but also excellent friends.

A third thanks goes to my family and friends, who have always supported me.

The last but not least thanks goes to the lovely Ms. Andie Pilot whose proofreading has improved this thesis significantly. She shall be rewarded as requested.

The research for this thesis was supported by the Swiss National Science Foundation grant 200021–117699.

Contents

1. Introduction

> 'Contrariwise,' continued Tweedledee, 'if it was so, it might be; and if it were so, it would be; but as it isn't, it ain't. That's logic.'
>
> *Lewis Carroll,* Through the Looking Glass *[Car98]*

Logic explores the rules of (sound) reasoning. Epistemic logic investigates the notions of knowledge and belief formally, often using modal logic. An interesting case is the presence of several subjects of knowledge, usually called agents. In these multi-agent systems, the notion of *common knowledge* is crucial. Informally, a fact is common knowledge if everyone knows the fact and everyone knows that everyone knows the fact and everyone knows that everyone knows that everyone knows the fact and so on, ad infinitum. Knowledge formalized using modal logic could be described as implicit, as we can only state the fact of knowledge, but not the reason for this knowledge. This is contrasted by the situation in *justification logic*, where not only the fact of knowledge is stated, but also a reason—or justification—for this knowledge is given. There is a very close relationship between many modal logics and justification logics, i.e., for a given modal logic we can often name its justification counterpart. The aim of this thesis is to establish a justification counterpart to modal multi-agent systems with commmon knowledge.[1]

This thesis is organized as follows. In Chapter 2 we introduce the basic notions for epistemic modal logic and in Chapter 3 we introduce the concept of common knowledge, in particular we discuss several proof systems for common knowledge. In Chapter 4, we present a co-inductive proof system for common knowledge in detail. In Chapter 5,

[1]See the Chapters mentioned in the following paragraph for proper references to the notions mentioned here.

we introduce the basic notions of justification logics. Chapters 6 and 7 are the core part of this thesis. Here, several justification logics with common knowledge are introduced and their properties are investigated. In Chapter 8, we discuss open problems and further research possibilities. In Chapter A, we define the notion of filtration for justification logics, a technical tool that was used in previous chapters. Chapter B introduces justified public announcement logics, which are another interesting research topic in epistemic justification logics, particularly if combined with justification logics for common knowledge.

In order to keep the technical machinery lightweight, notations and definitions are usually valid only within the scope of the chapter and exceptions are usually marked with excplicit references to other chapters. Where indicated, chapters are very closely based on the corresponding publication.

2. Epistemic Modal Logic

> Three ladies, A, B, C in a railway carriage all have dirty faces and are all laughing. It suddenly flashes on A: Why doesn't B realize C is laughing at her? — Heavens! *I* must be laughable. (Formally: if I, A, am not laughable, B will be arguing: if I, B, am not laughable, C has nothing to laugh at. Since B does not so argue, I, A, must be laughable.
>
> *John Littlewood,* Littlewood's Miscellany *[Lit86]*

Modalities are used to qualify the truth of a statement [FM98; Gar09]. Examples for modalities are:

Modal logic "it is necessary that. . . " or "it is possible that. . . "

Epistemic logic "it is known that. . . "

Doxastic logic "it is believed that. . . "

Provability logic "it is provable that. . . "

Deontic logic "it is obligatory that. . . "

Temporal logic "it has always been the case that. . . " or "it will be the case that. . . "

In this thesis we will mainly consider *epistemic logics*, i.e. logics of knowledge and belief [Hin62; HS09]. For a general overview of the history of modal logics see [Gol03; Bal10], an excellent introduction to modal logic can be found in [BRV02] and [BBW07] provides a survey of current results.

2.1. Syntax

Formulae of modal logic are given by the following grammar:

$$A ::= P \mid \neg A \mid A \to A \mid \Box A \ ,$$

where P is a proposition from a fixed, countable set of propositions Prop. Furthermore, we define:

$$\begin{aligned} A \vee B &:= \neg A \to B \\ A \wedge B &:= \neg(\neg A \vee \neg B) \\ \bot &:= A \wedge \neg A \\ A \leftrightarrow B &:= (A \to B) \wedge (B \to A) \\ \Diamond A &:= \neg \Box \neg A \end{aligned}$$

We will denote formulae by $A, B, C, \ldots$.

The basic modal logic K is given by the following axioms:

A1 all propositional tautologies

(k) $\Box(A \to B) \to (\Box A \to \Box B)$

and rules

$$\frac{A \quad A \to B}{B}\,(\mathrm{MP})\ , \qquad \frac{\vdash A}{\Box A}\,(\mathsf{Nec})\ ,$$

We will consider extensions by the following axioms

(d) $\Box\bot \to \bot$ (consistency)

(t) $\Box A \to A$ (truth)

(4) $\Box A \to \Box\Box A$ (positive introspection)

(5) $\neg\Box A \to \Box\neg\Box A$ (negative introspection)

See Table 2.1 for the modal logics we will consider in the following. Note that the logics $\mathsf{KT4}$ and $\mathsf{KT45}$ are also often called $\mathsf{S4}$ and $\mathsf{S5}$, respectively for historical reasons [Bal10].

Given the epistemic interpretation of $\Box A$ as "A is known", we read the axioms as follows:

	A1	(k)	(d)	(t)	(4)	(5)	(MP)	(Nec)
K	✓	✓					✓	✓
KD	✓	✓	✓				✓	✓
KT	✓	✓		✓			✓	✓
K4	✓	✓			✓		✓	✓
K5	✓	✓				✓	✓	✓
KD4	✓	✓	✓		✓		✓	✓
KD5	✓	✓	✓			✓	✓	✓
KT4	✓	✓		✓	✓		✓	✓
KT5	✓	✓		✓		✓	✓	✓
K45	✓	✓			✓	✓	✓	✓
KD45	✓	✓	✓		✓	✓	✓	✓
KT45	✓	✓		✓	✓	✓	✓	✓

Table 2.1.: Modal Logics

(k) knowledge is closed under logical consequences

(d) knowledge is consistent

(t) knowledge is true

(4) if A is known, then it is known that A is known

(5) if A is not known, then it is known that A is not known

Often (t) is used to distinguish knowledge and belief: without (t) we speak about *belief*, with (t) we speak about *knowledge*. However, for the sake of readability, we will always also mean belief when we say knowledge, e.g. when we say common knowledge, we actually mean common belief or common knowledge.

2.2. Semantics

To provide the logics defined above with semantics, we use so called Kripke models [Kri59]. The idea here is to interpret "A is known" (or "A is necesary") by

A is true in all circumstances imaginable.

In order to formalize this, consider a non-empty set W where $w \in W$ is called a possible world, a relation $R \subseteq W \times W$ between these worlds called accessibility relation and a function $\nu \colon \mathrm{Prop} \to \mathcal{P}(W)$ called valuation, which assigns to each proposition P the worlds that it is valid in. We call the triple $\mathcal{M} = (W, R, \nu)$ a *Kripke model.*

For a Kripke model $\mathcal{M} = (W, R, \nu)$ we define the ternary *satisfaction relation* $\mathcal{M}, w \Vdash A$ between *model*, $w \in W$ and formulae by induction on the formula A as follows:

- $\mathcal{M}, w \Vdash P$ if and only if $w \in \nu(P)$
- $\mathcal{M}, w \Vdash \neg A$ if and only if $\mathcal{M}, w \nVdash A$
- $\mathcal{M}, w \Vdash A \to B$ if and only if $\mathcal{M}, w \Vdash B$ whenever $\mathcal{M}, w \Vdash A$
- $\mathcal{M}, w \Vdash \Box A$ if and only if $\mathcal{M}, v \Vdash A$ for all $v \in W$ such that $R(w, v)$.

We write $\mathcal{M} \Vdash A$ if $\mathcal{M}, w \Vdash A$ for all $w \in W$.

Example 2.1. Consider the following example that illustrates knowledge of the weather at your current location, but uncertainty about the weather in other locations (assuming you do not have access to weather forecasts or similar means). We look at two propositions "It is sunny in Bern" and "It is sunny in Calgary" (and their negations). Assume it is actually sunny in both Bern and Calgary and we are located in Bern. Hence, we consider it possible that it is not sunny in Calgary. On the other hand, we do not consider it possible that it is not sunny in Bern. We get the following model $\mathcal{M} = (W, R, \nu)$:

We have four worlds $W = \{a, b, c, d\}$ which are related by $R = \{(a, a), (a, b), (b, a), (b, b), (c, c), (c, d), (d, c), (d, d)\}$ and the valuation is given as follows:

$$\begin{aligned} \nu(\text{“It is sunny in Bern”}) &= \{a, b\} \\ \nu(\text{“It is sunny in Calgary”}) &= \{a, d\} \end{aligned}$$

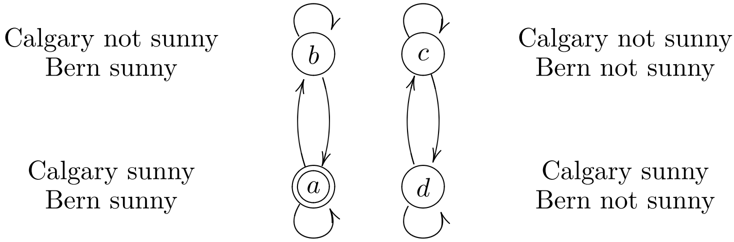

Figure 2.1.: A simple example of a Kripke model

See Figure 2.1 for a graphical representation of this model. We have

$$\begin{aligned} \mathcal{M}, a &\Vdash \Box\text{“It is sunny in Bern”} \\ \mathcal{M}, a &\Vdash \neg\Box\text{“It is sunny in Calgary”} \\ \mathcal{M}, a &\Vdash \neg\Box\neg\text{“It is sunny in Calgary”} \end{aligned}$$

as can be easily seen.

The basic modal logic K is sound and complete with respect to the class of all Kripke models. In order to establish soundness and completeness of the axiomatic extensions defined previously, we have to consider subclasses of the class of all Kripke models given by properties of the accessibility relation. We have the following correspondence between axioms and conditions on the accessibility relation

(d) R is serial, i.e., for all $w \in W$ there is a $v \in W$ such that $R(w, v)$

(t) R is reflexive, i.e., for all $w \in W$ we have $R(w, w)$

(4) R is transitive, i.e., for all $u, v, w \in W$ with $R(w, v)$ and $R(v, u)$ we have $R(w, u)$.

(5) R is Euclidean, i.e., for all $u, v, w \in W$ with $R(w, v)$ and $R(w, u)$ we have $R(v, u)$.

Let L be any of the logics defined previously. We define the class of all Kripke models for L, denoted $\mathcal{C}_{\mathsf{L}}$, to be the class of all Kripke models whose accessibility relation meets all conditions corresponding to the axioms of the logic L, e.g. the class $\mathcal{C}_{\mathsf{KT4}}$ of all Kripke models for $\mathsf{KT4}$ is the class of all Kripke models with reflexive and transitive accessibility relations.

The following theorem is standard. Soundness is established by an induction on the deriviation of the formula and completeness is obtained by a canonical model construction.

Theorem 2.2 (Soundness and completeness)**.** *Any logic* L *is sound and complete with respect to the class of its Kripke models, i.e.*

$$\mathsf{L} \vdash A \textit{ if and only if } \mathcal{M} \Vdash A \textit{ for all } \mathcal{M} \in \mathcal{C}_{\mathsf{L}}$$

This theorem can be strengthened and completeness with respect to classes of finite Kripke models can be shown and thus decidability of the logics is obtained.

2.3. Multi-Agent Systems

We obtain a variant of the systems presented above by using several modal operators $\Box_1, \Box_2, \ldots, \Box_h$ instead of one. Such multi-modal logics are also called a multi-agent systems, e.g. formulae $\Box_a \Box_b A$ can then be read as "agent a knows that agent b knows A". Usually, all agents will have the same modal strength, e.g. all $\Box_i$ will satisfy the axioms of $\mathsf{KT5}$. The notion of a model is also easily adapted to this case by using several accessibility relations instead of one. Let us have a look at a simple example of a multi-agent system.

Example 2.3. Consider a system of two agents. Agent 1 is located in Bern, agent 2 is located in Calgary. Analogous to the model from Example 2.1 we get the model displayed in Figure 2.2. The labels on the arrows indicate to which agent the accessibility relation is assigned. We now get e.g. that agent 1 knows that it is sunny in Bern, but also that agent 1 knows that agent 2 does not know it is sunny in Bern. Furthermore, agent 2 knows that agent 1 knows these facts, formally:

$$\begin{aligned}
\mathcal{M}, a &\Vdash \Box_1\text{"It is sunny in Bern"} \\
\mathcal{M}, a &\Vdash \Box_1\neg\Box_2\text{"It is sunny in Bern"} \\
\mathcal{M}, a &\Vdash \Box_2\Box_1\text{"It is sunny in Bern"} \\
\mathcal{M}, a &\Vdash \Box_2\Box_1\neg\Box_2\text{"It is sunny in Bern"}
\end{aligned}$$

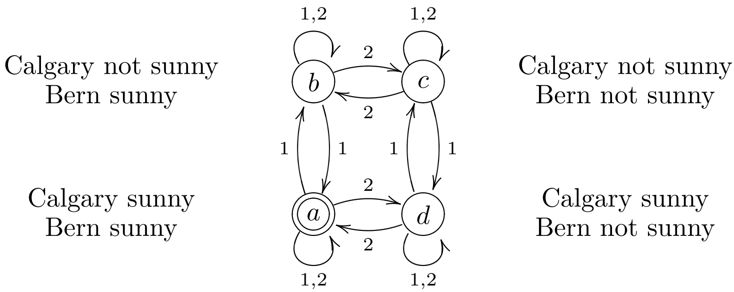

Figure 2.2.: A simple example of a Kripke model for two agents

We will give only this brief account of multi-agent systems here, as multi-agent systems will be a simple subsystem of the multi-agent systems *with common knowledge* introduced in the next chapter.

3. Common Knowledge

> Two men, who pull the oars of a boat, do it by an agreement or convention, though they have never given promises to each other
>
> *David Hume,* A Treatise of Human Nature *[Hum39]*

The notion of common knowledge is essential in the area of multi-agent systems [SLB09], where coordination among a set of agents is a central issue. Informally, common knowledge of a proposition A is defined as the infinitary conjunction *everybody knows A and everybody knows that everybody knows A and so on.*[1] This is equivalent to saying that common knowledge of A is the greatest fixed point of $\lambda X.$(*everybody knows A and everybody knows X*). The textbooks [Fag+95; MH95] provide excellent introductions to epistemic logics in general and common knowledge in particular. The standard approach to axiomatizing this property is by means of a co-closure axiom (see Definition 3.1) and the following induction rule (see, for instance, [Fag+95]):

$$\frac{A \to \mathsf{E}(A \wedge B)}{A \to \mathsf{C}B} \quad \text{(I-R1)}$$

For further views on common knowledge, see also [VS09] and [Bar88].

We now introduce syntax and semantics for common knowledge, the latter based on Kripke models. Furthermore we present two different Hilbert-style axiomatizations for common knowledge and investigate their relationship as published in [BKS10b]. Finally, we present a selection of sequent-style proof systems for common knowledge and discuss known restrictions.

[1] This view is attributed to Lewis [Lew69].

3.1. Syntax and Semantics

We consider a language with h agents for some $h > 0$. This language will be fixed throughout this chapter, and h will always denote the number of agents. We are given a countable set Prop of propositional variables P. Propositions P and their negations $\overline{P}$ are atoms.

To facilitate the proof-theoretic treatment, formulae are given in negation normal form, i.e. negations only occur in front of atoms. We denote formulae using A, B, C. They are defined by the following grammar:

$$A ::= P \mid \overline{P} \mid A \wedge A \mid A \vee A \mid \Box_i A \mid \Diamond_i A \mid \mathsf{C}A \mid \tilde{\mathsf{C}}A \ ,$$

where $1 \leq i \leq h$. The formula $\Box_i A$ is read as *agent i knows A*, and the formula $\mathsf{C}A$ is read as *A is common knowledge*. The connectives $\Box_i$ and C have $\Diamond_i$ and $\tilde{\mathsf{C}}$ as their respective duals. The negation $\neg A$ of a formula A is defined in the usual way by using De Morgan's laws, the law of double negation, and the duality laws for modal operators, i.e. we inductively define

$$\begin{array}{rcl} \neg P & := & \overline{P} \\ \neg \overline{P} & := & P \\ \neg A \wedge B & := & \neg A \vee \neg B \\ \neg A \vee B & := & \neg A \wedge \neg B \\ \neg \Box_i A & := & \Diamond_i \neg A \\ \neg \Diamond_i A & := & \Box_i \neg A \\ \neg \mathsf{C}A & := & \tilde{\mathsf{C}} \neg A \\ \neg \tilde{\mathsf{C}}A & := & \mathsf{C} \neg A \end{array}$$

We also define

$$\begin{array}{rcl} A \to B & := & \neg A \vee B \\ A \leftrightarrow B & := & (A \to B) \wedge (B \to A). \end{array}$$

The formula $\mathsf{E}A$ is an abbreviation for *everybody knows A*:

$$\mathsf{E}A := \Box_1 A \wedge \cdots \wedge \Box_h A \quad \text{and} \quad \tilde{\mathsf{E}}A := \Diamond_1 A \vee \cdots \vee \Diamond_h A \ .$$

A *Kripke model* $\mathcal{M}$ is a tuple $(W, R_1, \ldots, R_h, \nu)$, where W is a non-empty set of worlds, each R_i is a binary relation on W, and $\nu : \mathrm{Prop} \to \mathcal{P}(W)$ is a valuation function that assigns to each proposition a set of worlds.

Let $\mathcal{M} = (W, R_1 \ldots, R_h, \nu)$ be a Kripke model and $v, w \in W$ two worlds. We say that *v is reachable from w in n steps*, denoted $\mathsf{reach}(w, v, n)$, if there exist worlds $v_0, \ldots, v_n$ such that $v_0 = w$, $v_n = v$, and for all $0 \leq j \leq n-1$ there exists $1 \leq i \leq h$ with $R_i(s_j, s_{j+1})$. We say *v is reachable from w* if there exists an n with $\mathsf{reach}(w, v, n)$.

Let $\mathcal{M} = (W, R_1 \ldots, R_h, \nu)$ be a Kripke model and $w \in S$ be a world. We define the *satisfaction relation* $\mathcal{M}, w \Vdash A$ inductively on the structure of the formula A:

$$
\begin{array}{ll}
\mathcal{M}, w \Vdash P & \text{if } w \in \nu(P), \\
\mathcal{M}, w \Vdash \overline{P} & \text{if } w \in W \setminus \nu(P), \\
\mathcal{M}, w \Vdash A \wedge B & \text{if } \mathcal{M}, w \Vdash A \text{ and } \mathcal{M}, w \Vdash B, \\
\mathcal{M}, w \Vdash A \vee B & \text{if } \mathcal{M}, w \Vdash A \text{ or } \mathcal{M}, w \Vdash B, \\
\mathcal{M}, w \Vdash \Box_i A & \text{if } \mathcal{M}, v \Vdash A \text{ for all } v \text{ such that } R_i(w, v), \\
\mathcal{M}, w \Vdash \Diamond_i A & \text{if } \mathcal{M}, v \Vdash A \text{ for some } v \text{ with } R_i(w, v), \\
\mathcal{M}, w \Vdash \mathsf{C} A & \text{if } \mathcal{M}, v \Vdash A \text{ for all } v \text{ such that} \\
& \quad \text{there is an } n \geq 1 \text{ with } \mathsf{reach}(w, v, n), \\
\mathcal{M}, w \Vdash \tilde{\mathsf{C}} A & \text{if } \mathcal{M}, v \Vdash A \text{ for some } v \text{ with} \\
& \quad \mathsf{reach}(w, v, n) \text{ for some } n \geq 1.
\end{array}
$$

We write $\mathcal{M} \Vdash A$ if $\mathcal{M}, w \Vdash A$ for all $w \in W$. A formula A is called *valid*, denoted $\Vdash A$, if $\mathcal{M} \Vdash A$ for all Kripke models $\mathcal{M}$. A formula A is called *satisfiable* if $\mathcal{M}, w \Vdash A$ for some Kripke model $\mathcal{M}$ and some world w.

3.2. Hilbert-Style Axiomatizations

We will now present two Hilbert-style axiomatizations and prove their equivalence. The first axiomatization $\mathsf{H_R}$ is based on an induction rule and can be found in [Fag+95].

Definition 3.1 (The system $\mathsf{H_R}$)**.** The Hilbert calculus $\mathsf{H_R}$ for the logic of common knowledge is defined by the following axioms and inference rules:

Propositional axioms: All instances of propositional tautologies

Modus ponens: For all formulae A and B,

$$\frac{A \quad A \to B}{B} \quad \text{(MP)}$$

Modal axioms: For all formulae A and B and all indices $1 \leq i \leq h$,

$$\Box_i(A \to B) \to (\Box_i A \to \Box_i B) \quad (\mathsf{K})$$

Necessitation rule: For all formulae A and all indices $1 \leq i \leq h$,

$$\frac{A}{\Box_i A} \quad (\mathsf{Nec})$$

Co-closure axiom: For all formulae A,

$$\mathsf{C}A \to \mathsf{E}(A \land \mathsf{C}A) \quad (\mathsf{Co\text{-}Cl})$$

Induction rule: For all formulae A and B,

$$\frac{B \to \mathsf{E}(A \land B)}{B \to \mathsf{C}A} \quad (\mathsf{I\text{-}R1})$$

We have the following standard result using a canonical model construction, see [Fag+95].

Theorem 3.2 (Soundness and completeness of $\mathsf{H_R}$)**.** *For any formula A,*

$$\mathsf{H_R} \vdash A \quad \textit{if and only if} \quad A \textit{ is valid.}$$

We now introduce a deductive system for common knowledge where the induction rule is replaced by an induction axiom. The system is based on the system presented in [MH95], where an induction axiom is introduced as $A \land \mathsf{C}(A \to \mathsf{E}A) \to \mathsf{C}A$. However, in our setting, the axiom from [MH95] would not be sound since we do not define common knowledge to be reflexive. To obtain a complete system, we also need to include a normality axiom and a necessitation rule for the common knowledge operator.

Definition 3.3 (The system H_{Ax})**.** The Hilbert calculus H_{Ax} consists of the axioms and rules of H_{R} whereby (I-R1) is replaced by the following axioms and rule:

C-modal axiom: For all formulae A and B,

$$\mathsf{C}(A \to B) \to (\mathsf{C}A \to \mathsf{C}B) \quad \text{(C-K)}$$

C-necessitation rule: For all formulae A,

$$\frac{A}{\mathsf{C}A} \quad \text{(C-Nec)}$$

Induction axiom: For all formulae A,

$$\mathsf{E}A \wedge \mathsf{C}(A \to \mathsf{E}A) \to \mathsf{C}A \quad \text{(I-Ax)}$$

The soundness of H_{Ax} is easily obtained.

Theorem 3.4 (Soundness)**.** *For any formula A, if $\mathsf{H}_{\mathsf{Ax}} \vdash A$, then A is valid.*

Proof. We show soundness as usual by induction on the length of the derivation of $\mathsf{H}_{\mathsf{Ax}} \vdash A$. We only consider the case where A is the induction axiom. Let $\mathcal{M}$ be a Kripke model. We show by induction on n that for all $n \geq 1$, if $\mathcal{M}, w \Vdash \mathsf{E}A \wedge \mathsf{C}(A \to \mathsf{E}A)$, then for all worlds v with $\mathsf{reach}(w, v, n)$, we have $\mathcal{M}, v \Vdash A$. If $n = 1$, then $\mathcal{M}, w \Vdash \mathsf{E}A$ guarantees $\mathcal{M}, v \Vdash A$. For $n = m+1$, $m \geq 1$, let w be such that $\mathsf{reach}(w, v, n)$. Then there exists v' such that

(1) $\mathsf{reach}(w, v', m)$ and

(2) $\mathsf{reach}(v', v, 1)$.

From (1) and $\mathcal{M}, w \Vdash \mathsf{C}(A \to \mathsf{E}A)$ we obtain $\mathcal{M}, v' \Vdash A \to \mathsf{E}A$. By the induction hypothesis, we get $\mathcal{M}, v' \Vdash A$. Therefore, $\mathcal{M}, v' \Vdash \mathsf{E}A$. Thus, by (2), we get $\mathcal{M}, v \Vdash A$. □

In order to establish the completeness of H_{Ax}, we introduce an intermediate system $\mathsf{H}_{\mathsf{int}}$. We first reduce H_{R} to $\mathsf{H}_{\mathsf{int}}$ and then reduce $\mathsf{H}_{\mathsf{int}}$ to H_{Ax}. The completeness of H_{R} then implies the completeness of H_{Ax}.

Definition 3.5 (The system $\mathsf{H_{int}}$)**.** $\mathsf{H_{int}}$ consists of the axioms and rules of $\mathsf{H_R}$ whereby (I-R1) is replaced by the following axiom and rule:

C-distributivity: For all formulae A and B,

$$\mathsf{C}(A \wedge B) \to (\mathsf{C}A \wedge \mathsf{C}B) \quad \text{(C-Dis)}$$

Induction rule 2: For all formulae A,

$$\frac{A \to \mathsf{E}A}{\mathsf{E}A \to \mathsf{C}A} \quad \text{(I-R2)}$$

Lemma 3.6. *For each formula A, we have that $\mathsf{H_R} \vdash A$ implies $\mathsf{H_{int}} \vdash A$.*

Proof. It is sufficient to show that (I-R1) is derivable in $\mathsf{H_{int}}$. Assume

$$\mathsf{H_{int}} \vdash B \to \mathsf{E}(A \wedge B) \ . \tag{3.1}$$

Then $\mathsf{H_{int}} \vdash A \wedge B \to \mathsf{E}(A \wedge B)$. By (I-R2), we obtain that

$$\mathsf{H_{int}} \vdash \mathsf{E}(A \wedge B) \to \mathsf{C}(A \wedge B) \ .$$

Using (C-Dis), we get $\mathsf{H_{int}} \vdash \mathsf{E}(A \wedge B) \to \mathsf{C}A$. Finally, (3.1) yields $\mathsf{H_{int}} \vdash B \to \mathsf{C}A$, which completes the proof. □

Lemma 3.7. *For each formula A, we have that $\mathsf{H_{int}} \vdash A$ implies $\mathsf{H_{Ax}} \vdash A$.*

Proof. We first show that (C-Dis) is derivable in $\mathsf{H_{Ax}}$. The following formula is an instance of (C-K):

$$\mathsf{H_{Ax}} \vdash \mathsf{C}(A \wedge B \to B) \to (\mathsf{C}(A \wedge B) \to \mathsf{C}B) \ . \tag{3.2}$$

$\mathsf{H_{Ax}} \vdash A \wedge B \to B$ is a propositional axiom. By (C-Nec), $\mathsf{H_{Ax}} \vdash \mathsf{C}(A \wedge B \to B)$. By (3.2), we have $\mathsf{H_{Ax}} \vdash \mathsf{C}(A \wedge B) \to \mathsf{C}B$. A similar argument yields $\mathsf{H_{Ax}} \vdash \mathsf{C}(A \wedge B) \to \mathsf{C}A$. The last two statements together imply that (C-Dis) is derivable in $\mathsf{H_{Ax}}$.

It remains to show that (I-R2) is derivable in $\mathsf{H_{Ax}}$. Assume that $\mathsf{H_{Ax}} \vdash A \to \mathsf{E}A$. By (C-Nec), we get $\mathsf{H_{Ax}} \vdash \mathsf{C}(A \to \mathsf{E}A)$. Thus, the derivability of (I-R2) follows from (I-Ax). □

These two lemmas, together with the completeness of $\mathsf{H_R}$, give us the completeness of $\mathsf{H_{Ax}}$.

Corollary 3.8 (Completeness of $\mathsf{H_{Ax}}$)**.** *For all formulae A, if A is valid, then $\mathsf{H_{Ax}} \vdash A$.*

3.3. A Survey of Proof Systems

Besides the Hilbert-style axiomatizations from the previous section, different sequent systems for common knowledge have been presented. We are now giving a brief survey of existing sequent proof systems for common knowledge. A main aim of the research presented below was to find finitary, cut-free sequent systems that have desirable properties from a proof-theoretic perspective. Even though big progress has been made, no completely satisfactory solution has been found yet. As discussed in the next section, it might even be impossible to find a solution with all desired properties. We will only give a short description of the systems; for full details refer to the publications mentioned throughout this section.

A sequent is a finite set of formulae denoted by Γ, Δ, Σ. For a sequent $\Delta = \{A_1, \dots, A_n\}$, we denote the sequent $\{\Diamond_i A_1, \dots, \Diamond_i A_n\}$ by $\Diamond_i \Delta$ and the sequent $\{\tilde{\mathsf{E}} A_1, \dots, \tilde{\mathsf{E}} A_n\}$ by $\tilde{\mathsf{E}}\Delta$. In addition, $\mathcal{M}, w \Vdash \Delta$ is understood as $\mathcal{M}, w \Vdash A_1 \vee \dots \vee A_n$.

In [AJ05], Alberruci and Jäger present a Tait-style system $\overline{\mathrm{K}}_h(\mathsf{C})$:

Axioms and basic rules of inference

$$P, \overline{P}, \Gamma \;\; (\mathrm{ID}) \qquad \frac{A, B, \Gamma}{A \vee B, \Gamma} \;\; (\vee) \qquad \frac{A, \Gamma \qquad B, \Gamma}{A \wedge B, \Gamma} \;\; (\wedge)$$

Individual agents

$$\frac{A, \Gamma, \tilde{\mathsf{C}}\Delta}{\Box_i A, \Diamond_i \Gamma, \tilde{\mathsf{C}}\Delta, \Pi} \;\; (\Box_i)$$

Common knowledge

$$\frac{\tilde{\mathsf{E}} A, \Gamma}{\tilde{\mathsf{C}} A, \Gamma} \;\; (\tilde{\mathsf{C}}) \qquad \frac{\mathsf{E} A, \tilde{\mathsf{C}}\Delta}{\mathsf{C} A, \tilde{\mathsf{C}}\Delta, \Pi} \;\; (\mathsf{C})$$

Induction rule

$$\frac{\neg B, \mathsf{E} A, \tilde{\mathsf{C}}\Delta \qquad \neg A, \mathsf{E} B, \tilde{\mathsf{C}}\Delta}{\neg A, \mathsf{C} A, \tilde{\mathsf{C}}\Delta, \Pi} \;\; (\mathrm{Ind})$$

Cut

$$\frac{A, \Gamma \qquad \neg A, \Gamma}{\Gamma} \;\; (\mathrm{Cut})$$

They prove the soundness and completeness of the system with respect to standard Kripke-style semantics and furthermore show that the formulae allowed in the (Cut)-rule can be restricted to a certain form, namely disjunctive-conjunctive closure of the Fischer-Ladner closure of the sequent to be proved. However, full cut elimination is not possible.

Furthermore, in [AJ05], an infinitary system $\mathrm{K}_h^\omega(\mathsf{C})$ using an ω-rule for common knowledge is introduced, and soundness and completeness are shown:

Axioms and basic rules of inference

$$P, \overline{P}, \Gamma \quad (\mathrm{ID}) \qquad \frac{A, B, \Gamma}{A \vee B, \Gamma} \quad (\vee) \qquad \frac{A, \Gamma \qquad B, \Gamma}{A \wedge B, \Gamma} \quad (\wedge)$$

Individual agents

$$\frac{A, \Gamma, \tilde{\mathsf{C}}\Delta}{\Box_i A, \Diamond_i \Gamma, \tilde{\mathsf{C}}\Delta, \Pi} \quad (\Box_i)$$

Common knowledge

$$\frac{\tilde{\mathsf{E}}A, \Gamma}{\tilde{\mathsf{C}}A, \Gamma} \quad (\tilde{\mathsf{C}}) \qquad \frac{\mathsf{E}^m A, \Gamma \text{ for all } m \geq 1}{\mathsf{C}A, \Gamma} \quad (\omega\mathsf{C})$$

The infinitary system $\mathrm{K}_h^\omega(\mathsf{C})$ can be finitized using the finite model property of common knowledge, as shown in [JKS07]. In order to state the finite model property we first need to define the length of a formula.

Definition 3.9 (length of a formula)**.** Let A be a formula. Denote its length $\ell(A)$, given by:

- $\ell(P) := \ell(\overline{P}) := 1$,
- $\ell(B \wedge C) := \ell(B \vee C) := \ell(B) + \ell(C)$,
- $\ell(\mathsf{E}B) := \ell(\tilde{\mathsf{E}}B) := 1 + \ell(B)$,
- $\ell(\mathsf{C}B) := \ell(\tilde{\mathsf{C}}B) := \ell(B) \cdot h + h + 1$.

The finite model property says that a satisfiable formula is satisfiable in a finite model whose size is bound by the length of the formula. See e.g. [Fag+95; JKS07] for a proof of this theorem.

Theorem 3.10 (finite model property)**.** *If A is satisfiable, then there exists a model $\mathcal{M}$ and world w in that model, such that $\mathcal{M}, w \Vdash A$ and card($\mathcal{M}$) $\leq 2^{\ell(A)}$.*

The following corollary is easy to see. Given a formula $\mathsf{C}A$ that is not valid, we find a finite model as above witnessing this fact. Obviously now, there must be a reachable world within the given boundaries (depending on the size of the model and hence on the length of the formula) in which it does not hold.

Corollary 3.11. *If, for all $1 \leq m \leq 2^{\ell(\mathsf{C}A)+\ell(B_1)+\ldots+\ell(B_n)}$, we have*

$$\Vdash \mathsf{E}^m A \vee B_1 \vee \ldots \vee B_n$$

then also

$$\Vdash \mathsf{C}A \vee B_1 \vee \ldots \vee B_n$$

Based on this observation, the following finitized Tait-style system $\mathrm{K}_h^{<\omega}(\mathsf{C})$ is introduced in [JKS07] and again soundness and completeness are proved:

Axioms and basic rules of inference

$$P, \overline{P}, \Gamma \ \text{(ID)} \qquad \frac{A, B, \Gamma}{A \vee B, \Gamma} \ (\vee) \qquad \frac{A, \Gamma \qquad B, \Gamma}{A \wedge B, \Gamma} \ (\wedge)$$

Individual agents

$$\frac{A, \Gamma, \tilde{\mathsf{C}}\Delta}{\Box_i A, \Diamond_i \Gamma, \tilde{\mathsf{C}}\Delta, \Pi} \ (\Box_i)$$

Common knowledge

$$\frac{\tilde{\mathsf{E}}A, \Gamma}{\tilde{\mathsf{C}}A, \Gamma} \ (\tilde{\mathsf{C}})$$

$$\frac{\mathsf{E}^m A, \Gamma \text{ for all } 1 \leq m \leq 2^{\ell(\mathsf{C}A)+\sum_{B\in\Gamma}\ell(B)}}{\mathsf{C}A, \Gamma, \Pi} \ (<\omega\mathsf{C})$$

Even though this system is finitary and cut-free, the authors describe it as “somewhat unusual”, raising questions about structural properties and syntactic cut-elimination.

In [BS09], Brünnler and Studer introduce a nested sequent version of $\mathrm{K}^{\omega}_{h}(\mathsf{C})$. Nested sequents have an additional structural connective [.] corresponding to $\Box$ in a similar manner that commas in sequents correspond to disjunction. Roughly speaking, nested sequents allow sequent rules to be applied within formulae:[2]

Axioms and basic rules of inference

$$\Gamma\{P, \overline{P}\} \qquad \frac{\Gamma\{A, B\}}{\Gamma\{A \vee B\}}\ (\vee) \qquad \frac{\Gamma\{A\} \qquad \Gamma\{B\}}{\Gamma\{A \wedge B\}}\ (\wedge)$$

Individual agents

$$\frac{\Gamma\{[A]_i\}}{\Gamma\{\Box_i A\}}\ (\Box_i) \qquad \frac{\Gamma\{\Diamond_i A, [\Delta, A]_i\}}{\Gamma\{\Diamond_i A, [\Delta]_i\}}\ (\Diamond_i)$$

Common knowledge

$$\frac{\Gamma\{\tilde{\mathsf{C}}A, \tilde{\mathsf{E}}^k A\}}{\Gamma\{\tilde{\mathsf{C}}A\}}\ (\tilde{\mathsf{C}}) \qquad \frac{\Gamma\{\mathsf{E}^m A\} \quad (\text{for all } m \geq 1)}{\Gamma\{\mathsf{C}A\}}\ (\mathsf{C})$$

Cut rule

$$\frac{\Gamma\{A\} \qquad \Gamma\{\neg A\}}{\Gamma\{\emptyset\}}\ (\text{cut})$$

This system allows a syntactic cut-elimination procedure and there is an embedding from D_{C} into $\mathrm{K}^{\omega}_{h}(\mathsf{C})$ and vice versa, thus also yielding cut-elimination for the latter system.

Furthermore, in [Weh10] an annotated sequent system, which is an adpation of the system for temporal logics presented in [BL08] to common knowledge, is introduced. This system seems to be a finitary version of the system presented in the next chapter, but no formal comparison has been presented yet.

Finally, we also mention [AGW07] where a cut-free tableaux system is presented, which is mainly aimed towards implementations and thus has a different focus than the systems presented above.

[2]We will not present all formalities necessary to introduce nested sequent systems here. See [Brü10] for a survey of nested sequents in modal logic and the recent [Fit12] that shows that nested sequents are notational variants of prefixed tableaus in the same way that classical semantic tableaus are notational variants of classical Gentzen sequent calculae.

3.4. Known Restrictions

A reason why it might be hard to come up with a satisfactory proof system for common knowledge was found by Studer in [Stu09]. Let X, Y be propositional variables and $A(\vec{P}, X)$ a formula such that A contains at most the displayed propositional variables $\vec{P}, X = P_1, \ldots, P_n, X$. A logic has the Beth property (B1) [Bet53] if implicit definitions can be made explicit, i.e. whenever

$$\Vdash A(\vec{P}, X) \land A(\vec{P}, Y) \to (X \leftrightarrow Y),$$

then there exists a formula $B(\vec{P})$ such that

$$\Vdash A(\vec{P}, X) \to (X \leftrightarrow B(\vec{P})).$$

The Beth property is implied by the Craig Interpolation Property CIP [Cra57]. The latter can be stated as follows. Let $A(\vec{P}, \vec{Q})$ and $B(\vec{P}, \vec{R})$ be formulae. If

$$\Vdash A(\vec{P}, \vec{Q}) \to B(\vec{P}, \vec{R})$$

then there exists a formula $C(\vec{P})$ that contains only common propositional variables of $A(\vec{P}, \vec{Q})$ and $B(\vec{P}, \vec{R})$ such that

$$\Vdash A(\vec{P}, \vec{Q}) \to C(\vec{P}) \text{ and } C(\vec{P}) \to B(\vec{P}, \vec{R}).$$

For many logics, Craig Interpolation is a consequence of cut elimination, see e.g. [ST96]. In [Stu09], Studer shows that common knowledge (over arbitrary frames or over transitive frames with at least two agents) lacks the Beth property (B1) and hence also Craig interpolation (CIP). This might indicate that a finitary, cut-free sequent system for common knowledge with the usual desirable properties might not exist.

3.5. Variants and Axiomatic Extensions

We will now introduce variants and axiomatic extensions of the logic of common knowledge presented above. These variants will be used in Chapters 6 and 7. The first variant we present uses a stronger base logic, namely $\mathsf{S4}$. The second variant allows us to speak about groups of agents. The latter will also be presented with different base logics.

The first system we are presenting is $\mathsf{S4}_h^{\mathsf{C}}$, a system of h agents, and uses the same language as the system for common knowledge presented previously. The Hilbert system $\mathsf{S4}_h^{\mathsf{C}}$ is given by the following axioms:

Propositional axioms All instances of propositional tautologies.

Modal axioms for agents For any agent $i \in \{1, \ldots, h\}$

$$\Box_i(A \to B) \to (\Box_i A \to \Box_i B),$$
$$\Box_i A \to A, \qquad \Box_i A \to \Box_i \Box_i A.$$

Axioms for common knowlege

$$\mathsf{C}(A \to B) \to (\mathsf{C}A \to \mathsf{C}B),$$
$$A \wedge \mathsf{C}(A \to \mathsf{E}A) \to \mathsf{C}A, \qquad \mathsf{C}A \to \mathsf{E}(A \wedge \mathsf{C}A).$$

and rules:

$$\frac{A \quad A \to B}{B}\ (\mathrm{MP}), \qquad \frac{A}{\Box_i A}\ (\mathsf{Nec}), \qquad \frac{A}{\mathsf{C}A}\ (\mathsf{C\text{-}Nec})$$

for all agents $i \in \{1, \ldots, h\}$.

Soundness and completeness for this system with respect to the usual Kripke-style semantics (with reflexive and transitive accessibility relations for individual agents) is established as previously by induction on the derivation and a canonical model construction, respectively, see [Fag+95; MH95].

In order to deal with groups of agents, we need to extend our language. By *a group of agents* $\mathsf{G} = \langle i_1, \ldots, i_k \rangle$ we mean a (non-empty) tuple of $i_j \in \{1, \ldots, h\}$ with $i_1 < i_2 < \ldots < i_k$.[3] We will use set-notation $i_j \in \mathsf{G}$ and $\mathsf{G} \subseteq \mathsf{H}$ to state that i_j occurs in G and all $i_j \in \mathsf{G}$ occur in H, respectively.

Formulae in the language of common knowledge with groups of agents are given by the following grammar:

$$A ::= P_j \mid \neg A \mid (A \to A) \mid \Box_i A \mid \mathsf{C}_{\mathsf{G}} A \ ,$$

[3] Using (ordered) tuples instead of sets is not necessary for the modal language but will facilitate notations and technical details in the case of justification logic in Chapter 7.

where $P_j \in \text{Prop}$ and G is a group of agents.

Furthermore, for a group of agents $\mathsf{G} = \langle i_1, \ldots, i_k \rangle$, we define

$$\mathsf{E_G}A := \Box_{i_1} A \wedge \ldots \wedge \Box_{i_k} A \ .$$

The Hilbert system $\mathsf{K}_h^{\mathsf{C}}$ is given by the following axioms:

Propositional axioms All instances of propositional tautologies.

Modal axioms for agents For any agent $i \in \{1, \ldots, h\}$

$$\Box_i(A \to B) \to (\Box_i A \to \Box_i B). \tag{k}$$

Modal axiom for common knowledge For any group of agents G

$$\mathsf{C_G}(A \to B) \to (\mathsf{C_G}A \to \mathsf{C_G}B). \tag{C-k}$$

Group restriction axioms For any groups of agents $\mathsf{G}_2 \subseteq \mathsf{G}_1$

$$\mathsf{C}_{\mathsf{G}_1} A \to \mathsf{C}_{\mathsf{G}_2} A. \tag{res}$$

Co-closure axioms For any group of agents G

$$\mathsf{C_G}A \to \mathsf{E_G}(A \wedge \mathsf{C_G}A). \tag{ccl}$$

Induction axioms For any group of agents G

$$\mathsf{C_G}(A \to \mathsf{E_G}A) \to (\mathsf{E_G}A \to \mathsf{C_G}A). \tag{ind}$$

and rules

$$\frac{A \quad A \to B}{B}\ (\text{MP}) \qquad \frac{A}{\Box_i A}\ (\mathsf{Nec}) \qquad \frac{A}{\mathsf{C_H}A}\ (\mathsf{C\text{-}Nec})$$

for all agents $i \in \{1, \ldots, h\}$ and for the group of all agents $H = \langle 1, \ldots, h \rangle$.

We will also consider several extensions of $\mathsf{K}_h^{\mathsf{C}}$ given in table 3.5 using the axioms from table 3.5 for each agent $i \in \{1, \ldots, h\}$.

Note that the rule (Nec) can be easily derived using (C-Nec), (ccl) and propositional reasoning, but was included in order to keep the presentation in line with the system from [MH95]. However, the system from [MH95] does not use groups of agents, whereas the system from [Fag+95] does and uses the induction rule instead of the axiom. Hence, some remarks on the setup of the above logics are required. Note first that C-necessitation for the group of all agents is sufficient:

(d)	$\Box_i \bot \to \bot$
(t)	$\Box_i A \to A$
(4)	$\Box_i A \to \Box_i \Box_i A$
(5)	$\neg \Box_i A \to \Box_i (\neg \Box_i A)$

Table 3.1.: Additional Axioms

Lemma 3.12. *In any of the logics* $\mathsf{L}_h^{\mathsf{C}}$ *defined above,* C*-necessitation for arbitrary groups of agents is derivable.*

Proof. Let $\mathsf{H} = \langle 1, \dots, h \rangle$ be the group of all agents and $\mathsf{G} \subsetneq \mathsf{H}$. Suppose the formula A has already been derived. Then

1. $\mathsf{C}_\mathsf{H} A$ by (C-Nec)
2. $\mathsf{C}_\mathsf{H} A \to \mathsf{C}_\mathsf{G} A$ by (res)
3. $\mathsf{C}_\mathsf{G} A$ by (MP)

□

On the other hand, we could also have included C-necessitation for arbitrary groups and would thus have been able to omit the group restriction axiom (res).

Lemma 3.13. *The group restriction axiom (res) is derivable from the* C*-necessitation rule for arbitrary groups.*

Proof. Let $\mathsf{G}_1 \subseteq \mathsf{G}_2$ be two groups of agents.

1. $\mathsf{C}_{\mathsf{G}_2} A \to \mathsf{E}_{\mathsf{G}_2} \mathsf{C}_{\mathsf{G}_2} A$ (ccl)
2. $\mathsf{E}_{\mathsf{G}_2} \mathsf{C}_{\mathsf{G}_2} A \to \mathsf{E}_{\mathsf{G}_1} \mathsf{C}_{\mathsf{G}_2} A$ propositional reasoning
3. $\mathsf{C}_{\mathsf{G}_2} A \to \mathsf{E}_{\mathsf{G}_1} \mathsf{C}_{\mathsf{G}_2} A$ propositional reasoning
4. $\mathsf{C}_{\mathsf{G}_1} (\mathsf{C}_{\mathsf{G}_2} A \to \mathsf{E}_{\mathsf{G}_1} \mathsf{C}_{\mathsf{G}_2} A)$ C-necessitation
5. $\mathsf{C}_{\mathsf{G}_1} (\mathsf{C}_{\mathsf{G}_2} A \to \mathsf{E}_{\mathsf{G}_1} \mathsf{C}_{\mathsf{G}_2} A) \to (\mathsf{E}_{\mathsf{G}_1} \mathsf{C}_{\mathsf{G}_2} A \to \mathsf{C}_{\mathsf{G}_1} \mathsf{C}_{\mathsf{G}_2} A)$ (ind)
6. $\mathsf{E}_{\mathsf{G}_1} \mathsf{C}_{\mathsf{G}_2} A \to \mathsf{C}_{\mathsf{G}_1} \mathsf{C}_{\mathsf{G}_2} A$ (MP)

L_h^C	K_h^C	KD_h^C	$\mathsf{KD4}_h^\mathsf{C}$	$\mathsf{KD5}_h^\mathsf{C}$	$\mathsf{KD45}_h^\mathsf{C}$	KT_h^C	$\mathsf{KT4}_h^\mathsf{C}$	$\mathsf{KT5}_h^\mathsf{C}$	$\mathsf{KT45}_h^\mathsf{C}$	$\mathsf{K4}_h^\mathsf{C}$	$\mathsf{K45}_h^\mathsf{C}$	$\mathsf{K5}_h^\mathsf{C}$
(taut)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
(k)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
(d)		✓	✓	✓	✓							
(t)						✓	✓	✓	✓			
(4)			✓		✓		✓		✓	✓	✓	
(5)				✓	✓			✓	✓		✓	✓
(C-k)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
(res)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
(ccl)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
(ind)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
(MP)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
(Nec)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
(C-Nec)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 3.2.: Extensions. (taut) is used as an abbreviation for propositional tautologies.

7. $\mathsf{C}_{\mathsf{G}_2} A \to \mathsf{C}_{\mathsf{G}_1} \mathsf{C}_{\mathsf{G}_2} A$ — propositional reasoning
8. $\mathsf{C}_{\mathsf{G}_2} A \to \mathsf{E}_{\mathsf{G}_2} A$ — (ccl)
9. $\mathsf{E}_{\mathsf{G}_2} A \to \mathsf{E}_{\mathsf{G}_1} A$ — propositional reasoning
10. $\mathsf{C}_{\mathsf{G}_2} A \to \mathsf{E}_{\mathsf{G}_1} A$ — propositional reasoning
11. $\mathsf{E}_{\mathsf{G}_1} A \to (A \to \mathsf{E}_{\mathsf{G}_1} A)$ — propositional axiom
12. $\mathsf{C}_{\mathsf{G}_2} A \to (A \to \mathsf{E}_{\mathsf{G}_1} A)$ — propositional reasoning
13. $\mathsf{C}_{\mathsf{G}_1}(\mathsf{C}_{\mathsf{G}_2} A \to (A \to \mathsf{E}_{\mathsf{G}_1} A))$ — C-necessitation
14. $\mathsf{C}_{\mathsf{G}_1}(\mathsf{C}_{\mathsf{G}_2} A \to (A \to \mathsf{E}_{\mathsf{G}_1} A))$
    $\to (\mathsf{C}_{\mathsf{G}_1} \mathsf{C}_{\mathsf{G}_2} A \to \mathsf{C}_{\mathsf{G}_1}(A \to \mathsf{E}_{\mathsf{G}_1} A))$ — (C-k)
15. $\mathsf{C}_{\mathsf{G}_1} \mathsf{C}_{\mathsf{G}_2} A \to \mathsf{C}_{\mathsf{G}_1}(A \to \mathsf{E}_{\mathsf{G}_1} A)$ — (MP)
16. $\mathsf{C}_{\mathsf{G}_2} A \to \mathsf{C}_{\mathsf{G}_1}(A \to \mathsf{E}_{\mathsf{G}_1} A)$ — propositional reasoning
17. $\mathsf{C}_{\mathsf{G}_1}(A \to \mathsf{E}_{\mathsf{G}_1} A) \to (\mathsf{E}_{\mathsf{G}_1} A \to \mathsf{C}_{\mathsf{G}_1} A)$ — (ind)
18. $\mathsf{C}_{\mathsf{G}_2} A \to (\mathsf{E}_{\mathsf{G}_1} A \to \mathsf{C}_{\mathsf{G}_1} A)$ — propositional reasoning
19. $\mathsf{C}_{\mathsf{G}_2} A \to \mathsf{C}_{\mathsf{G}_1} A$ — propositional reasoning

□

However, we can not omit both.

Lemma 3.14. *The logic without the group restriction axiom (res) and with* C*-necessitation restricted to the group* $\mathsf{H} = \langle 1, \dots, h \rangle$ *of all agents is incomplete.*

Proof. We define the following translation $^\circ$:

$$
\begin{aligned}
P^\circ &:= P \\
(\neg A)^\circ &:= \neg(A^\circ) \\
(A \to B)^\circ &:= A^\circ \to B^\circ \\
(\Box_i A)^\circ &:= \Box_i A^\circ \\
(\mathsf{C}_\mathsf{H} A)^\circ &:= \mathsf{C}_\mathsf{H} A^\circ \\
(\mathsf{C}_\mathsf{G} A)^\circ &:= \bot, \text{ if } \mathsf{G} \neq \mathsf{H}
\end{aligned}
$$

Let $\mathsf{L}_h^{\mathsf{C}}$ be any of the logics defined above, but without the group restriction axiom (res). If A is a theorem of $\mathsf{L}_h^{\mathsf{C}}$, then so is A°. We prove this by induction on the derivation.

- If A is an instance of a propositional tautology or a modal axiom for the agents (k), then so is A°, as the common knowledge operator is not explicitly mentioned here.
- If A is an instance of the modal axiom for common knowledge (C-k), then A° is either another instance of that axiom (if $\mathsf{G} = \mathsf{H}$) or $\bot \to (\bot \to \bot)$, a propositional tautology.
- If A is an instance of the co-closure axiom (ccl), then A° is also either another instance of that axiom (if $\mathsf{G} = \mathsf{H}$) or of the form $\bot \to \mathsf{E}_\mathsf{G}(B^\circ \wedge \bot)$, a propositional tautology.
- If A is an instance of the induction axiom (ind), then again A° is either another instance of that axiom (if $\mathsf{G} = \mathsf{H}$) or of the form $\bot \to \mathsf{E}_\mathsf{G}(B^\circ \to \bot)$, a propositional tautology.
- If A is derived by modus ponens (MP) or the necessitation rule (Nec), then by induction hypothesis, we can also derive A° in the same way.
- If A is derived by common knowledge necessitation (C-Nec), then A is of the form $\mathsf{C}_\mathsf{H}B$ and we can use the induction hypothesis to derive B°. Using (C-Nec) we obtain $A^\circ = \mathsf{C}_\mathsf{H}B^\circ$, as we only have necessitation for the group of all agents.

Now assume (towards a contradiction) the group restriction axiom (res) was derivable in $\mathsf{L}_h^{\mathsf{C}}$. Then in particular,

$$\mathsf{C}_\mathsf{H}A \to \mathsf{C}_\mathsf{G}A$$

for a group of agents $\mathsf{G} \subsetneq \mathsf{H}$ would be derivable. But then also

$$(\mathsf{C}_\mathsf{H}A \to \mathsf{C}_\mathsf{G}A)^\circ = \mathsf{C}_\mathsf{H}A^\circ \to \bot$$

would be derivable. But the latter is in general not derivable and thus we arrive at a contradiction. Hence the group restriction axiom (res) is not derivable in $\mathsf{L}_h^{\mathsf{C}}$. □

3.6. The Muddy Children Puzzle

Let us conclude this chapter with the famous Muddy Children Puzzle (see [DHK07; Fag+95; MH95]) in order to illustrate common knowledge. Suppose three children, let us call them 1, 2, and 3, are playing outside. Whilst playing, some of the children get mud on their foreheads, say child 1 and 3 get muddy, while child 2 remains clean. While they can see perfectly well, whether another child is muddy, they do not know whether they themselves are muddy, as they have no means of looking at their foreheads. In order to formalize this, assume we have three propositions muddy_1, muddy_2 and muddy_3 in order to say child 1, 2, or 3, respectively, is muddy. Considering all possibilities, this gives rise to eight possible worlds, which we will name by three-digit binary numbers, e.g. 101, and thereby signal which propositions are evaluated to true in which world. For example, in world 101, propositions muddy_1 and muddy_3 evaluate to true, while muddy_2 evaluates to false. This situation is depicted in Figure 3.1, where the actual world 101 is highlighted. The accessibility relations are labelled by the children they are assigned to and for the sake of presentation, we omit the arrows for the reflexive part of the accessibility relation (i.e. we have $R_i(w, w)$ for each $i = 1, 2, 3$ and each world w even though this is not depicted) and the non-directed arrow indicates accessibility in both directions. For example, we have $R_2(101, 111)$ and $R_2(111, 101)$. Let us now look at some statements about the actual world. It is easy to see that in this model each child knows that at least one child is muddy. Furthermore, no child knows about its own state of muddiness.

Now the father of the children enters the scene and announces to the children, that at least one of the children is muddy. This leads to the situation in Figure 3.2, where the world with no muddy children is not considered possible anymore. Note that this public announcement generated common knowledge of the fact that at least one child is muddy. Also, compare this to the previous situation where we only had mutual knowledge of this fact, wheras common knowledge failed, as, e.g. child 3 considered it possible that child 1 considers the possibility of no child being muddy.

After his first announcement the father then asks the children whether they know which of them is muddy. All children answer “no”, as they all are not sure about their own state of muddiness. However, such

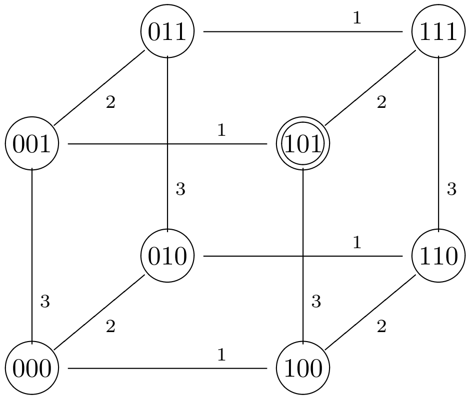

Figure 3.1.: Muddy Children Puzzle: Initial state

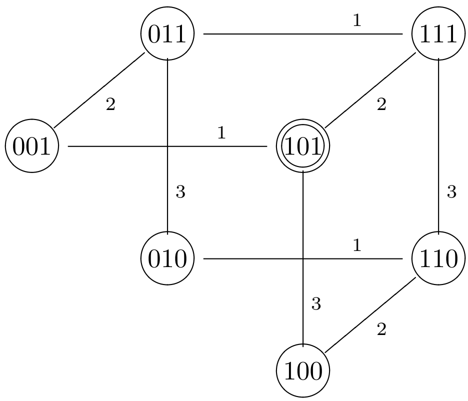

Figure 3.2.: Muddy Children Puzzle: The father announces “At least one child is muddy”.

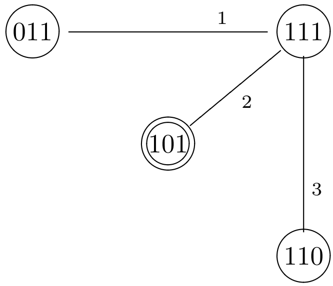

Figure 3.3.: Muddy Children Puzzle: Final state.

an announcement of non-knowledge also bears information. What can the children derive from their statements? Child 3 still considers the possibility of only child 1 being muddy. But if this was the case, then child 1 would have answered “yes” to their father’s question, as it knew that at least one child is muddy and could clearly see, that neither of the other two children is muddy and hence must be muddy itself. As this did not happen, child 3 deduces that not only child 1 is muddy. The other children reason in a similar manner and we obtain the situation as given in Figure 3.3. Now, when the father asks again whether the children know who is muddy, child 1 and 3 can step forward and proclaim “yes”.

A notable fact about this puzzle is that the father did not announce a new or unknown fact. Clearly, each child knew that at least one child was muddy. However, the announcement made this fact common knowledge and only thus led finally to the solution. See also Chapter B for more information on public announcements and also [DHK07] where more puzzles of this kind are treated.

4. A Co-Inductive Proof System for Common Knowledge

Σωκράτης
λέγει που Ἡράκλειτος ὅτι πάντα χωρεῖ καὶ οὐδὲν μένει, καὶ ποταμοῦ ῥοῇ ἀπεικάζων τὰ ὄντα λέγει ὡς δὶς ἐς τὸν αὐτὸν ποταμὸν οὐκ ἂν ἐμβαίης.[1]

Plato, Cratylus 402 *[Pla03]*

We introduce a co-inductive system S for common knowledge. In this formal system, proofs are finitely branching trees that may have infinitely long branches. This contrasts the system $K_h^\omega(C)$ from [AJ05], which was presented in the previous chapters, where proofs are infinitely branching trees (due to the ω-rule) but all the branches are finitely long. Such systems have previously been studied, for example, for the μ-calculus [NW96; Stu08] and the linear time μ-calculus [DHL06]. The underlying idea of this approach is based on the fundamental semantic theorem of the modal μ-calculus [BS07] (due to Streett and Emerson [SE89]). A similar result was also developed in [SW91]. Note that for the modal μ-calculus also systems using an ω-rule are available, see [JKS08].

We establish the soundness and completeness of the infinitary system S along the lines of [NW96] by employing techniques from the proof of the fundamental semantic theorem and utilizing the determinacy of certain infinite games. Alternatively, we could use the completeness

[1] **Socrates**
Heracleitus says, you know, that all things move and nothing remains still, and he likens the universe to the current of a river, saying that you cannot step twice into the same stream.
Translation from [Pla21].

of the common knowledge system with an ω-rule as in [AJ05]. The transformation from ω-rules to infinite branches then would yield the completeness of S (see [Stu08] for this approach in the context of the μ-calculus).

We now introduce the infinitary system S for common knowledge as published in [BKS10b]. As mentioned above, in this formal system, proofs are finitely branching trees that may have infinitely long branches while all finite branches must still end in an axiom. In order to obtain a sound deductive system, we have to impose a global constraint on such infinite branches. Roughly, we require that on every infinite branch in a proof, there be a greatest fixed point, i.e. a common knowledge formula, unfolded infinitely often.

Definition 4.1. A *preproof* for a sequent Γ is a possibly infinite tree whose root is labeled with Γ and which is built according to the following axioms and rules:

Axioms For all sequents Γ and all propositions P,

$$\Gamma, P, \overline{P} \quad (\mathsf{ax})$$

Propositional rules For all sequents Γ and all formulae A and B,

$$\frac{\Gamma, A, B}{\Gamma, A \lor B} \ (\lor) \qquad\qquad \frac{\Gamma, A \qquad \Gamma, B}{\Gamma, A \land B} \ (\land)$$

Modal rules For all sequents Γ and Σ, all formulae A, and all indices $1 \leq i \leq h$,

$$\frac{\Gamma, A}{\Diamond_i \Gamma, \Box_i A, \Sigma} \ (\Box)$$

Fixed point rules For all sequents Γ and all formulae A,

$$\frac{\Gamma, \tilde{\mathsf{E}} A \lor \tilde{\mathsf{E}}\tilde{\mathsf{C}} A}{\Gamma, \tilde{\mathsf{C}} A} \ (\tilde{\mathsf{C}}) \qquad\qquad \frac{\Gamma, \mathsf{E} A \land \mathsf{E}\mathsf{C} A}{\Gamma, \mathsf{C} A} \ (\mathsf{C})$$

We now introduce the notion of a thread in a branch of a proof tree.

Definition 4.2. The *principal formula* of a rule is the formula that is explicitly displayed in the conclusion of the rule. The *active formulae* of a rule are those formulae that are explicitly displayed in the premise(s) of the rule. The formulae in Γ and Σ are called the *side formulae* of a rule.

Definition 4.3. Consider a preproof for some sequent. For all rule applications r that occur in this tree, we define a *connection relation* $\mathsf{Con}(r)$ on formulae as follows:

1. If r is not an application of ($\Box$), we define $(A, B) \in \mathsf{Con}(r)$ if $A = B$ and A is a side formula of r or if A is the principal formula and B is an active formula of r.

2. If r is an application of ($\Box$), we define $(\Box_i A, A) \in \mathsf{Con}(r)$ if $\Box_i A$ is the principal formula of r and we define $(\Diamond_i B, B) \in \mathsf{Con}(r)$ if $\Diamond_i B \in \Diamond_i \Gamma$.

Definition 4.4. Consider a finite or infinite branch $\Gamma_0, \Gamma_1, \ldots$ in a preproof. Let r_i be the rule application where Γ_i is the conclusion and Γ_{i+1} is a premise. A *thread* in this branch is a sequence of formulae $A_0, A_1, \ldots$ such that $(A_i, A_{i+1}) \in \mathsf{Con}(r_i)$ and $A_i \in \Gamma_i$ for every i. Note that a thread in an infinite branch may be finite or infinite.

Definition 4.5. Consider an infinite branch of a preproof for a sequent Γ. An infinite thread in this branch is called a C-*thread* if infinitely many of its formulae are the principal formulae of applications of (C).

Definition 4.6. An S-*proof* for a sequent Γ is a preproof for Γ such that every finite branch ends in an axiom and every infinite branch contains a C-thread. We write $\mathsf{S} \vdash \Gamma$ if there exists an S-proof for Γ.

We will illustrate how S-proofs work by deriving three theorems in the following example. In order to present this derivations in a compact form, we need to state some properties of the system first. It should be noted that the proof of Lemma 4.7(2) requires infinite derivations, e.g., in the case of $A = \mathsf{C}B$.

Lemma 4.7. *1. For all formulae A and all sequents Γ and Σ, the following analog of the ($\Box$)-rule is derivable in S:*

$$\frac{\Gamma, A}{\tilde{\mathsf{E}}\Gamma, \mathsf{E}A, \Sigma} \quad (\mathsf{E})$$

2. *For all formulae A and all sequents Γ, the following generalized form of axioms* (ax) *is derivable:*

$$\mathsf{S} \vdash \Gamma, A, \neg A \quad (\mathsf{ax}')$$

Proof. 1. Using several applications of $(\vee)$ we can rewrite the sequent $\tilde{\mathsf{E}}\Gamma, \mathsf{E}A, \Sigma$ as

$$\Diamond_1\Gamma, \dots, \Diamond_h\Gamma, \Box_1 A \wedge \dots \wedge \Box_h A, \Sigma$$

and we get the following proof tree

$$\dfrac{\dfrac{\Gamma, A}{\Diamond_1\Gamma, \Box_1 A, \Sigma_1}(\Box) \qquad \dfrac{\dfrac{\Gamma, A}{\Diamond_2\Gamma, \Box_2 A, \Sigma_2}(\Box) \qquad \dfrac{\dfrac{\Gamma, A}{\Diamond_h\Gamma, \Box_h, \Sigma_h}(\Box)}{\vdots \; (*)}}{\Diamond_1\Gamma, \dots, \Diamond_h\Gamma, \Box_2 A \wedge \dots \wedge \Box_h A, \Sigma}(\wedge)}{\Diamond_1\Gamma, \dots, \Diamond_h\Gamma, \Box_1 A \wedge \dots \wedge \Box_h A, \Sigma}(\wedge)$$

where $(*)$ denotes further applications of $(\wedge)$ and $(\Box)$ in the same manner as to deal with $\Box_1 A$ and $\Box_2 A$.

2. The proof is by induction of the complexity of the formula A. If A is a propositional variable then the claim is obviously an instance of (ax). The cases where the main connective of A is $\wedge$, $\vee$, $\Box_i$ or $\Diamond_i$ are straightforward. Let us therefore look at the case where A is $\mathsf{C}B$. Then $\neg A$ is $\tilde{\mathsf{C}}\neg B$. We have

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\tilde{\mathsf{C}}\neg B, B, \neg B}(\mathrm{IH})}{\Gamma, \mathsf{E}B, \tilde{\mathsf{E}}\neg B, \tilde{\mathsf{E}}\tilde{\mathsf{C}}\neg B}(\mathsf{E}) \qquad \dfrac{\dfrac{\vdots \; (*)}{\underline{\mathsf{C}B}, \neg B, \tilde{\mathsf{C}}\neg B}(\tilde{\mathsf{C}})}{\Gamma, \underline{\mathsf{EC}B}, \tilde{\mathsf{E}}\neg B, \tilde{\mathsf{E}}\tilde{\mathsf{C}}\neg B}(\mathsf{E})}{\Gamma, \underline{\mathsf{E}B \wedge \mathsf{EC}B}, \tilde{\mathsf{E}}\neg B, \tilde{\mathsf{E}}\tilde{\mathsf{C}}\neg B}(\wedge)}{\Gamma, \underline{\mathsf{C}B}, \tilde{\mathsf{E}}\neg B, \tilde{\mathsf{E}}\tilde{\mathsf{C}}\neg B}(\mathsf{C})}{\Gamma, \underline{\mathsf{C}B}, \tilde{\mathsf{E}}\neg B \vee \tilde{\mathsf{E}}\tilde{\mathsf{C}}\neg B}(\vee)}{\Gamma, \underline{\mathsf{C}B}, \tilde{\mathsf{C}}\neg B}(\tilde{\mathsf{C}})$$

On the left branch we can apply our induction hypothesis (IH) to prove $\tilde{\mathsf{C}}\neg B, B, \neg B$. The right branch $(*)$ is continued infinitely long in the same manner as the part of the tree that is displayed starting with the sequent $\mathsf{C}B, \neg B, \tilde{\mathsf{C}}\neg B$ instead of $\Gamma, \underline{\mathsf{C}B}, \tilde{\mathsf{C}}\neg B$. The underlined formulae form a C-thread and thus the global condition is met and we indeed have a proof tree.

The case where A is $\tilde{\mathsf{C}}B$ follows by symmetry. $\square$

As our proof system uses a non-local correctness criteria, some remarks are due on the above derived rules. First, the notion of C-threads can be easily extended to include the (E)-rule. Now assume we are given a S-proof using (E) as an additional rule, in particular this means the global correctness criteria is satisfied, i.e., every infinite branch contains a C-thread. We can turn this proof into a proper S-proof, i.e., a proof not using the (E)-rule by simply replacing each occurrence of the (E)-rule by the proof tree given above (and the additional applications of ($\vee$) mentioned), thereby h-times copying the proof of Γ, A. Note that the proof tree used to replace the (E)-rule only contains applications of ($\vee$), ($\wedge$) and ($\square$). It can now be easily seen that any infinite branch in the proof without the (E)-rule corresponds to an infinite branch in the proof with the (E)-rule and the C-thread witnessing the correctness of the proof with the (E)-rule for this branch can be transformed into a C-thread witnessing the correctness of the proof without the (E)-rule by taking into consideration the additional applications of ($\vee$), ($\wedge$) and ($\square$).

The case for (ax') is equally simple, as it easily can be seen that combining two S-proofs by ($\wedge$) or extending a S-proof by application of any of the other rules yields another S-proof, i.e., application of the rules preserves the global correctness criteria. This immediately justifies the usage of (ax') in proofs.

Example 4.8. We present three sample S-proofs. The topmost sequents labeled (ax') are derivable by Lemma 4.7(2). The only infinite branches outside of (ax')-derivations are marked ($*$). They are built in the same manner as the part of the proof tree displayed, i.e. by applying the same rule again, which is possible, as they start with a sequent of the same form as the one at the root of the tree. To show that these preproofs are indeed S-proofs, it is sufficient to find a C-thread in these branches. The C-threads are given by the underlined formulae.

1. Figure 4.1 contains the bottom part of an infinite S-proof for $\mathsf{C}A \to \mathsf{C}\mathsf{C}A$ expressed in a sequent form as $\tilde{\mathsf{C}}\neg A, \mathsf{C}\mathsf{C}A$.

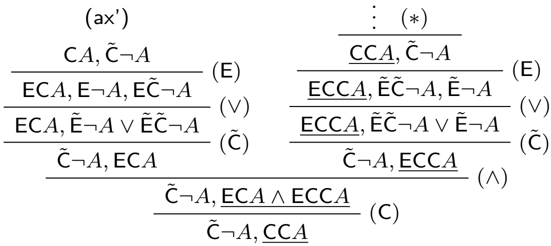

Figure 4.1.: A sample S-proof for $\mathsf{C}A \to \mathsf{CC}A$ with a highlighted C-thread.

2. Figure 4.2 contains the bottom part of an infinite S-proof for $\mathsf{C}A \to \mathsf{CE}A$ expressed in a sequent form as $\tilde{\mathsf{C}}\neg A, \mathsf{CE}A$.
3. Fig. 4.3 contains the bottom part of an infinite S-proof for the induction axiom (I-Ax) expressed in a sequent form as $\tilde{\mathsf{E}}\neg A, \tilde{\mathsf{C}}(A \wedge \tilde{\mathsf{E}}\neg A), CA$.

4.1. Soundness

The soundness proof essentially uses the idea that underlies the fundamental semantic theorem of the modal μ-calculus.

Definition 4.9. $\delta(A)$ denotes the maximal number of nested C operators in the formula A

Example 4.10. For instance, we have $\delta(\mathsf{C}(\mathsf{C}P \vee \mathsf{C}Q)) = 2$.

Definition 4.11. Let $m \geq 1$ and $\sigma = (\sigma_m, \ldots, \sigma_1)$ be a sequence of ordinals with $\sigma_i \leq \omega$. For all formulae A such that $\delta(A) \leq m$, we define the satisfaction relation $\Vdash^\sigma_{\mathsf{C}}$ in the same way as $\Vdash$ except in the case of C, where we set $\mathcal{M}, w \Vdash^\sigma_{\mathsf{C}} \mathsf{C}B$ if $\mathcal{M}, v \Vdash^\sigma_{\mathsf{C}} B$ for all v with $\mathsf{reach}(w, v, n)$ where n is a natural number with $\sigma_{\delta(\mathsf{C}B)} \geq n \geq 1$.

We immediately obtain the following simple facts

Lemma 4.12. *1. Let A be a formula and $\sigma = (\sigma_m, \ldots, \sigma_1)$ with $m \geq \delta(A)$. Then $\mathcal{M}, w \Vdash A$ implies $\mathcal{M}, w \Vdash^\sigma_{\mathsf{C}} A$.*

$$
\dfrac{\dfrac{
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\overset{(\mathsf{ax'})}{A,\neg A}}{\mathsf{E}A,\tilde{\mathsf{E}}\neg A,\tilde{\mathsf{E}}\tilde{\mathsf{C}}\neg A}\,(\mathsf{E})}{\mathsf{E}A,\tilde{\mathsf{E}}\neg A\lor\tilde{\mathsf{E}}\tilde{\mathsf{C}}\neg A}\,(\lor)}{\mathsf{E}A,\tilde{\mathsf{C}}\neg A}\,(\tilde{\mathsf{C}})}{\mathsf{EE}A,\tilde{\mathsf{E}}\neg A,\tilde{\mathsf{E}}\tilde{\mathsf{C}}\neg A}\,(\mathsf{E})}{\mathsf{EE}A,\tilde{\mathsf{E}}\neg A\lor\tilde{\mathsf{E}}\tilde{\mathsf{C}}\neg A}\,(\lor)}{\tilde{\mathsf{C}}\neg A,\mathsf{EE}A}\,(\tilde{\mathsf{C}})
\qquad
\dfrac{\dfrac{\dfrac{\dfrac{\vdots\;(*)}{\underline{\mathsf{CE}A},\tilde{\mathsf{C}}\neg A}}{\underline{\mathsf{ECE}A},\tilde{\mathsf{E}}\neg A,\tilde{\mathsf{E}}\tilde{\mathsf{C}}\neg A}\,(\mathsf{E})}{\underline{\mathsf{ECE}A},\tilde{\mathsf{E}}\neg A\lor\tilde{\mathsf{E}}\tilde{\mathsf{C}}\neg A}\,(\lor)}{\tilde{\mathsf{C}}\neg A,\underline{\mathsf{ECE}A}}\,(\tilde{\mathsf{C}})
}{\tilde{\mathsf{C}}\neg A,\underline{\mathsf{ECE}A\land\mathsf{EE}A}}\,(\land)}{\tilde{\mathsf{C}}\neg A,\underline{\mathsf{CE}A}}\,(\mathsf{C})
$$

Figure 4.2.: A sample S-proof for $\mathsf{C}A \to \mathsf{CE}A$ with a highlighted C-thread.

$$
\dfrac{\dfrac{
\dfrac{\overset{(\mathsf{ax'})}{\neg A, A}}{\tilde{\mathsf{E}}\neg A,\tilde{\mathsf{C}}(A\land\tilde{\mathsf{E}}\neg A),\mathsf{E}A}\,(\mathsf{E})
\qquad
\dfrac{\dfrac{\dfrac{\dfrac{\overset{(\mathsf{ax'})}{\neg A,A,\tilde{\mathsf{C}}(A\land\tilde{\mathsf{E}}\neg A),\mathsf{C}A}\qquad\dfrac{\vdots\;(*)}{\neg A,\tilde{\mathsf{E}}\neg A,\tilde{\mathsf{C}}(A\land\tilde{\mathsf{E}}\neg A),\underline{\mathsf{C}A}}\,(\mathsf{C})}{\neg A,A\land\tilde{\mathsf{E}}\neg A,\tilde{\mathsf{C}}(A\land\tilde{\mathsf{E}}\neg A),\underline{\mathsf{C}A}}\,(\land)}{\tilde{\mathsf{E}}\neg A,\tilde{\mathsf{E}}(A\land\tilde{\mathsf{E}}\neg A),\tilde{\mathsf{E}}\tilde{\mathsf{C}}(A\land\tilde{\mathsf{E}}\neg A),\underline{\mathsf{EC}A}}\,(\mathsf{E})}{\tilde{\mathsf{E}}\neg A,\tilde{\mathsf{E}}(A\land\tilde{\mathsf{E}}\neg A)\lor\tilde{\mathsf{E}}\tilde{\mathsf{C}}(A\land\tilde{\mathsf{E}}\neg A),\underline{\mathsf{EC}A}}\,(\lor)}{\tilde{\mathsf{E}}\neg A,\tilde{\mathsf{C}}(A\land\tilde{\mathsf{E}}\neg A),\underline{\mathsf{EC}A}}\,(\tilde{\mathsf{C}})
}{\tilde{\mathsf{E}}\neg A,\tilde{\mathsf{C}}(A\land\tilde{\mathsf{E}}\neg A),\underline{\mathsf{E}A\land\mathsf{EC}A}}\,(\land)}{\tilde{\mathsf{E}}\neg A,\tilde{\mathsf{C}}(A\land\tilde{\mathsf{E}}\neg A),\underline{\mathsf{C}A}}\,(\mathsf{C})
$$

Figure 4.3.: A sample S-proof for the induction axiom $(\mathsf{I\text{-}Ax})$ with a highlighted C-thread.

2. *$\mathcal{M}, w \not\Vdash A$ implies that there exists σ such that $\mathcal{M}, w \not\Vdash^{\sigma}_{\mathsf{C}} A$.*

3. *Let $\sigma = (\sigma_m, \dots, \sigma_{\delta(\mathsf{C}B)}, \dots, \sigma_1)$, $\sigma_{\delta(\mathsf{C}B)} < \omega$ and*

$$\sigma' = (\sigma_m, \dots, \sigma_{\delta(\mathsf{C}B)} + 1, \dots, \sigma_1) \ .$$

We have $\mathcal{M}, w \Vdash^{\sigma'}_{\mathsf{C}} \mathsf{C}B$ if and only if $\mathcal{M}, w \Vdash^{\sigma}_{\mathsf{C}} \mathsf{E}B \wedge \mathsf{EC}B$.

Proof. 1. This is immediate by the definition of $\Vdash^{\sigma}_{\mathsf{C}}$.

2. We prove the contrapositive. If $\mathcal{M}, w \Vdash^{\sigma}_{\mathsf{C}} A$ for all σ, then in particular for $\sigma = (\omega, \dots, \omega)$, but then the definitions of $\Vdash^{\sigma}_{\mathsf{C}}$ and $\Vdash$ coincide. Note that this is the case where ω itself as a possible element of a sequence σ is necessary, as we explain below.

3. We have $\mathcal{M}, w \Vdash^{\sigma}_{\mathsf{C}} \mathsf{E}B \wedge \mathsf{EC}B$ if and only if for all v with $\mathsf{reach}(w, v, 1)$ we have $\mathcal{M}, w \Vdash^{\sigma}_{\mathsf{C}} B$ and $\mathcal{M}, w \Vdash^{\sigma}_{\mathsf{C}} \mathsf{C}B$. This again holds if and only if for all v with $\mathsf{reach}(w, v, 1)$ we have $\mathcal{M}, v \Vdash^{\sigma}_{\mathsf{C}} B$ and for all v' with $\mathsf{reach}(v, v', n)$ where $n \leq \sigma_{\delta(\mathsf{C}B)}$ we have $\mathcal{M}, v' \Vdash^{\sigma}_{\mathsf{C}} B$. This is equivalent to saying that for all v with $\mathsf{reach}(w, v, n)$ where $n \leq \sigma_{\delta(\mathsf{C}B)} + 1$ we have $\mathcal{M}, v' \Vdash^{\sigma}_{\mathsf{C}} B$. By definition, this is $\mathcal{M}, w \Vdash^{\sigma'}_{\mathsf{C}} \mathsf{C}B$. □

As stated above, ω is not only sufficient but also necessary as an element in sequences σ. In order to see this, consider the following example. Let $\mathcal{M}$ be given by

$$\begin{aligned} W &:= \{w\} \cup \{v_{k,l} \mid k \in \mathbb{N}, l \leq k\}, \\ R_i &:= \{(w, v_{k,1}) \mid k \in \mathbb{N}\} \cup \{(v_{k,l}, v_{k,l+1}) \mid k \in \mathbb{N}, l < k\}, \\ \nu(P) &:= \{w\} \cup \{v_{k,l} \mid k \in \mathbb{N}, k \neq l\}. \end{aligned}$$

Figure 4.4 depicts a part of this model. It is easy to see that we have $\mathcal{M}, w \not\Vdash \Diamond_1 \mathsf{C}P$, but $\mathcal{M}, w \Vdash^{\sigma}_{\mathsf{C}} \Diamond_1 \mathsf{C}P$ for any $\sigma = (n)$ with $n < \omega$.

Lemma 4.13. *Let A be a formula, Δ be a sequent, σ be a sequence of ordinals, $\mathcal{M} = (W, R_1, \dots, R_h, \nu)$ be a Kripke structure, $w \in W$ be a world, and $1 \leq i \leq h$. If $\mathcal{M}, w \not\Vdash \Box_i A, \Diamond_i \Delta$ and $\mathcal{M}, w \not\Vdash^{\sigma}_{\mathsf{C}} \Box_i A$, then there exists a world $v \in W$ with $R_i(w, v)$ such that $\mathcal{M}, v \not\Vdash A, \Delta$ and $\mathcal{M}, v \not\Vdash^{\sigma}_{\mathsf{C}} A$.*

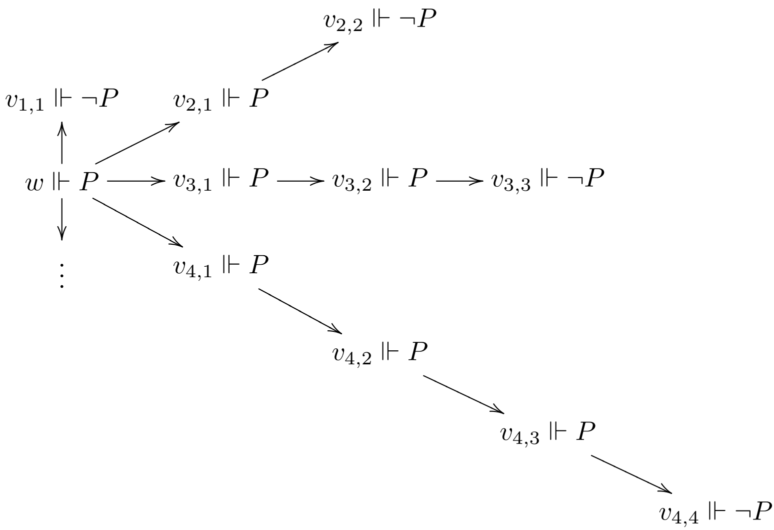

Figure 4.4.: A simple model showing the necessity of ω as an element of sequents σ.

Proof. Suppose (towards a contradiction) we have for all $v \in W$ with $R_i(w, v)$, that at least one of the claims $\mathcal{M}, v \Vdash A, \Delta$ or $\mathcal{M}, v \Vdash^{\sigma}_{\mathsf{C}} A$ holds. We distinguish the following two cases:

1. $\mathcal{M}, v \Vdash^{\sigma}_{\mathsf{C}} A$ holds for all $v \in W$ with $R_i(w, v)$. Then we have $\mathcal{M}, w \Vdash^{\sigma}_{\mathsf{C}} \Box_i A$. Contradiction.
2. There is at least one $v \in W$ with $R_i(w, v)$ such that $\mathcal{M}, v \nVdash^{\sigma}_{\mathsf{C}} A$. Then $\mathcal{M}, v \nVdash A$ by Lemma 4.12(1). Hence, there must be a formula $B \in \Delta$ such that $\mathcal{M}, v \Vdash B$. However, this means $\mathcal{M}, w \Vdash \Diamond_i B$ and, therefore, $\mathcal{M}, w \Vdash \Diamond_i \Delta$. Contradiction. □

Definition 4.14. Given two sequences σ and τ of equal length m, we say $\sigma < \tau$ if σ is smaller than τ with respect to the lexicographic ordering.

Since we consider sequences of a fixed length, the relation $<$ is a well-ordering.

Theorem 4.15 (Soundness). *For all formulae A, if A is not valid, then $\mathsf{S} \nvdash A$.*

Proof. Suppose (towards a contradiction) A is not valid yet there is an S-proof $\mathcal{T}$ for it. As A is not valid, there is a Kripke structure $\mathcal{M}$ and a world w such that $\mathcal{M}, w \nVdash A$. We will construct a branch $\Gamma_0, \Gamma_1, \dots$ with the corresponding inferences $r_0, r_1, \dots$ in $\mathcal{T}$ and a sequence $w_0, w_1, \dots$ of worlds in $\mathcal{M}$ such that

(a) $\mathcal{M}, w_i \nVdash \Gamma_i$ and

(b) if $(B, C) \in \mathsf{Con}(r_i)$, $B \in \Gamma_i$, $C \in \Gamma_{i+1}$, and $\mathcal{M}, w_i \nVdash^{\sigma}_{\mathsf{C}} B$, then $\mathcal{M}, w_{i+1} \nVdash^{\sigma}_{\mathsf{C}} C$.

Let $\Gamma_0 := A$ and $w_0 := w$. If Γ_i and w_i are given, we construct Γ_{i+1} and w_{i+1} according to the different cases for r_i. Note that because of a, Γ_i cannot be axiomatic and thus must have been inferred by some rule.

1. $r_i = (\Box)$: Let $\Box_i B \in \Gamma_i$ be the principal formula of r_i. Let σ be the least sequence such that $\mathcal{M}, w_i \nVdash^{\sigma}_{\mathsf{C}} \Box_i B$. We apply Lemma 4.13 for this σ to find a state w_{i+1} such that a and b hold. We let Γ_{i+1} be the unique premise of r_i.

2. $r_i = (\wedge)$: Let $B_1 \wedge B_2 \in \Gamma_i$ be the principal formula of r_i. Let σ be the least sequence such that $\mathcal{M}, w_i \not\Vdash^{\sigma}_{\mathsf{C}} B_1 \wedge B_2$. Let Γ_{i+1} be the j-th premise of r_i such that $\mathcal{M}, w_i \not\Vdash^{\sigma}_{\mathsf{C}} B_j$. Further, set $w_{i+1} := w_i$. This construction guarantees a and b.

3. In all other cases, r_i has a unique premise Δ. We set $w_{i+1} := w_i$ and $\Gamma_{i+1} := \Delta$. Again a and b hold.

We have constructed an infinite branch in $\mathcal{T}$. Since $\mathcal{T}$ is an S-proof, this branch must contain a C-thread $A_0, A_1, \ldots$. For each natural number j, we define σ^j to be the least sequence such that $\mathcal{M}, w_j \not\Vdash^{\sigma^j}_{\mathsf{C}} A_j$. Note that σ^j exists by Lemma 4.12(2). It follows from b that $\sigma^{j+1} \leq \sigma^j$ for all j. Moreover, because we consider a C-thread, there are infinitely many applications of (C), which, according to Lemma 4.12(3),[2] means that there are infinitely many j's with $\sigma^{j+1} < \sigma^j$. This contradicts the well-foundedness of $<$. □

4.2. Completeness

The completeness proof for the infinitary system S is based on the similar result for the modal μ-calculus from [NW96]. For a given formula A, we define an infinite game such that player I has a winning strategy if and only if there is an S-proof for A and player II has a winning strategy if and only if there is a countermodel for A. It is possible to show that this game is *determined*, i.e., one of the players has a winning strategy. Hence, the completeness of S follows.

Definition 4.16. A sequent Γ is *saturated* if all of the following conditions hold:

1. if $A \wedge B \in \Gamma$, then $A \in \Gamma$ or $B \in \Gamma$,

2. if $A \vee B \in \Gamma$, then $A \in \Gamma$ and $B \in \Gamma$,

3. if $\mathsf{C}A \in \Gamma$, then $\mathsf{E}A \wedge \mathsf{E}\mathsf{C}A \in \Gamma$, and

4. if $\tilde{\mathsf{C}}A \in \Gamma$, then $\tilde{\mathsf{E}}A \vee \tilde{\mathsf{E}}\tilde{\mathsf{C}}A \in \Gamma$.

[2]For σ minimal with $\mathcal{M}, w \not\Vdash^{\sigma}_{\mathsf{C}} \mathsf{C}A$, we have $\sigma_{\delta(\mathsf{C}A)} < \omega$ as a very simple consequence of the definition of $\Vdash^{\sigma}_{\mathsf{C}}$.

Definition 4.17. The *system* $\mathsf{S}_{\mathsf{Game}}$ consists of the rules of S with the rule $(\Box)$ being replaced by the rule $(\Box')$ in Figure 4.5:

An $\mathsf{S}_{\mathsf{Game}}$-*tree for a sequent* Γ is built by iterating the following two steps until one reaches a saturated sequent which is either axiomatic or to which $(\Box')$ cannot be applied:

1. Apply the rules $(\vee)$, $(\wedge)$, (C), and $(\tilde{\mathsf{C}})$ backwards until a saturated sequent is reached.[3] While applying the rules, make sure that the conclusion always remains a subset of the premise.[4]

2. Apply $(\Box')$ backwards, if possible.

Note that the rule $(\Box')$ explores all possible applications of $(\Box)$.

We now introduce a system $\mathsf{S}_{\mathsf{Dis}}$ for establishing unprovability. Accordingly, its rules should not be read as sound, i.e., preserving validity, but rather as "dis-sound," i.e., preserving invalidity.

Definition 4.18. The *system* $\mathsf{S}_{\mathsf{Dis}}$ consists of the rules of $\mathsf{S}_{\mathsf{Game}}$ whereby $(\wedge)$ is replaced by the following two rules:

Alternative $(\wedge)$**:** For all sequents Γ and all formulae A and B,

$$\frac{\Gamma, A}{\Gamma, A \wedge B}\ (\wedge_{\mathsf{Dis}}1) \qquad\qquad \frac{\Gamma, B}{\Gamma, A \wedge B}\ (\wedge_{\mathsf{Dis}}2)$$

An $\mathsf{S}_{\mathsf{Dis}}$-*tree* is built in the same way as an $\mathsf{S}_{\mathsf{Game}}$-tree except that $(\wedge_{\mathsf{Dis}}1)$ and $(\wedge_{\mathsf{Dis}}2)$ are used instead of $(\wedge)$.

Note that an $\mathsf{S}_{\mathsf{Dis}}$-tree for a sequent Γ is not unique.

The notions of a *thread* and a C-*thread* are extended to $\mathsf{S}_{\mathsf{Game}}$- and $\mathsf{S}_{\mathsf{Dis}}$-trees. A $\tilde{\mathsf{C}}$-*thread* is a thread that contains infinitely many principal formulae of applications of $(\tilde{\mathsf{C}})$. Note that any infinite thread is either a C- or a $\tilde{\mathsf{C}}$-thread but not both.

[3] As we explore all possible saturations of a given sequent, the order of application of rules is not relevant. If a deterministic procedure is preferred, one could simply fix an enumeration of all formulae in the language and then apply the rules backwards to the formula with least index in the sequent that has not been treated yet.

[4] As sequents are set of formulae, we have for example $\Gamma, A \vee B = \Gamma, A \vee B, A \vee B$. Thus, we can get $\Gamma, A \vee B, A, B$ from $\Gamma, A \vee B$ by applying the $(\vee)$-rule backwards. We deal accordingly with the other rules.

Let $1 \leq m \leq h$, $H = \{h_1, \ldots, h_m\} \subseteq \{1, \ldots, h\}$, and $n_{h_1}, \ldots, n_{h_m}$ be positive integers. For all saturated sequents Σ that contain neither formulae that start with $\Diamond_j$, $j \in H$, nor formulae that start with $\Box_i$, $1 \leq i \leq h$, all sequents Γ_j, $j \in H$, and all formulae $A_{j,1}, \ldots, A_{j,n_j}$, $j \in H$, the rule can be applied.

$$\frac{\Gamma_{h_1}, A_{h_1,1} \quad \ldots \quad \Gamma_{h_1}, A_{h_1,n_{h_1}} \quad \ldots \quad \Gamma_{h_m}, A_{h_m,1} \quad \ldots \quad \Gamma_{h_m}, A_{h_m,n_{h_m}}}{\Diamond_{h_1}\Gamma_{h_1}, \Box_{h_1}A_{h_1,1}, \ldots, \Box_{h_1}A_{h_1,n_{h_1}}, \ldots, \Diamond_{h_m}\Gamma_{h_m}, \Box_{h_m}A_{h_m,1}, \ldots, \Box_{h_m}A_{h_m,n_{h_m}}, \Sigma} \; (\Box')$$

Note that this rule has $n_{h_1} + \cdots + n_{h_m}$ many premises.

Figure 4.5.: **Alternative modal rules**.

Definition 4.19. We say that an $\mathsf{S}_{\mathsf{Dis}}$-tree $\mathcal{T}$ for a sequent Γ *disproves* Γ if

1. no branch ends with an axiom and
2. any infinite thread in any branch is a $\tilde{\mathsf{C}}$-thread.

Example 4.20. In order to disprove $\tilde{\mathsf{C}}\overline{P} \to \tilde{\mathsf{C}}\mathsf{C}P$, we construct an $\mathsf{S}_{\mathsf{Dis}}$-tree $\mathcal{T}$ for a corresponding sequent $\mathsf{C}P, \tilde{\mathsf{C}}\mathsf{C}P$ (see Fig. 4.6). In this tree, $\Diamond_1\mathsf{C}P, \Diamond_1\tilde{\mathsf{C}}\mathsf{C}P, \Box_1 P, \Sigma$ is a saturation of the sequent

$$\mathsf{C}P, \mathsf{E}P \wedge \mathsf{E}\mathsf{C}P, \mathsf{E}P, \tilde{\mathsf{C}}\mathsf{C}P, \tilde{\mathsf{E}}\mathsf{C}P \vee \tilde{\mathsf{E}}\tilde{\mathsf{C}}\mathsf{C}P, \tilde{\mathsf{E}}\mathsf{C}P, \tilde{\mathsf{E}}\tilde{\mathsf{C}}\mathsf{C}P \ . \tag{4.1}$$

The saturation process is abbreviated as $(*)$. It involves exactly $2h-2$ applications of $(\vee)$ to saturate the disjunctions $\tilde{\mathsf{E}}\mathsf{C}P$ and $\tilde{\mathsf{E}}\tilde{\mathsf{C}}\mathsf{C}P$. In addition, the conjunction $\mathsf{E}P$ is saturated by at most $h-1$ applications of $(\wedge_{\mathsf{Dis}}1)$ and $(\wedge_{\mathsf{Dis}}2)$ in such a way that $\Box_1 P$ is the only resulting formula that starts with $\Box_i$. Most formulae that result from this saturation are disjunctions, conjunctions, or are already present in (4.1), with the exception of

$$\Diamond_1\mathsf{C}P, \ldots, \Diamond_h\mathsf{C}P, \Diamond_1\tilde{\mathsf{C}}\mathsf{C}P, \ldots, \Diamond_h\tilde{\mathsf{C}}\mathsf{C}P, \text{ and } \Box_1 P \ .$$

Thus, Σ contains neither formulae that start with $\Box_i$ nor formulae that start with $\Diamond_1$, which enables us to apply $(\Box')$. The tree $\mathcal{T}$ extends upward indefinitely with infinitely many repetitions of the sequent $\mathsf{C}P, \tilde{\mathsf{C}}\mathsf{C}P, P$. This tree has only one branch, which is infinite. And this branch contains only one infinite thread, the one that consists of the underlined formulae in Fig. 4.6. And this thread is indeed a $\tilde{\mathsf{C}}$-thread.

It may seem that this branch also contains a C-thread because there are infinitely many applications of (C) in the branch. However, the principal formulae of these (C)-rules do not belong to one thread. In particular, the thread that starts from $\mathsf{C}P$ in the root sequent does not pass through $\mathsf{C}P$ in the premise of the $(\Box')$-rule shown in Fig. 4.6. Instead, this thread passes through $\mathsf{E}P \wedge \mathsf{E}\mathsf{C}P$, $\mathsf{E}P$, …, $\Box_1 P$, and P and eventually disappears after the next application of $(\Box')$.

Now we are going to show that any sequent Γ has either an S-tree that proves it or an $\mathsf{S}_{\mathsf{Dis}}$-tree that disproves it.

Let $\mathcal{T}$ be an $\mathsf{S}_{\mathsf{Game}}$-tree for Γ. We define an infinite game for two players on $\mathcal{T}$. Intuitively, player I will try to show that Γ is provable

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\vdots \quad (*)}{\mathsf{C}P, \mathsf{E}P \wedge \mathsf{E}\mathsf{C}P, \mathsf{E}P, \tilde{\mathsf{C}}\mathsf{C}P, \tilde{\mathsf{E}}\mathsf{C}P \vee \tilde{\mathsf{E}}\tilde{\mathsf{C}}\mathsf{C}P, \tilde{\mathsf{E}}\mathsf{C}P, \underline{\tilde{\mathsf{E}}\tilde{\mathsf{C}}\mathsf{C}P}}}{\mathsf{C}P, \mathsf{E}P \wedge \mathsf{E}\mathsf{C}P, \mathsf{E}P, \tilde{\mathsf{C}}\mathsf{C}P, \underline{\tilde{\mathsf{E}}\mathsf{C}P \vee \tilde{\mathsf{E}}\tilde{\mathsf{C}}\mathsf{C}P}}(\vee)}{\mathsf{C}P, \mathsf{E}P \wedge \mathsf{E}\mathsf{C}P, \mathsf{E}P, \underline{\tilde{\mathsf{C}}\mathsf{C}P}}(\tilde{\mathsf{C}})}{\mathsf{C}P, \mathsf{E}P \wedge \mathsf{E}\mathsf{C}P, \underline{\tilde{\mathsf{C}}\mathsf{C}P}}(\wedge_{\mathsf{Dis}}1)}{\mathsf{C}P, \underline{\tilde{\mathsf{C}}\mathsf{C}P}}(\mathsf{C})}{}
$$

where the part above $(*)$ is:

$$
\dfrac{\dfrac{\vdots}{\mathsf{C}P, \underline{\tilde{\mathsf{C}}\mathsf{C}P}, P}(\mathsf{C})}{\Diamond_1\mathsf{C}P, \underline{\Diamond_1\tilde{\mathsf{C}}\mathsf{C}P}, \Box_1 P, \Sigma}(\Box')
$$

Figure 4.6.: A sample $\mathsf{S}_{\mathsf{Dis}}$-disproof for $\tilde{\mathsf{C}}\overline{P} \to \tilde{\mathsf{C}}\mathsf{C}P$ with a highlighted $\tilde{\mathsf{C}}$-thread.

while player II will try to show the opposite. The game is played as follows:

1. the game starts at the root of $\mathcal{T}$,
2. at any $(\Box')$ node, player I chooses one of the children,
3. at any $(\wedge)$ node, player II chooses one of the children,
4. at all other non-leaf nodes, the only child is chosen by default.

Such a game results in a path in $\mathcal{T}$. In the case of a finite path, player I wins if the path ends in an axiom; otherwise, player II wins. In the case of an infinite path, player I wins if the path contains a C-thread; otherwise, player II wins.

Theorem 4.21. *1. There is a winning strategy for player I if and only if there is an* S*-proof for* Γ *contained in* $\mathcal{T}$*.*

2. There is a winning strategy for player II if and only if there is an $\mathsf{S}_{\mathsf{Dis}}$*-disproof for* Γ *contained in* $\mathcal{T}$*.*

Proof. For the first claim, if there is an S-proof for Γ contained in $\mathcal{T}$, then the winning strategy for player I is to stay in the nodes that belong to this proof. For the other direction, consider a winning strategy for player I. It induces an S-proof for Γ as follows: the root of $\mathcal{T}$ is the root of the proof; if a node is included in the proof and player I has to perform the next move, then we select the child prescribed by the winning strategy; if it is player II's move, then we include all the children in our proof. The proof of the second claim is similar. $\square$

With the help of

Theorem 4.22 (Martin's theorem [Mar75])**.** *All Borel games are determined.*

we can show that this game is determined, i.e., one of the players has a winning strategy. For details of this argument, see [DHL06; NW96]. We obtain the following as a corollary:

Theorem 4.23. *Let $\mathcal{T}$ be an $\mathsf{S}_{\mathsf{Game}}$-tree for Γ. Then there exists either an S-proof for Γ in $\mathcal{T}$ or an $\mathsf{S}_{\mathsf{Dis}}$-disproof for Γ in $\mathcal{T}$.*

It remains to show that from a given $\mathsf{S}_{\mathsf{Dis}}$-disproof for Γ, we can construct a countermodel for Γ.

Definition 4.24. Consider an $\mathsf{S}_{\mathsf{Dis}}$-tree $\mathcal{T}$ disproving a sequent Γ. The Kripke structure $\mathcal{M}^{\mathcal{T}} = (W^{\mathcal{T}}, R_1^{\mathcal{T}}, \ldots, R_h^{\mathcal{T}}, \nu^{\mathcal{T}})$ *induced* by $\mathcal{T}$ is defined as follows:

1. $W^{\mathcal{T}}$ consists of all occurrences of sequents in the conclusions of applications of $(\Box')$ in $\mathcal{T}$ as well as of all occurrences of sequents in the leaves of $\mathcal{T}$,
2. $R_i^{\mathcal{T}}(\Gamma, \Delta)$ holds if there is exactly one application of $(\Box')$ in between Γ and Δ and if there is a thread through Γ and Δ that contains $\Box_i A \in \Gamma$ and $A \in \Delta$ for some formula A,
3. $\nu^{\mathcal{T}}(P) := \{\Gamma \in W^{\mathcal{T}} : P \notin \Gamma\}$.

We can assign to each sequent Δ in $\mathcal{T}$ the corresponding world in $W^{\mathcal{T}}$ simply by finding the closest saturated descendant. We will denote this world by $\mathsf{sat}(\Delta)$.

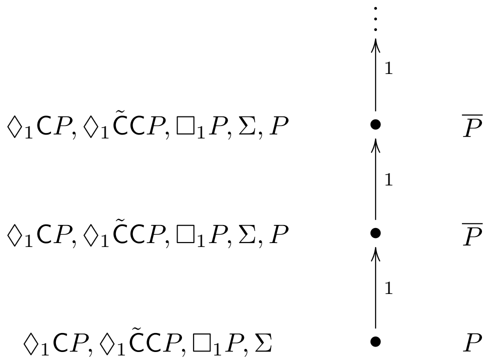

Figure 4.7.: The Kripke structure $\mathcal{M}^{\mathcal{T}}$ induced by the $\mathsf{S}_{\mathsf{Dis}}$-tree $\mathcal{T}$ from Example 4.20.

Example 4.25. The $\mathsf{S}_{\mathsf{Dis}}$-tree $\mathcal{T}$ for $\tilde{\mathsf{C}}\overline{P} \to \tilde{\mathsf{C}}\mathsf{C}P$ constructed in Example 4.20 induces a Kripke structure $\mathcal{M}^{\mathcal{T}}$ shown in Fig. 4.7. It is easy to see that

$$\mathcal{M}^{\mathcal{T}}, \quad \Diamond_1 \mathsf{C}P, \Diamond_1 \tilde{\mathsf{C}}\mathsf{C}P, \Box_1 P, \Sigma \quad \not\Vdash \quad \tilde{\mathsf{C}}\overline{P} \to \tilde{\mathsf{C}}\mathsf{C}P \ .$$

Lemma 4.29 states that this is a general phenomenon: the root of the Kripke structure induced by a given $\mathsf{S}_{\mathsf{Dis}}$-tree falsifies the sequent at the root of the tree.

Definition 4.26. $\tilde{\delta}(A)$ denotes the maximal number of nested $\tilde{\mathsf{C}}$ operators in A.

Definition 4.27. Let $\mathcal{M}$ be a Kripke structure, w a world in $\mathcal{M}$, and A a formula. We define the $\tilde{\mathsf{C}}$-*signature* $\mathsf{sig}_{\tilde{\mathsf{C}}}(A, w)$ to be the least sequence $\sigma = (\sigma_{\tilde{\delta}(A)}, \ldots, \sigma_1)$ such that $\mathcal{M}, w \Vdash^{\sigma}_{\tilde{\mathsf{C}}} A$

Here $\Vdash^{\sigma}_{\tilde{\mathsf{C}}}$ is defined in the same way as $\Vdash$ except in the case of $\tilde{\mathsf{C}}$, where we set $\mathcal{M}, w \Vdash^{\sigma}_{\tilde{\mathsf{C}}} \tilde{\mathsf{C}}B$ if $\mathcal{M}, v \Vdash^{\sigma}_{\tilde{\mathsf{C}}} B$ for some w for which there exists n with $\sigma_{\tilde{\delta}(\tilde{\mathsf{C}}B)} \geq n \geq 1$ and $\mathsf{reach}(w, v, n)$.

Remark 4.28. We have $\mathsf{sig}_{\tilde{\mathsf{C}}}(\tilde{\mathsf{C}}A, w) > \mathsf{sig}_{\tilde{\mathsf{C}}}(\tilde{\mathsf{E}}A \vee \tilde{\mathsf{E}}\tilde{\mathsf{C}}A, w)$ as we can show

$$\mathcal{M}, w \Vdash^{\sigma'}_{\tilde{\mathsf{C}}} \tilde{\mathsf{C}}A \text{ if and only if } \mathcal{M}, w \Vdash^{\sigma}_{\tilde{\mathsf{C}}} \tilde{\mathsf{E}}A \vee \tilde{\mathsf{E}}\tilde{\mathsf{C}}A$$

for $\sigma = (\sigma_m, \ldots, \sigma_{\tilde{\delta}(A)}, \ldots, \sigma_1)$ and $\sigma' = (\sigma_m, \ldots, \sigma_{\tilde{\delta}(A)} + 1, \ldots, \sigma_1)$ analogous to Lemma 4.12(3).

Lemma 4.29. *Let $\mathcal{T}$ be an $\mathsf{S_{Dis}}$-tree disproving the sequent $\Gamma = \{A\}$ for some formula A. Then $\mathcal{M}^{\mathcal{T}}, \mathsf{sat}(\Gamma) \not\Vdash A$.*

Proof. Suppose (towards a contradiction) that $\mathcal{M}^{\mathcal{T}}, \mathsf{sat}(\Gamma) \Vdash A$. We show that we can construct a C-thread in some branch of $\mathcal{T}$, which contradicts the assumption that $\mathcal{T}$ disproves A. In order to do so, we will simultaneously construct a branch $\Gamma_0, \Gamma_1, \ldots$ and a thread $A_0, A_1, \ldots$ in $\mathcal{T}$ such that

$$\mathcal{M}^{\mathcal{T}}, \mathsf{sat}(\Gamma_n) \Vdash A_n \text{ for all } n. \tag{4.2}$$

We start with $\Gamma_0 := \Gamma$ and $A_0 := A$. Now assume that we have constructed the thread up to some element $A_n \in \Gamma_n$ with $\mathcal{M}^{\mathcal{T}}, \mathsf{sat}(\Gamma_n) \Vdash A_n$. The next element is selected as follows:

1. If a rule different from $(\Box')$ has been applied, then there is only one child of Γ_n and we let Γ_{n+1} be that child. We have $\mathsf{sat}(\Gamma_n) = \mathsf{sat}(\Gamma_{n+1})$ and distinguish the following cases:
   a) A_n is not the principal formula. We set $A_{n+1} := A_n$.
   b) $A_n = B \vee C$ is the principal formula. We set $A_{n+1} := B$ if
   $$\mathsf{sig}_{\tilde{\mathsf{C}}}(B \vee C, \mathsf{sat}(\Gamma_n)) = \mathsf{sig}_{\tilde{\mathsf{C}}}(B, \mathsf{sat}(\Gamma_{n+1})) \ ;$$
   otherwise, we set $A_{n+1} := C$.
   c) $A_n = B \wedge C$ is the principal formula. We set $A_{n+1} := B$ if B occurs in Γ_{n+1}; otherwise, we set $A_{n+1} := C$.
   d) $A_n = \mathsf{C}B$ is the principal formula. Let $A_{n+1} := \mathsf{E}B \wedge \mathsf{E}\mathsf{C}B$.
   e) $A_n = \tilde{\mathsf{C}}B$ is the principal formula. Let $A_{n+1} := \tilde{\mathsf{E}}B \vee \tilde{\mathsf{E}}\tilde{\mathsf{C}}B$.
2. If $(\Box')$ has been applied, then we have $\mathsf{sat}(\Gamma_n) = \Gamma_n$. We distinguish the following cases:
   a) $A_n = \Box_i B$. There is a child where B is the active formula. Let Γ_{n+1} be that child and set $A_{n+1} := B$.
   b) $A_n = \Diamond_i B$. Because of $\mathcal{M}^{\mathcal{T}}, \mathsf{sat}(\Gamma_n) \Vdash A_n$, there exists a world w with
   $$R_i^{\mathcal{T}}(\mathsf{sat}(\Gamma_n), w) \text{ and } \mathsf{sig}_{\tilde{\mathsf{C}}}(B, w) \leq \mathsf{sig}_{\tilde{\mathsf{C}}}(\Diamond_i B, \mathsf{sat}(\Gamma_n)) \ .$$

The definition of $\mathcal{M}^{\mathcal{T}}$ implies that there is a child Γ' of Γ_n with $\mathsf{sat}(\Gamma') = t$. We set $\Gamma_{n+1} := \Gamma'$ and $A_{n+1} := B$.

c) A_n is not of the form $\Box_i B$ or $\Diamond_i B$. Then there exists $A'_n \in \Gamma_n$ that is of this form such that $\mathcal{M}^{\mathcal{T}}, \Gamma_n \Vdash A'_n$, as the rule has been applied and thus there is another world accessible from Γ_n. We drop the thread constructed so far and continue instead with the thread from A to A'_n.

If the constructed thread were finite, then the last element Γ_n of the path would necessarily be a saturated sequent which would not contain formulae of the form $\Box_i B$. Then the definition of $\mathcal{M}^{\mathcal{T}}$ and an easy induction on the structure of A_n would imply that $\mathcal{M}^{\mathcal{T}}, \Gamma_n \nVdash A_n$, which would contradict (4.2). Hence, the constructed thread is infinite. By construction we have

$$\mathsf{sig}_{\tilde{\mathsf{C}}}(A_0, \mathsf{sat}(\Gamma_0)) \geq \mathsf{sig}_{\tilde{\mathsf{C}}}(A_1, \mathsf{sat}(\Gamma_1)) \geq \dots \ .$$

Assuming that the constructed thread is a $\tilde{\mathsf{C}}$-thread, we can use Remark 4.28 (and hence an argument about signatures similar to the one used in the proof of the soundness theorem 4.15 for S) to find a contradiction to the wellfoundedness of the order on sequences of ordinals of fixed length $\tilde{\delta}(A_0)$. The constructed thread is thus a C-thread. This contradicts the assumption that $\mathcal{T}$ disproves Γ. □

Theorem 4.30 (Completeness of S)**.** *If A is a valid formula, then there exists an S-proof for it.*

Proof. Let A be a formula that is not provable in S. By Theorem 4.23, there exists an $\mathsf{S}_{\mathsf{Dis}}$-tree $\mathcal{T}$ that disproves A. Thus, by Lemma 4.29, there exists a countermodel $\mathcal{M}^{\mathcal{T}}$ for A . Hence, A is not valid. □

5. Justification Logics

Θεαίτητος
ὅ γε ἐγώ, ὦ Σώκρατες, εἰπόντος του ἀκούσας ἐπελελήσμην, νῦν δ᾽ ἐννοῶ· ἔφη δὲ τὴν μὲν μετὰ λόγου.[1]

Plato, Theaetetus 201 *[Pla03]*

Justification logics [AF11] are epistemic logics that explicitly include justifications for an agent's knowledge. Instead of a statement A *is known*, denoted $\Box A$, justification logics reason about justifications for knowledge by using the construct $t : A$ to formalize statements t *is a justification for* A, where, dependent on the application, the *evidence term* t can be viewed as an informal justification or a formal mathematical proof.

Evidence terms are built by means of operations that correspond to the axioms of modal logics, e.g. the justification logic LP corresponds to the modal logic S4, as is illustrated in Fig. 5.1.

Artemov [Art01] has shown that the Logic of Proofs LP is an *explicit*[2] *counterpart* of the modal logic S4 in the following formal sense: *each theorem of* LP *becomes a theorem of* S4 *if all the terms are replaced with the modality* $\Box$*; and, vice versa, each theorem of* S4 *can be transformed into a theorem of* LP *if the occurrences of modality are replaced with suitable evidence terms.* The latter process is called *realization*, and the statement of correspondence is called a *realization theorem.* Note that

[1] **Theaetetus**
Oh yes, I remember now, Socrates, having heard someone make the distinction, but I had forgotten it. He said that knowledge was true opinion accompanied by reason.
Translation from [Pla21].

[2] For other meanings of "explicit" see Sect. 6.8.

S4 **axioms**	LP **axioms**	
$\Box(A \to B) \to (\Box A \to \Box B)$	$t : (A \to B) \to (s : A \to t \cdot s : B)$	(application)
$\Box A \to A$	$t : A \to A$	(reflexivity)
$\Box A \to \Box\Box A$	$t : A \to !t : t : A$	(inspection)
	$t : A \vee s : A \to t + s : A$	(sum)

Figure 5.1.: Axioms of S4 and LP

the operation + introduced by the sum axiom in Fig. 5.1 does not have a modal analog, but it is an essential part of the proof of the realization theorem in [Art01]. Explicit counterparts for many normal modal logics between K and $\mathsf{S5}$ have been developed (see a recent survey in [Art08] and a uniform proof of realization theorems for all justification logics corresponding to logics in the modal cube in [BGK10]).

Historically, Artemov [Art95; Art01] developed the first of these logics, the Logic of Proofs LP, to solve the problem of provability semantics for $\mathsf{S4}$, a longstanding open question by Gödel [Göd33]. Using the realization theorem, theorems of $\mathsf{S4}$ first are transformed into theorems of LP. Formulae A of LP are then translated into formulae A^* in the language of arithmetic such that

$$\mathsf{LP} \vdash A \text{ if and only if } \mathsf{PA} \vdash A^* \ ,$$

where PA denotes Peano arithmetic. The general idea is to translate formulae of the form $t : B$ in such a way that

$$t : B \ \mapsto \ \mathsf{Proof}\left(\overline{t^*}, \overline{\ulcorner B^* \urcorner}\right)$$

where Proof is a proof predicate and t^* is the code of a derivation of the formula B^* with code $\ulcorner B^* \urcorner$, thus giving formal meaning to the interpretation of t as a proof term.

Fitting's model construction [Fit05] provides a natural epistemic semantics for the Logic of Proofs, which can be generalized to the whole family of justification logics. It augments Kripke models with a function that specifies admissible evidence for each formula in a given world. A formula of the form $t : A$ is considered true in a given world w, if

1) A is true in all worlds accessible from w

2) t is admissible evidence for A in w

Single world Fitting models are called Mkrtychev models [Mkr97]. In this special case due to the absence of accessibility relations, the validity of formulae depends solely on the admissible evidence function and the valuation function.

This novel approach has many applications in epistemic logic, see [Art06; Art08; Art10] where amongst others, the famous Gettier example [Get63] is discussed. Furthermore, it makes it possible to tackle the logical omniscience problem [AK09] and to deal with certain forms of self-referentiality [Kuz10].

We will now briefly present some standard justification logics, properties and semantics. For the sake of brevity, only a small selection of justification logics is presented, further logics will be mentioned in Section 5.4

5.1. Syntax

Justification terms are built from constants c_i and variables x_i according to the following grammar:

$$t ::= c_i \mid x_i \mid t \cdot t \mid t + t \mid !t \quad .$$

We denote the set of terms by Tm. A term that does not contain variables is called *ground*.

Formulae are built from atomic propositions p_i according to the following grammar:

$$F ::= p_i \mid \neg F \mid (F \rightarrow F) \mid t : F \quad .$$

Prop denotes the set of atomic propositions and Fm denotes the set of formulae.

The axioms of J consist of all instances of the following schemes:

A1 finitely many schemes axiomatizing classical propositional logic

A2 $t : (A \rightarrow B) \rightarrow (s : A \rightarrow t \cdot s : B)$

A3 $t : A \vee s : A \rightarrow t + s : A$

We will consider extensions of J by the following axiom schemes. We denote the various extensions by L.

(d) $t : \bot \rightarrow \bot$

(t) $t : A \rightarrow A$

(4) $t : A \rightarrow !t : t : A$

A *constant specification* $\mathcal{CS}$ for a logic L is any subset

$$\mathcal{CS} \subseteq \{c : A \mid c \text{ is a constant and } A \text{ is an axiom of } \mathsf{L}\}.$$

A constant specification $\mathcal{CS}$ for a logic L is called

1. *axiomatically appropriate* if for each axiom A of L there is a constant c_i such that $c_i : A \in \mathcal{CS}$

2. *schematic* if for each constant c_i the set $\{A \mid c_i : A \in \mathcal{CS}\}$ consists of instances of several (possible none) axiom schemes (as given above), i.e. every constant justifies certain axiom schemes.

For a constant specification $\mathcal{CS}$, the deductive system $\mathsf{J}_{\mathcal{CS}}$ is the Hilbert system given by the axioms A1–A3 and by the rules modus ponens and axiom necessitation:

$$\frac{A \quad A \to B}{B}\ (\mathrm{MP}) \enspace ,$$

$$\frac{}{\underbrace{!! \cdots !}_{n} c : \underbrace{! \cdots !}_{n-1} c : \cdots :!!c :!c : c : A}\ (\mathrm{AN!}), \text{ where } c : A \in \mathcal{CS} \enspace .$$

In the presence of the (4) axiom, a simplified axiom necessitation rule can be used:

$$\frac{}{c : A}\ (\mathrm{AN}), \text{ where } c : A \in \mathcal{CS} \enspace .$$

Table 5.1 defines the various logics we are going to consider, we will use $\mathsf{L}_{\mathcal{CS}}$ to denote these logics.

5.2. Basic Properties

We will now illustrate the basic justification logic $\mathsf{J}_{\mathcal{CS}}$ using an often cited example from [Art08] which gives a justification logic counterpart to the modal theorem

$$\Box A \vee \Box B \to \Box(A \vee B) \enspace .$$

	A1	A2	A3	(d)	(t)	(4)	(MP)	(AN!)	(AN)
$\mathsf{J}_{\mathcal{CS}}$	✓	✓	✓				✓	✓	
$\mathsf{JD}_{\mathcal{CS}}$	✓	✓	✓	✓			✓	✓	
$\mathsf{JT}_{\mathcal{CS}}$	✓	✓	✓		✓		✓	✓	
$\mathsf{JD4}_{\mathcal{CS}}$	✓	✓	✓	✓		✓	✓		✓
$\mathsf{J4}_{\mathcal{CS}}$	✓	✓	✓			✓	✓		✓
$\mathsf{LP}_{\mathcal{CS}}$	✓	✓	✓		✓	✓	✓		✓

Table 5.1.: Deductive Systems $\mathsf{L}_{\mathcal{CS}}$.

Example 5.1. Assume we are given $\mathsf{J}_{\mathcal{CS}}$ with $a : [A \to (A \vee B)] \in \mathcal{CS}$ and $b : [B \to (A \vee B)]$. Then the following is a theorem of $\mathsf{J}_{\mathcal{CS}}$

$$(x : A \vee y : B) \to (a \cdot x + b \cdot y) : (A \vee B) \ .$$

Proof. From (AN!) we get

$$a : [A \to (A \vee B)] \text{ and } b : [B \to (A \vee B)] \ .$$

Using A2 and (MP), we obtain

$$x : A \to (a \cdot x) : (A \vee B) \text{ and } y : B \to (b \cdot y) : (A \vee B)$$

Finally, from A3 we have

$$(a \cdot x) : (A \vee B) \to (a \cdot x + b \cdot y) : (A \vee B)]$$

and

$$(b \cdot y) : (A \vee B) \to (a \cdot x + b \cdot y) : (A \vee B)] \ .$$

Using propositional reasoning, we obtain the desired result. □

Theorem 5.2 (Constructive necessitation)**.** *Let $\mathcal{CS}$ be an axiomatically appropriate constant specification for L. Then $\mathsf{L}_{\mathcal{CS}}$ enjoys the internalization property, i.e.*

$$\textit{If } \vdash A \textit{ then there is a ground term } t \textit{ such that } \vdash t : A \ .$$

Justification logic	Corresponding modal logic
$\mathsf{J}_{\mathcal{CS}}$	K
$\mathsf{JD}_{\mathcal{CS}}$	KD
$\mathsf{JT}_{\mathcal{CS}}$	KT
$\mathsf{JD4}_{\mathcal{CS}}$	KD4
$\mathsf{J4}_{\mathcal{CS}}$	K4
$\mathsf{LP}_{\mathcal{CS}}$	S4

Table 5.2.: Corresponding modal logics

Definition 5.3 (Forgetful projection, realization)**.** Forgetful projection $.^\circ$ is a mapping from formulae of justification logic to formulae of modal logic, defined by

- $P^\circ := P$,
- $.^\circ$ commutes with propositional connectives,
- $(t : A)^\circ := \Box A^\circ \quad .$

A realization $.^r$ is a mapping from formulae of modal logic to formulae of justification logic such that for all formulae A we have $(A^r)^\circ = A$.

See [BGK10; GK12] for a uniform proof of the following theorem (and more realization results).

Theorem 5.4 (Realization theorem)**.** *Let $\mathcal{CS}$ be axiomatially appropriate. Then $\mathsf{L}^\circ_{\mathcal{CS}}$ is exactly the corresponding modal logic given in Table 5.2, i.e., the forgetful projection of each theorem of $\mathsf{L}_{\mathcal{CS}}$ is a theorem of the corresponding modal logic; and, vice-versa, for each theorem A of the corresponding modal logic, there is a realization r such that A^r is a theorem of $\mathsf{L}_{\mathcal{CS}}$.*

Note that while it is easy to prove that the forgetful projection of the justification logic is included in the corresponding modal logic, the reverse inclusion is much more involved and usually requires a cut-free sequent system for the modal logic (see [BGK10; GK12] for a survey). Notable exceptions to this can be found in [Fit05; Fit10] where a (non-constructive) semantic method is used and [Fit11] where reduction to another logic is the main tool. A comparable reduction technique is also used for the realization result presented in Chapter B, Section B.5.

5.3. Semantics

The semantics of our logics are given by so-called Fitting-models $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ where (W, R) is a usual Kripke-frame, i.e. W is a non-empty set of possible worlds and $R \subset W \times W$ is an accessibility relation.

Definition 5.5 (Evidence relation)**.** Let (W, R) be a Kripke frame and $\mathcal{CS}$ a constant specification. An admissible evidence relation $\mathcal{E}$ for a logic $\mathsf{L}_{\mathcal{CS}}$ is a subset of $\mathrm{Tm} \times \mathrm{Fm} \times W$ that satisfies the closure conditions:

1. if $(s, A, w) \in \mathcal{E}$ or $(t, A, w) \in \mathcal{E}$, then $(s + t, A, w) \in \mathcal{E}$,

2. if $(s, A \to B, w) \in \mathcal{E}$ and $(t, A, w) \in \mathcal{E}$, then $(s \cdot t, B, w) \in \mathcal{E}$.

Depending on whether the logic $\mathsf{L}_{\mathcal{CS}}$ contains the (4) axiom, the evidence function has to satisfy one of the two following sets of closure conditions. If $\mathsf{L}_{\mathcal{CS}}$ does not include the (4) axiom, then the additional requirement is:

3. if $c : A \in CS$ and $w \in W$, then

$$(\underbrace{! \cdots !!}_{n} c, \underbrace{! \cdots !!}_{n-1} c : \cdots : !c : c : A, w) \in \mathcal{E} \ .$$

If $\mathsf{L}_{\mathcal{CS}}$ includes the (4) axiom, then the additional requirement is:

4. if $c : A \in CS$ and $w \in W$, then $(c, A, w) \in \mathcal{E}$,

5. if $(t, A, w) \in \mathcal{E}$, then $(!t, t : A, w) \in \mathcal{E}$,

6. if $(t, A, w) \in \mathcal{E}$ and wRv, then $(t, A, v) \in \mathcal{E}$.

Sometimes we write $\mathcal{E}(s, A, w)$ for $(s, A, w) \in \mathcal{E}$.

Remark 5.6. To facilitate notations in Chapter A we use an evidence *relation* $\mathcal{E}^r \subseteq \mathrm{Tm} \times \mathrm{Fm} \times W$ contrary to the usual approach of an evidence *function* $\mathcal{E}^f : W \times \mathrm{Tm} \to \mathcal{P}(\mathrm{Fm})$. Obviously, these two approaches are interchangeable:

$$\mathcal{E}^r(t, A, w) \text{ if and only if } A \in \mathcal{E}^f(w, t) \ .$$

On the other hand, in Chapters 6, 7 and B, logics will be presented using the "traditional" form of evidence functions.

Definition 5.7 (Model)**.** Let $\mathcal{CS}$ be a constant specification. A *Fitting model* for a logic $\mathsf{L}_{\mathcal{CS}}$ is a quadruple $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ where

- (W, R) is a Kripke frame such that
  - if $\mathsf{L}_{\mathcal{CS}}$ includes the (4) axiom, then R is transitive,
  - if $\mathsf{L}_{\mathcal{CS}}$ includes the (t) axiom, then R is reflexive,
  - if $\mathsf{L}_{\mathcal{CS}}$ includes the (d) axiom, then R is serial
- $\mathcal{E}$ is an admissible evidence relation for $\mathsf{L}_{\mathcal{CS}}$ over the frame (W, R),
- $\nu : \mathrm{Prop} \to \mathcal{P}(W)$ is a valuation function.

Definition 5.8 (Satisfaction relation)**.** The relation of formula A being satisfied in a Fitting model $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ at a world $w \in W$ is defined by induction on the structure of A by

- $\mathcal{M}, w \Vdash p_i$ if and only if $w \in \nu(p_i)$
- $\Vdash$ commutes with Boolean connectives
- $\mathcal{M}, w \Vdash t : B$ if and only if
  1) $\mathcal{M}, v \Vdash B$ for all $v \in W$ with wRv and
  2) $(t, B, w) \in \mathcal{E}$

We say a formula A is valid in a Fitting model $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ if for all $w \in W$ we have $\mathcal{M}, w \Vdash A$. We say a formula A is valid for a logic $\mathsf{L}_{\mathcal{CS}}$ if for all Fitting models $\mathcal{M}$ for $\mathsf{L}_{\mathcal{CS}}$ we have that A is valid in $\mathcal{M}$.

Example 5.9. Let us also briefly consider an example of a Fitting model given by the following situation: the meteorology textbooks tell us that chinooks (the North American version of the foehn winds) lead to warm

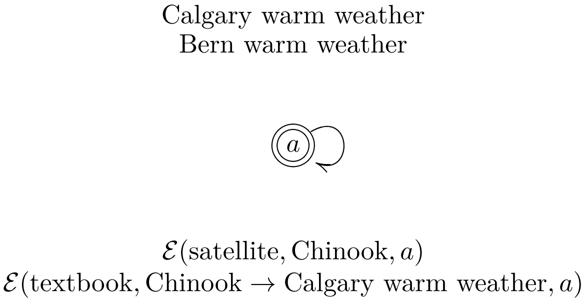

Figure 5.2.: A simple example of a Fitting model.

weather in Calgary. As it happens, we have satellite evidence for a chinook and we conclude that the weather is warm in Calgary. On the other hand, while staring at the computer screen with the satellite pictures, we neglected to have a look outside. So, even though we are very certain that the weather in Bern is also warm (when we last looked out the window the sun was shining and we can not imagine the situation to have changed meanwhile), we have no evidence for this. This very simple example is depicted in Figure 5.2.[3] We can easily see that

$$\mathcal{M}, a \Vdash (\text{textbook} \cdot \text{satellite}) : (\text{Calgary warm weather}),$$

but $\mathcal{M}, a \nVdash t : (\text{Bern warm weather})$ for any evidence term t. We might also consider the situation depicted in Figure 5.3. Here the radio station announced warm weather in Bern, but we also know that the radio station has a particularly bad reputation when it comes to weather forecasts. Again, we have $\mathcal{M}, a \nVdash t : (\text{Bern warm weather})$ for any evidence term t, but this time for different reasons.

The logics defined above are sound and complete (with a restriction in case of the logics containing the (d) axiom). See [Fit05; Pac05; Art08] for the full proofs of the following results.

[3] The special case of single world Fitting models is called M-models, see Definition 6.19 in Chapter 6.

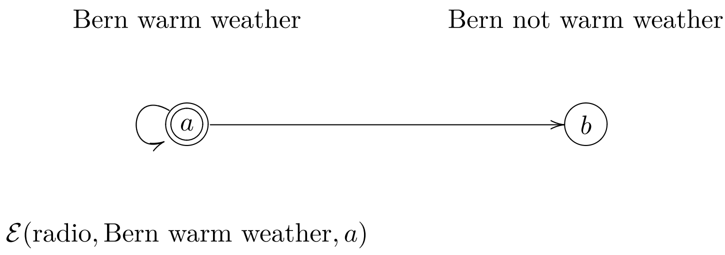

Figure 5.3.: Another simple example of a Fitting model.

Soundness can be obtained by an easy induction on the derivation of the formula.

Theorem 5.10 (Soundness)**.** *Let $\mathcal{CS}$ be a constant specification. If a formula A is derivable in a logic $\mathsf{L}_{\mathcal{CS}}$, then A is valid for $\mathsf{L}_{\mathcal{CS}}$.*

For completeness, a canonical model construction is used. The axiomatical appropriateness of the constant specification in case the logic contains the (d) axiom is necessary to show the seriality condition on the accessibility relation (see also the proof of Corollary 7.26).

Theorem 5.11 (Completeness)**.**

1. *Let $\mathcal{CS}$ be an arbitrary constant specification. If a formula A is is not derivable in the logic $\mathsf{L}_{\mathcal{CS}} \in \{\mathsf{J}_{\mathcal{CS}}, \mathsf{JT}_{\mathcal{CS}}, \mathsf{J4}_{\mathcal{CS}}, \mathsf{LP}_{\mathcal{CS}}\}$, then there exists a Fitting model $\mathcal{M}$ for $\mathsf{L}_{\mathcal{CS}}$ with $\mathcal{M}, w \not\Vdash A$ for some world w in $\mathcal{M}$.*

2. *Let $\mathcal{CS}$ be an axiomatically appropriate constant specification. If a formula A is not derivable in $\mathsf{L}_{\mathcal{CS}} \in \{\mathsf{JD}_{\mathcal{CS}}, \mathsf{JD4}_{\mathcal{CS}}\}$, then there exists a Fitting model $\mathcal{M}$ for $\mathsf{L}_{\mathcal{CS}}$ with $\mathcal{M}, w \not\Vdash A$ for some world w in $\mathcal{M}$.*

5.4. Further Justification Logics

A further axiomatic extension corresponding to the negative inspection axiom (5) is

$$\neg t : A \rightarrow ?t : (\neg t : A)$$

introduced in [Pac05; Rub06]. In order to simplify this introduction, we did not include this axiom yet, but we will use it in the logics presented in Chapter 7, and thus refer there for further information.

In [AN05], modal and justification logics are combined, the connecting princple being "what is known for a reason is known" or "whatever is known explictly is known implicitly", formally

$$t : A \rightarrow \Box A \ .$$

In [Art06], a variant of this is considered, namely a combination of multi-modal and justification logic where "whatever is known for a reason is known by each agent"

$$t : A \rightarrow \Box_i A \qquad \text{for all agents } i \ .$$

Here the main interest lies in the forgetful projection of this logic, giving rise to a modality J named "justified common knowledge", which we will discuss in the following chapter.

In [Yav08], a two-agent justification logic with certain communication principles among agents is presented. We will also discuss this logic in the following chapter where we will present a generalized (multi-agent) version of it.

Furthermore, justification logics with quantifiers are available but we will not discuss them here, as they are beyond the scope of this thesis (see [Yav01] and [Fit08a]).

See [AF11] for an overview of justification logics so far, and the extensive justification logic bibliography[Kuz12].

6. A Justification Logic with Common Knowledge

RECONSIDER, *v.* To seek a justification for a decision already made.

Ambrose Bierce, The Devil's Dictonary [Bie11]

In this chapter, we develop a multi-agent justification logic with evidence terms for individual agents as well as for common knowledge, with the intention to provide an explicit counterpart of the h-agent modal logic of traditional common knowledge $\mathsf{S4}_h^{\mathsf{C}}$ as presented in Chapter 3. For the sake of compactness and readability, we will not yet treat groups of agents and different base logics in this chapter, but postpone this to Chapter 7.

Let us remember some basic concepts first and give a short survey of related logics. We have already seen that common knowledge of A is defined as the infinitary conjunction *everybody knows A and everybody knows that everybody knows A and so on.* This is equivalent to saying that common knowledge of A is the greatest fixed point of

$$\lambda X.(\text{everybody knows } A \text{ and everybody knows } X) \ . \qquad (6.1)$$

An explicit counterpart of McCarthy's *any fool knows* common knowledge modality [McC+78], where common knowledge of A is defined as an arbitrary fixed point of (6.1), is presented in [Art06]. The relationship between the traditional common knowledge as presented in [Fag+95; MH95] and McCarthy's version is studied in [Ant07].

Multi-agent justification logics with evidence terms for each agent are considered in [Yav08; Ren09a; Art10], but common knowledge is not present in any of them. Renne's system combines features of modal and dynamic epistemic logics [Ren09a] and hence cannot be directly compared to our system. Artemov's interest lies mostly in exploring a case of two agents with unequal epistemic powers: e.g., Artemov's Observer has sufficient evidence to reproduce the Object Agent's thinking, but not vice versa [Art10]. Yavorskaya studies various operations of evidence transfer between agents [Yav08]. Yavorskaya's minimal[1] two-agent justification logic LP^2, which is an explicit counterpart of $\mathsf{S4}_2$, is the closest to our system. We will show that in the case of two agents our system is a conservative extension of LP^2.

On the semantic side we have Fittings *F-models* [Fit05] and Mkrtychev's independently proved stronger completeness result for LP with respect to singleton F-models [Mkr97], now known as *M-models*. Artemov extends F-models to the language with both evidence terms for McCarthy's common knowledge modality and ordinary modalities for the individual agents [Art06], creating the most general type of epistemic models, sometimes called *AF-models*, where common evidence terms are given their own accessibility relation, which does not directly depend on the accessibility relations for individual modalities. The absence of ordinary modalities in Yavorskaya's two-agent justification systems provides for a stronger completeness result with respect to M-models [Yav08].

This chapter is organized as follows. In Section 6.1, we introduce a language and give an axiomatization of a family of multi-agent justification logics with common knowledge. In Section 6.2, we prove their basic properties including the internalization property, which is characteristic of all justification logics. In Section 6.3, we develop an epistemic semantics and prove soundness and completeness with respect to this semantics as well as with respect to singleton models, thereby demonstrating the finite model property. In Section 6.5, we show that for the two-agent case, our logic is a conservative extension of Yavorskaya's minimal two-agent justification logic. In Section 6.6, we demonstrate how our logic is related to the modal logic of traditional common knowledge and discuss the problem of realization. In Section 6.7, we provide an analysis of the coordinated attack problem in our logic. Finally, in Section 6.8, we discuss how the newly introduced terms affect the agents, including their ability to communicate information in various

[1]Minimality here is understood in the sense of the minimal transfer of evidence.

communication modes.

The material in this chapter was published as [BKS10a; BKS11a].

6.1. Syntax

To create an explicit counterpart of the modal logic of common knowledge $\mathsf{S4}_h^{\mathsf{C}}$, we use its axiomatization via the induction axiom from [MH95] rather than via the induction rule (see Chapter 3, Section 3.2) to facilitate proving the internalization property for the resulting justification logic. We supply each agent with its own copy of terms from the Logic of Proofs, while terms for common and mutual knowledge employ additional operations. The fact that each agent has its own set of operations makes our framework more flexible. For instance, agents may be thought of as representing different arithmetical proof systems that use different encodings (cf. [Yav08]).

We also slightly change notation from $t :_x A$ from the previous chapter to $[t]_x A$ in order to indicate that "t is evidence of type x for A", but this change is of purely cosmetic nature.

Motivated by the coinductive proof system S from Chapter 4, a proof of $\mathsf{C}A$ can be viewed as an infinite list of proofs of the conjuncts $\mathsf{E}^m A$ from the representation of common knowledge through an infinite conjunction. To generate a finite representation of this infinite list, we use an explicit counterpart of the induction axiom (I-Ax)

$$[t]_{\mathsf{C}}(A \to [r]_{\mathsf{E}}A) \to ([s]_{\mathsf{E}}A \to [\mathsf{ind}(t, s)]_{\mathsf{C}}A)$$

with a binary operation $\mathsf{ind}(\cdot, \cdot)$. To facilitate access to the elements of the list, explicit counterparts of the co-closure axiom (Co-Cl)provide evidence terms that can be seen as splitting the infinite list into its head and tail,

$$[t]_{\mathsf{C}}A \to [\mathsf{head}(t)]_{\mathsf{E}}A \ , \quad [t]_{\mathsf{C}}A \to [\mathsf{tail}(t)]_{\mathsf{E}}\, [t]_{\mathsf{C}}A \ ,$$

by means of two unary co-closure operations $\mathsf{head}(\cdot)$ and $\mathsf{tail}(\cdot)$. One might raise the question why the term r from the antecedent of the induction rule does not appear in the conclusion, namely in the $\mathsf{ind}(t, s)$ term. This design choice can be justified by noting that the term t in the antecedent already encompasses the term r and can be seen like a function producing the term r for a given formula A. With the aim of keeping the logic as simple as possible we therefore omit this term in the conclusion (see also the discussion about operations in the introduction to Chapter B).

Evidence terms for mutual knowledge are viewed as tuples of the individual agents' evidence terms. The standard tupling operation and h unary projections are employed as means of translation between the individual agents' and mutual knowledge evidence. Note that, strictly speaking, evidence terms for mutual knowledge are not necessary because they could be defined, just like the modality for mutual knowledge can be defined in the modal case. However, the resulting system would be very cumbersome in notation and usage.

While only two of the three operations on LP terms (see Chapter 5, Fig. 5.1) are adopted for common knowledge evidence and none is adopted for mutual knowledge evidence, it will be shown in Section 6.2 that three out of the four remaining operations are definable, with a notable exception of inspection for mutual knowledge, as is to be expected. While the usage of the application operation for common knowledge evidence terms is justifiable on the grounds of the corresponding modal (K) axiom for common knowledge (see 3), the necessity of the sum operation for common knowledge evidence terms is less clear and can only be shown once the realization theorem is proved (see Section 6.6 for details).

We consider a system of h agents. Throughout this chapter, i always denotes an element of $\{1, \dots, h\}$, $*$ always denotes an element of $\{1, \dots, h, \mathsf{C}\}$, and $\circledast$ always denotes an element of $\{1, \dots, h, \mathsf{E}, \mathsf{C}\}$.

Let $\mathrm{Cons}_{\circledast} := \{c_1^{\circledast}, c_2^{\circledast}, \dots\}$ and $\mathrm{Var}_{\circledast} := \{x_1^{\circledast}, x_2^{\circledast}, \dots\}$ be countable sets of *proof constants* and *proof variables*, respectively, for each type of knowledge $\circledast$. The sets $\mathrm{Tm}_1, \dots, \mathrm{Tm}_h$, Tm_{E}, and Tm_{C} of *evidence terms for individual agents* and for *mutual* and *common knowledge* respectively are inductively defined as follows:

1. $\mathrm{Cons}_{\circledast} \subseteq \mathrm{Tm}_{\circledast}$ and $\mathrm{Var}_{\circledast} \subseteq \mathrm{Tm}_{\circledast}$;

2. $!_i t \in \mathrm{Tm}_i$ for any $t \in \mathrm{Tm}_i$;

3. $t +_* s \in \mathrm{Tm}_*$ and $t \cdot_* s \in \mathrm{Tm}_*$ for any $t, s \in \mathrm{Tm}_*$;

4. $\langle t_1, \dots, t_h \rangle \in \mathrm{Tm}_{\mathsf{E}}$ for any $t_1 \in \mathrm{Tm}_1, \dots, t_h \in \mathrm{Tm}_h$;

5. $\pi_i t \in \mathrm{Tm}_i$ for any $t \in \mathrm{Tm}_{\mathsf{E}}$;

6. $\mathsf{head}(t) \in \mathrm{Tm}_{\mathsf{E}}$ and $\mathsf{tail}(t) \in \mathrm{Tm}_{\mathsf{E}}$ for any $t \in \mathrm{Tm}_{\mathsf{C}}$;

7. $\mathsf{ind}(t, s) \in \mathrm{Tm}_{\mathsf{C}}$ for any $t \in \mathrm{Tm}_{\mathsf{C}}$ and any $s \in \mathrm{Tm}_{\mathsf{E}}$.

$\mathrm{Tm} := \mathrm{Tm}_1 \cup \ldots \cup \mathrm{Tm}_h \cup \mathrm{Tm}_{\mathsf{E}} \cup \mathrm{Tm}_{\mathsf{C}}$ denotes the set of all evidence terms. The indices of the operations !, +, and $\cdot$ will most often be omitted if they can be inferred from the context. A term is called *ground* if no proof variables occur in it.

Let $\mathrm{Prop} := \{P_1, P_2, \ldots\}$ be a countable set of *propositional variables*. *Formulae* are denoted by $A, B, C, \ldots$ and are defined by the grammar

$$A ::= P_j \mid \neg A \mid (A \to A) \mid [t]_{\circledast} A \ ,$$

where $t \in \mathrm{Tm}_{\circledast}$ and $P_j \in \mathrm{Prop}$. The set of all formulae is denoted by $\mathrm{Fm}_{\mathsf{LP}_h^{\mathsf{C}}}$. Conjunction, disjunction and equivalence is defined as usual. We adopt the following convention: whenever a formula $[t]_{\circledast} A$ is used, it is assumed to be well-formed: i.e., it is implicitly assumed that term $t \in \mathrm{Tm}_{\circledast}$. This enables us to omit the explicit typification of terms.

The axioms of $\mathsf{LP}_h^{\mathsf{C}}$ are given by:

1. all propositional tautologies
2. $[t]_*(A \to B) \to ([s]_* A \to [t \cdot s]_* B)$ (application)
3. $[t]_* A \vee [s]_* A \to [t + s]_* A$ (sum)
4. $[t]_i A \to A$ (reflexivity)
5. $[t]_i A \to [!t]_i \, [t]_i A$ (inspection)
6. $[t_1]_1 A \wedge \ldots \wedge [t_h]_h A \to [\langle t_1, \ldots, t_h \rangle]_{\mathsf{E}} A$ (tupling)
7. $[t]_{\mathsf{E}} A \to [\pi_i t]_i A$ (projection)
8. $[t]_{\mathsf{C}} A \to [\mathsf{head}(t)]_{\mathsf{E}} A, \quad [t]_{\mathsf{C}} A \to [\mathsf{tail}(t)]_{\mathsf{E}} \, [t]_{\mathsf{C}} A$ (co-closure)
9. $[t]_{\mathsf{C}}(A \to [r]_{\mathsf{E}} A) \to ([s]_{\mathsf{E}} A \to [\mathsf{ind}(t, s)]_{\mathsf{C}} A)$ (induction)

A *constant specification* $\mathcal{CS}$ is any subset

$$\mathcal{CS} \subseteq \bigcup_{\circledast \in \{1,\dots,h,\mathsf{E},\mathsf{C}\}} \left\{ [c]_{\circledast} A : c \in \mathrm{Cons}_{\circledast} \text{ and } A \text{ is an axiom of } \mathsf{LP}_h^{\mathsf{C}} \right\}.$$

A constant specification $\mathcal{CS}$ is called C-*axiomatically appropriate* if, for each axiom A, there is a proof constant $c \in \mathrm{Cons}_{\mathsf{C}}$ such that $[c]_{\mathsf{C}} A \in \mathcal{CS}$. A constant specification $\mathcal{CS}$ is called *homogeneous*, if $\mathcal{CS} \subseteq \{[c]_{\circledast} A : c \in \mathrm{Cons}_{\circledast} \text{ and } A \text{ is an axiom}\}$ for some fixed $\circledast$: i.e., if for all $[c]_{\circledast} A \in \mathcal{CS}$ the constants c are of the same type.

For a constant specification $\mathcal{CS}$, the deductive system $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS})$ is the Hilbert system given by the axioms of $\mathsf{LP}_h^{\mathsf{C}}$ above and by the rules modus ponens and axiom necessitation:

$$\frac{A \quad A \to B}{B}\ (\mathrm{MP})\ , \qquad \frac{}{[c]_{\circledast} A}\ (\mathrm{AN})\ , \text{ where } [c]_{\circledast} A \in \mathcal{CS}.$$

By $\mathsf{LP}_h^{\mathsf{C}}$ we denote the system $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS})$ with

$$\mathcal{CS} = \left\{ [c]_{\mathsf{C}} A \ : \ c \in \mathrm{Cons}_{\mathsf{C}} \text{ and } A \text{ is an axiom of } \mathsf{LP}_h^{\mathsf{C}} \right\} \ . \qquad (6.2)$$

As usual in Hilbert-style systems, we say a formula A is derivable from a set of formulae Δ, if there is a proof of A using elements of Δ as additional axioms. However, note that constant specifications are not extended in order to also include elements of Δ. For an arbitrary $\mathcal{CS}$, we write $\Delta \vdash_{\mathcal{CS}} A$ to state that A is derivable from a set of formulae Δ in $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS})$ and omit $\mathcal{CS}$ when working with the constant specification from (6.2) by writing $\Delta \vdash A$. We also omit Δ when $\Delta = \varnothing$ and write $\vdash_{\mathcal{CS}} A$ or $\vdash A$, in which case A is called a theorem of $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS})$ or of $\mathsf{LP}_h^{\mathsf{C}}$ respectively. In order to emphasize this, we also sometimes write $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash A$ and $\mathsf{LP}_h^{\mathsf{C}} \vdash A$, repsectively. We use Δ, A to mean $\Delta \cup \{A\}$.

6.2. Basic Properties

In this section, we show that our logic possesses the standard properties expected of any justification logic. In addition, we show that the operations on terms introduced in the previous section are sufficient to express the operations of sum and application for mutual knowledge evidence and the operation of inspection for common knowledge evidence. This is the reason why $+_{\mathsf{E}}$, $\cdot_{\mathsf{E}}$, and $!_{\mathsf{C}}$ are not primitive connectives in the language. It should be noted that no inspection operation for mutual evidence terms can be defined, which follows from Lemma 6.29 in Section 6.6 and the fact that $\mathsf{E}A \to \mathsf{E}\mathsf{E}A$ is not a valid modal formula.

Lemma 6.1. *For any constant specification $\mathcal{CS}$ and any formulae A and B:*

1. $\vdash_{\mathcal{CS}} [t]_{\mathsf{E}}A \to A$ *for all* $t \in \mathrm{Tm}_{\mathsf{E}}$; (E-*reflexivity*)
2. *for any* $t, s \in \mathrm{Tm}_{\mathsf{E}}$, *there is a term* $t \cdot_{\mathsf{E}} s \in \mathrm{Tm}_{\mathsf{E}}$ *such that* $\vdash_{\mathcal{CS}} [t]_{\mathsf{E}}(A \to B) \to ([s]_{\mathsf{E}}A \to [t \cdot_{\mathsf{E}} s]_{\mathsf{E}}B)$; (E-*application*)
3. *for any* $t, s \in \mathrm{Tm}_{\mathsf{E}}$, *there is a term* $t +_{\mathsf{E}} s \in \mathrm{Tm}_{\mathsf{E}}$ *such that* $\vdash_{\mathcal{CS}} [t]_{\mathsf{E}}A \vee [s]_{\mathsf{E}}A \to [t +_{\mathsf{E}} s]_{\mathsf{E}}A$; (E-*sum*)
4. *for any* $t \in \mathrm{Tm}_{\mathsf{C}}$ *and any* $i \in \{1, \ldots, h\}$, *there is a term* $\downarrow_i t \in \mathrm{Tm}_i$ *such that* $\vdash_{\mathcal{CS}} [t]_{\mathsf{C}}A \to [\downarrow_i t]_i A$; ($i$-*conversion*)
5. $\vdash_{\mathcal{CS}} [t]_{\mathsf{C}}A \to A$ *for all* $t \in \mathrm{Tm}_{\mathsf{C}}$. (C-*reflexivity*)

Proof.

1. Immediate by the projection and reflexivity axioms.
2. Set $t \cdot_{\mathsf{E}} s := \langle \pi_1 t \cdot_1 \pi_1 s, \ldots, \pi_h t \cdot_h \pi_h s \rangle$.
3. Set $t +_{\mathsf{E}} s := \langle \pi_1 t +_1 \pi_1 s, \ldots, \pi_h t +_h \pi_h s \rangle$.
4. Set $\downarrow_i t := \pi_i \mathsf{head}(t)$.
5. Immediate by 4. and the reflexivity axiom. □

Unlike Lemma 6.1, Lemma 6.2 requires that a constant specification $\mathcal{CS}$ be C-axiomatically appropriate.

Lemma 6.2. *Let $\mathcal{CS}$ be* C*-axiomatically appropriate and A be a formula.*

1. *For any* $t \in \mathrm{Tm}_{\mathsf{C}}$, *there is a term* $!_{\mathsf{C}}t \in \mathrm{Tm}_{\mathsf{C}}$ *such that* $\vdash_{\mathcal{CS}} [t]_{\mathsf{C}}A \to [!_{\mathsf{C}}t]_{\mathsf{C}}\,[t]_{\mathsf{C}}A$. (C-*inspection*)
2. *For any* $t \in \mathrm{Tm}_{\mathsf{C}}$, *there is a term* $\Leftarrow t \in \mathrm{Tm}_{\mathsf{C}}$ *such that* $\vdash_{\mathcal{CS}} [t]_{\mathsf{C}}A \to [\Leftarrow t]_{\mathsf{C}}\,[\mathsf{head}(t)]_{\mathsf{E}}A$. (C-*shift*)

Proof. 1. As $\mathcal{CS}$ is C-axiomatically appropriate, there is a c such that

$$[c]_{\mathsf{C}}([t]_{\mathsf{C}}A \to [\mathsf{tail}(t)]_{\mathsf{E}}\,[t]_{\mathsf{C}}A) \in \mathcal{CS} \ .$$

We then have $[c]_{\mathsf{C}}([t]_{\mathsf{C}}A \to [\mathsf{tail}(t)]_{\mathsf{E}}\,[t]_{\mathsf{C}}A)$ by axiom necessitation. From this, modus ponens and the following instance of the induction axiom

$$[c]_{\mathsf{C}}([t]_{\mathsf{C}}A \to [\mathsf{tail}(t)]_{\mathsf{E}}\,[t]_{\mathsf{C}}A) \\ \to ([\mathsf{head}(t)]_{\mathsf{E}}A \to [\mathsf{ind}(c, \mathsf{head}(t))]_{\mathsf{C}}A)$$

we get $[\mathsf{head}(t)]_{\mathsf{E}}A \to [\mathsf{ind}(c, \mathsf{head}(t))]_{\mathsf{C}}A$. Furthermore, we have the following instance of the co-closure axiom

$$[t]_{\mathsf{C}}A \to [\mathsf{head}(t)]_{\mathsf{E}}A$$

and by propositional reasoning we finally get

$$[t]_{\mathsf{C}}A \to [\mathsf{ind}(c, \mathsf{head}(t))]_{\mathsf{C}}A \ .$$

We can thus set $!_{\mathsf{C}}t := \mathsf{ind}(c, \mathsf{head}(t))$.

2. As in the previous case, the C-axiomatical appropriateness guarantees the existence of a c' such that

$$[c']_{\mathsf{C}}([t]_{\mathsf{C}}A \to [\mathsf{head}(t)]_{\mathsf{E}}A) \in \mathcal{CS} \ .$$

Using modus ponens and the following instance of the application axiom

$$[c']_{\mathsf{C}}([t]_{\mathsf{C}}A \to [\mathsf{head}(t)]_{\mathsf{E}}A) \\ \to ([!_{\mathsf{C}}t]_{\mathsf{C}}\,[t]_{\mathsf{C}}A \to [c' \cdot !_{\mathsf{C}}]_{\mathsf{C}}\,[\mathsf{head}(t)]_{\mathsf{E}}A)$$

we get

$$[!_{\mathsf{C}}t]_{\mathsf{C}}\,[t]_{\mathsf{C}}A \to [c' \cdot !_{\mathsf{C}}]_{\mathsf{C}}\,[\mathsf{head}(t)]_{\mathsf{E}}A$$

Using propositional reasoning and $[t]_{\mathsf{C}}A \to [!_{\mathsf{C}}t]_{\mathsf{C}}\,[t]_{\mathsf{C}}A$ we obtain the desired result and can thus set $\Lleftarrow t := c' \cdot_{\mathsf{C}} (!_{\mathsf{C}}t)$. ☐

The following two lemmas are standard in justification logics. Their proofs can be taken almost word for word from [Art01] and are, therefore, omitted here.

Lemma 6.3 (Deduction Theorem)**.** *Let $\mathcal{CS}$ be a constant specification and $\Delta \cup \{A, B\} \subseteq \mathrm{Fm}_{\mathsf{LP}^{\mathsf{C}}_h}$. Then*

$$\Delta, A \vdash_{\mathcal{CS}} B \textit{ if and only if } \Delta \vdash_{\mathcal{CS}} A \to B \enspace .$$

Lemma 6.4 (Substitution)**.** *For any constant specification $\mathcal{CS}$, any propositional variable P, any $\Delta \cup \{A, B\} \subseteq \mathrm{Fm}_{\mathsf{LP}^{\mathsf{C}}_h}$, any $x \in \mathrm{Var}_{\circledast}$, and any $t \in \mathrm{Tm}_{\circledast}$,*

$$\textit{if } \Delta \vdash_{\mathcal{CS}} A, \qquad \textit{then } \Delta(x/t, P/B) \vdash_{\mathcal{CS}(x/t,P/B)} A(x/t, P/B) \enspace ,$$

where $A(x/t, P/B)$ denotes the formula obtained by simultaneously replacing all occurrences of x in A with t and all occurrences of P in A with B and $\Delta(x/t, P/B)$ and $\mathcal{CS}(x/t, P/B)$ are defined accordingly.

The following lemma states that our logic can internalize its own proofs, which is an important property of justification logics.

Lemma 6.5 (C-lifting)**.** *Let $\mathcal{CS}$ be a homogeneous C-axiomatically appropriate constant specification. For any formulae A, $B_1, \ldots, B_n$, $C_1, \ldots, C_m$ and any terms $s_1, \ldots, s_n \in \mathrm{Tm}_{\mathsf{C}}$, if*

$$[s_1]_{\mathsf{C}} B_1, \ldots, [s_n]_{\mathsf{C}} B_n, C_1, \ldots, C_m \vdash_{\mathcal{CS}} A \enspace ,$$

then for each $\circledast$ there is a term $t_{\circledast}(x_1^{\mathsf{C}}, \ldots, x_n^{\mathsf{C}}, y_1^{\circledast}, \ldots, y_m^{\circledast}) \in \mathrm{Tm}_{\circledast}$ such that

$$\begin{aligned} &[s_1]_{\mathsf{C}} B_1, \ldots, [s_n]_{\mathsf{C}} B_n, [y_1]_{\circledast} C_1, \ldots, [y_m]_{\circledast} C_m \\ &\qquad\qquad\qquad\qquad \vdash_{\mathcal{CS}} [t_{\circledast}(s_1, \ldots, s_n, y_1, \ldots, y_m)]_{\circledast} A \end{aligned}$$

for fresh variables $x_1, \ldots, x_n \in \mathrm{Var}_{\mathsf{C}}$ and $y_1, \ldots, y_m \in \mathrm{Var}_{\circledast}$.

Proof. We proceed by induction on the derivation of A.

If A is an axiom, there is a constant $c \in \mathrm{Cons}_{\mathsf{C}}$ such that $[c]_{\mathsf{C}} A \in \mathcal{CS}$ because $\mathcal{CS}$ is C-axiomatically appropriate. Then take

$$t_{\mathsf{C}} := c, \qquad t_i := \downarrow_i c, \qquad t_{\mathsf{E}} := \mathsf{head}(c)$$

and use axiom necessitation, axiom necessitation and i-conversion, or axiom necessitation and the co-closure axiom, respectively.

For $A = [s_j]_{\mathsf{C}} B_j, 1 \leq j \leq n$, take

$$t_{\mathsf{C}} :=!_{\mathsf{C}} x_j, \qquad t_i :=\downarrow_i !_{\mathsf{C}} x_j, \qquad t_{\mathsf{E}} := \mathsf{tail}(x_j)$$

for a fresh variable $x_j \in \mathrm{Var}_{\mathsf{C}}$ and, after x_j is replaced with s_j, use C-inspection, C-inspection and i-conversion, or the co-closure axiom, respectively.

For $A = C_j, 1 \leq j \leq m$, take $t_{\circledast} := y_j$ for a fresh variable $y_j \in \mathrm{Var}_{\circledast}$.

For A derived by modus ponens from $D \to A$ and D, by induction hypothesis there are terms $r_{\circledast}, s_{\circledast} \in \mathrm{Tm}_{\circledast}$ such that $[r_{\circledast}]_{\circledast}(D \to A)$ and $[s_{\circledast}]_{\circledast} D$ are derivable. Take $t_{\circledast} := r_{\circledast} \cdot_{\circledast} s_{\circledast}$ and use $\circledast$-application, which is an axiom for $\circledast = i$ and for $\circledast = \mathsf{C}$ or follows from Lemma 6.1 for $\circledast = \mathsf{E}$.

For $A = [c]_{\mathsf{C}} E \in \mathcal{CS}$ derived by axiom necessitation, take

$$t_{\mathsf{C}} :=!_{\mathsf{C}} c, \qquad t_i :=\downarrow_i !_{\mathsf{C}} c, \qquad t_{\mathsf{E}} := \mathsf{tail}(c)$$

and, as before, use C-inspection, C-inspection and i-conversion, or the co-closure axiom respectively. No other instances of the axiom necessitation rule are possible. Indeed, $\mathcal{CS}$ must contain formulae of the type $[c]_{\mathsf{C}} E$ because of C-axiomatic appropriateness. The homogeneity of $\mathcal{CS}$ then means that formulae neither of type $[c]_i E$ nor of type $[c]_{\mathsf{E}} E$ can occur in $\mathcal{CS}$. □

Corollary 6.6 (Constructive necessitation). *Let $\mathcal{CS}$ be a homogeneous C-axiomatically appropriate constant specification. For any formula A, if $\vdash_{\mathcal{CS}} A$, then for each $\circledast$ there is a ground term $t \in \mathrm{Tm}_{\circledast}$ such that $\vdash_{\mathcal{CS}} [t]_{\circledast} A$.*

The following two lemmas show that our system $\mathsf{LP}_h^{\mathsf{C}}$ can internalize versions of the induction rule (I-R1) and (I-R2) used in axiomatizations of $\mathsf{S4}_h^{\mathsf{C}}$ in Chapter 3, Section 3.2. We name them accordingly and first give a proof of the analogue of the simpler (I-R2) which we then use to prove the analogue of the general (I-R1).

Lemma 6.7 (Internalized induction rule 2). *Let $\mathcal{CS}$ be a homogeneous C-axiomatically appropriate constant specification. For any term $s \in \mathrm{Tm}_{\mathsf{E}}$ and any formula A, if $\vdash_{\mathcal{CS}} A \to [s]_{\mathsf{E}} A$, there is $t \in \mathrm{Tm}_{\mathsf{C}}$ such that $\vdash_{\mathcal{CS}} A \to [\mathsf{ind}(t, s)]_{\mathsf{C}} A$.*

Proof. By constructive necessitation, $\vdash_{\mathcal{CS}} [t]_{\mathsf{C}}(A \to [s]_{\mathsf{E}}A)$ for some $t \in \mathrm{Tm}_{\mathsf{C}}$. We then can use the following instance of the induction axiom

$$[t]_{\mathsf{C}}(A \to [s]_{\mathsf{E}}A) \to ([s]_{\mathsf{E}}A \to [\mathsf{ind}(t,s)]_{\mathsf{C}}A)$$

to obtain $[s]_{\mathsf{E}}A \to [\mathsf{ind}(t,s)]_{\mathsf{C}}A$. Using propositional reasoning and $A \to [s]_{\mathsf{E}}A$ again, we get $A \to [\mathsf{ind}(t,s)]_{\mathsf{C}}A$. □

Lemma 6.8 (Internalized induction rule 1)**.** *Let $\mathcal{CS}$ be a homogeneous* C*-axiomatically appropriate constant specification. For any formulae A and B and any term $s \in \mathrm{Tm}_{\mathsf{E}}$, if we have $\vdash_{\mathcal{CS}} B \to [s]_{\mathsf{E}}(A \wedge B)$, then there exists $t \in \mathrm{Tm}_{\mathsf{C}}$ and $c \in \mathrm{Cons}_{\mathsf{C}}$ such that $\vdash_{\mathcal{CS}} B \to [c \cdot \mathsf{ind}(t,s)]_{\mathsf{C}}A$, where $[c]_{\mathsf{C}}(A \wedge B \to A) \in \mathcal{CS}$.*

Proof. Assume

$$\vdash_{\mathcal{CS}} B \to [s]_{\mathsf{E}}(A \wedge B) \ . \tag{6.3}$$

From this we immediately get $\vdash_{\mathcal{CS}} A \wedge B \to [s]_{\mathsf{E}}(A \wedge B)$. Thus, by Lemma 6.7, there is a $t \in \mathrm{Tm}_{\mathsf{C}}$ with

$$\vdash_{\mathcal{CS}} A \wedge B \to [\mathsf{ind}(t,s)]_{\mathsf{C}}(A \wedge B) \ . \tag{6.4}$$

Since $\mathcal{CS}$ is C-axiomatically appropriate, there is a constant $c \in \mathrm{Cons}_{\mathsf{C}}$ such that

$$\vdash_{\mathcal{CS}} [c]_{\mathsf{C}}(A \wedge B \to A) \ . \tag{6.5}$$

Making use of C-application, we find by (6.4) and (6.5) that

$$\vdash_{\mathcal{CS}} A \wedge B \to [c \cdot \mathsf{ind}(t,s)]_{\mathsf{C}}A \ . \tag{6.6}$$

From (6.3) we get by E-reflexivity that $\vdash_{\mathcal{CS}} B \to A \wedge B$. This, together with (6.6), finally yields $\vdash_{\mathcal{CS}} B \to [c \cdot \mathsf{ind}(t,s)]_{\mathsf{C}}A$. □

6.3. Soundness and Completeness

Definition 6.9. An *(epistemic) model meeting a constant specification $\mathcal{CS}$* is a structure $\mathcal{M} = (W, R, \mathcal{E}, \nu)$, where (W, R, ν) is a Kripke model for $\mathsf{S4}_h$ with a *set of possible worlds* $W \neq \varnothing$, with a function $R\colon \{1,\dots,h\} \to \mathcal{P}(W \times W)$ that assigns a reflexive and transitive *accessibility relation* on W to each agent $i \in \{1,\dots,h\}$, and with a *truth* *valuation* $\nu\colon \mathrm{Prop} \to \mathcal{P}(W)$. We always write R_i instead of $R(i)$ and define the accessibility relations for mutual and common knowledge in the standard way: $R_{\mathsf{E}} := R_1 \cup \ldots \cup R_h$ and $R_{\mathsf{C}} := \bigcup_{n=1}^{\infty}(R_{\mathsf{E}})^n$.

An *evidence function* $\mathcal{E}\colon W \times \mathrm{Tm} \to \mathcal{P}\left(\mathrm{Fm}_{\mathsf{LP}_h^{\mathsf{C}}}\right)$ determines the formulae evidenced by a term at a world. We define $\mathcal{E}_{\circledast} := \mathcal{E} \restriction (W \times \mathrm{Tm}_{\circledast})$. Note that whenever $A \in \mathcal{E}_{\circledast}(w,t)$, it follows that $t \in \mathrm{Tm}_{\circledast}$. The evidence function $\mathcal{E}$ must satisfy the following closure conditions: for any worlds $w, v \in W$,

1. $\mathcal{E}_*(w,t) \subseteq \mathcal{E}_*(v,t)$ whenever $(w,v) \in R_*$; (monotonicity)
2. if $[c]_{\circledast}A \in \mathcal{CS}$, then $A \in \mathcal{E}_{\circledast}(w,c)$; (constant specification)
3. if $(A \to B) \in \mathcal{E}_*(w,t)$ and $A \in \mathcal{E}_*(w,s)$, then $B \in \mathcal{E}_*(w, t\cdot s)$; (application)
4. $\mathcal{E}_*(w,s) \cup \mathcal{E}_*(w,t) \subseteq \mathcal{E}_*(w, s+t)$; (sum)
5. if $A \in \mathcal{E}_i(w,t)$, then $[t]_i A \in \mathcal{E}_i(w, !t)$; (inspection)
6. if $A \in \mathcal{E}_i(w,t_i)$ for all $1 \le i \le h$, then $A \in \mathcal{E}_{\mathsf{E}}(w, \langle t_1, \ldots, t_h\rangle)$; (tupling)
7. if $A \in \mathcal{E}_{\mathsf{E}}(w,t)$, then $A \in \mathcal{E}_i(w, \pi_i t)$; (projection)
8. if $A \in \mathcal{E}_{\mathsf{C}}(w,t)$, then $A \in \mathcal{E}_{\mathsf{E}}(w, \mathsf{head}(t))$ and $[t]_{\mathsf{C}}A \in \mathcal{E}_{\mathsf{E}}(w, \mathsf{tail}(t))$; (co-closure)
9. if $A \in \mathcal{E}_{\mathsf{E}}(w,s)$ and $(A \to [r]_{\mathsf{E}}A) \in \mathcal{E}_{\mathsf{C}}(w,t)$, then $A \in \mathcal{E}_{\mathsf{C}}(w, \mathsf{ind}(t,s))$. (induction)

When the model is clear from the context, we will directly refer to $R_1, \ldots, R_h$, R_{E}, R_{C}, $\mathcal{E}_1, \ldots, \mathcal{E}_h$, $\mathcal{E}_{\mathsf{E}}$, $\mathcal{E}_{\mathsf{C}}$, W, and ν.

Definition 6.10. A ternary relation $\mathcal{M}, w \Vdash A$ for *formula A being satisfied at a world $w \in W$ in a model $\mathcal{M} = (W, R, \mathcal{E}, \nu)$* is defined by induction on the structure of the formula A:

1. $\mathcal{M}, w \Vdash P_n$ if and only if $w \in \nu(P_n)$;
2. $\Vdash$ behaves classically with respect to the propositional connectives;

3. $\mathcal{M}, w \Vdash [t]_{\circledast} A$ if and only if
   1) $A \in \mathcal{E}_{\circledast}(w, t)$ and
   2) $\mathcal{M}, v \Vdash A$ for all $v \in W$ with $(w, v) \in R_{\circledast}$.

We write $\mathcal{M} \Vdash A$ if $\mathcal{M}, w \Vdash A$ for all $w \in W$. We write $\mathcal{M}, w \Vdash \Delta$ for $\Delta \subseteq \mathrm{Fm}_{\mathsf{LP}^{\mathsf{C}}_h}$ if $\mathcal{M}, w \Vdash A$ for all $A \in \Delta$. We write $\Vdash_{\mathcal{CS}} A$ and say that formula A is *valid with respect to* $\mathcal{CS}$ if $\mathcal{M} \Vdash A$ for all epistemic models $\mathcal{M}$ meeting $\mathcal{CS}$.

Lemma 6.11 (Soundness)**.** *All theorems are valid, i.e.,* $\vdash_{\mathcal{CS}} A$ *implies* $\Vdash_{\mathcal{CS}} A$*.*

Proof. Let $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ be a model meeting $\mathcal{CS}$ and let $w \in W$. We show soundness by induction on the derivation of A. The cases for propositional tautologies, for the application, sum, reflexivity, and inspection axioms, and for the modus ponens rule are the same as for the single-agent case in [Fit05] and are, therefore, omitted. We show the remaining five cases:

(tupling) Assume $\mathcal{M}, w \Vdash [t_i]_i A$ for all $1 \leq i \leq h$. Then for all $1 \leq i \leq h$, we have 1) $\mathcal{M}, v \Vdash A$ whenever $(w, v) \in R_i$ and 2) $A \in \mathcal{E}_i(w, t_i)$. By the tupling closure condition, it follows from 2) that $A \in \mathcal{E}_{\mathsf{E}}(w, \langle t_1, \ldots, t_h \rangle)$. Since $R_{\mathsf{E}} = \bigcup_{i=1}^{h} R_i$ by definition, it follows from 1) that $\mathcal{M}, v \Vdash A$ whenever $(w, v) \in R_{\mathsf{E}}$. Hence, $\mathcal{M}, w \Vdash [\langle t_1, \ldots, t_h \rangle]_{\mathsf{E}} A$.

(projection) Assume $\mathcal{M}, w \Vdash [t]_{\mathsf{E}} A$. Then 1) $\mathcal{M}, v \Vdash A$ whenever $(w, v) \in R_{\mathsf{E}}$ and 2) $A \in \mathcal{E}_{\mathsf{E}}(w, t)$. By the projection closure condition, it follows from 2) that $A \in \mathcal{E}_i(w, \pi_i t)$. In addition, since $R_{\mathsf{E}} = \bigcup_{i=1}^{h} R_i$, it follows from 1) that $\mathcal{M}, v \Vdash A$ whenever $(w, v) \in R_i$. Thus, $\mathcal{M}, w \Vdash [\pi_i t]_i A$.

(co-closure) Assume $\mathcal{M}, w \Vdash [t]_{\mathsf{C}} A$. Then 1) $\mathcal{M}, v \Vdash A$ whenever $(w, v) \in R_{\mathsf{C}}$ and 2) $A \in \mathcal{E}_{\mathsf{C}}(w, t)$. It follows from 1) that $\mathcal{M}, v' \Vdash A$ whenever $(w, v') \in R_{\mathsf{E}}$ since $R_{\mathsf{E}} \subseteq R_{\mathsf{C}}$; also, due to the monotonicity closure condition, $\mathcal{M}, v' \Vdash [t]_{\mathsf{C}} A$ since $R_{\mathsf{E}} \circ R_{\mathsf{C}} \subseteq R_{\mathsf{C}}$. By the co-closure closure condition, it follows from 2) that $A \in \mathcal{E}_{\mathsf{E}}(w, \mathsf{head}(t))$ and $[t]_{\mathsf{C}} A \in \mathcal{E}_{\mathsf{E}}(w, \mathsf{tail}(t))$. Hence, $\mathcal{M}, w \Vdash [\mathsf{head}(t)]_{\mathsf{E}} A$ and $\mathcal{M}, w \Vdash [\mathsf{tail}(t)]_{\mathsf{E}} [t]_{\mathsf{C}} A$.

(induction) Assume $\mathcal{M}, w \Vdash [t]_{\mathsf{C}}(A \to [r]_{\mathsf{E}}A)$. and $\mathcal{M}, w \Vdash [s]_{\mathsf{E}}A$. So $A \in \mathcal{E}_{\mathsf{E}}(w, s)$ and $A \to [s]_{\mathsf{E}}A \in \mathcal{E}_{\mathsf{C}}(w, t)$. By the induction closure condition, we have $A \in \mathcal{E}_{\mathsf{C}}(w, \mathsf{ind}(t, s))$. To show that $\mathcal{M}, v \Vdash A$ whenever $(w, v) \in R_{\mathsf{C}}$, we prove that $\mathcal{M}, v \Vdash A$ whenever $(w, v) \in (R_{\mathsf{E}})^n$ by induction on the positive integer n.

The **base case** $n = 1$ immediately follows from $\mathcal{M}, w \Vdash [s]_{\mathsf{E}}A$.

Induction step. If $(w, v) \in (R_{\mathsf{E}})^{n+1}$, there must exist $v' \in W$ such that $(w, v') \in (R_{\mathsf{E}})^n$ and $(v', v) \in R_{\mathsf{E}}$. By induction hypothesis, $\mathcal{M}, v' \Vdash A$. Since $\mathcal{M}, w \Vdash [t]_{\mathsf{C}}(A \to [r]_{\mathsf{E}}A)$, we get $\mathcal{M}, v' \Vdash A \to [r]_{\mathsf{E}}A$. Thus, $\mathcal{M}, v' \Vdash [r]_{\mathsf{E}}A$, which yields $\mathcal{M}, v \Vdash A$.

Finally, we conclude that $\mathcal{M}, w \Vdash [\mathsf{ind}(t, s)]_{\mathsf{C}}A$.

(axiom necessitation) Let $[c]_{\circledast}A \in \mathcal{CS}$. Since A must be an axiom, $\mathcal{M}, w \Vdash A$ for all $w \in W$, as shown above. Since $\mathcal{M}$ is a model meeting $\mathcal{CS}$, we also have $A \in \mathcal{E}_{\circledast}(w, c)$ for all $w \in W$ by the constant specification closure condition. Thus, $\mathcal{M}, w \Vdash [c]_{\circledast}A$ for all $w \in W$. □

Definition 6.12. Let $\mathcal{CS}$ be a constant specification. A set Φ of formulae is called *$\mathcal{CS}$-consistent* if $\Phi \nvdash_{\mathcal{CS}} \phi$ for some formula ϕ. A set Φ is called *maximal $\mathcal{CS}$-consistent* if it is $\mathcal{CS}$-consistent and has no $\mathcal{CS}$-consistent proper extensions.

Whenever safe, we do not mention the constant specification and only talk about consistent and maximal consistent sets. It can be easily shown that maximal consistent sets contain all axioms of $\mathsf{LP}_h^{\mathsf{C}}$ and are closed under modus ponens.

Definition 6.13. For a set Φ of formulae, we define

$$\Phi/\circledast := \{A \ : \ \text{there is a } t \in \mathrm{Tm}_{\circledast} \text{ such that } [t]_{\circledast}A \in \Phi\} \ .$$

Definition 6.14. Let $\mathcal{CS}$ be a constant specification. The *canonical (epistemic) model $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ meeting $\mathcal{CS}$* is defined as follows:

1. $W := \{w \subseteq \mathrm{Fm}_{\mathsf{LP}_h^{\mathsf{C}}} \ : \ w \text{ is a maximal } \mathcal{CS}\text{-consistent set}\}$;
2. $R_i := \{(w, v) \in W \times W \ : \ w/i \subseteq v\}$;

3. $\mathcal{E}_{\circledast}(w,t) := \{A \in \mathrm{Fm}_{\mathsf{LP}^{\mathsf{C}}_h} \ : \ [t]_{\circledast} A \in w\}$;

4. $\nu(P_n) := \{w \in W \ : \ P_n \in w\}$.

Lemma 6.15. *Let $\mathcal{CS}$ be a constant specification. The canonical epistemic model meeting $\mathcal{CS}$ is an epistemic model meeting $\mathcal{CS}$.*

Proof. The proof of the reflexivity and transitivity of each R_i, as well as the argument for the constant specification, application, sum, and inspection closure conditions, is the same as in the single-agent case (see [Fit05]). We show the remaining five closure conditions:

(tupling) Assume $A \in \mathcal{E}_i(w, t_i)$ for all $1 \le i \le h$. By definition of $\mathcal{E}_i$, we have $[t_i]_i A \in w$ for all $1 \le i \le h$. Therefore, by the tupling axiom and maximal consistency, $[\langle t_1, \ldots, t_h \rangle]_{\mathsf{E}} A \in w$. Thus, $A \in \mathcal{E}_{\mathsf{E}}(w, \langle t_1, \ldots, t_h \rangle)$.

(projection) Assume $A \in \mathcal{E}_{\mathsf{E}}(w,t)$. By definition of $\mathcal{E}_{\mathsf{E}}$, we have $[t]_{\mathsf{E}} A \in w$. Therefore, by the projection axiom and maximal consistency, $[\pi_i t]_i A \in w$. Thus, $A \in \mathcal{E}_i(w, \pi_i t)$.

(co-closure) Assume $A \in \mathcal{E}_{\mathsf{C}}(w,t)$. By definition of $\mathcal{E}_{\mathsf{C}}$, we have $[t]_{\mathsf{C}} A \in w$. Therefore, by the co-closure axioms and maximal consistency, $[\mathsf{head}(t)]_{\mathsf{E}} A \in w$ and $[\mathsf{tail}(t)]_{\mathsf{E}}\, [t]_{\mathsf{C}} A \in w$. Thus, $A \in \mathcal{E}_{\mathsf{E}}(w, \mathsf{head}(t))$ and $[t]_{\mathsf{C}} A \in \mathcal{E}_{\mathsf{E}}(w, \mathsf{tail}(t))$.

(induction) Assume $A \in \mathcal{E}_{\mathsf{E}}(w,s)$ and $(A \to [r]_{\mathsf{E}} A) \in \mathcal{E}_{\mathsf{C}}(w,t)$. By definition of $\mathcal{E}_{\mathsf{E}}$ and $\mathcal{E}_{\mathsf{C}}$, we have $[s]_{\mathsf{E}} A \in w$ and $[t]_{\mathsf{C}}(A \to [r]_{\mathsf{E}} A) \in w$. By the induction axiom, it follows by maximal consistency that $A \in \mathcal{E}_{\mathsf{C}}(w, \mathsf{ind}(t,s))$.

(monotonicity) We show only the case of $* = \mathsf{C}$ since the other cases are the same as in [Fit05]. It is sufficient to prove by induction on the positive integer n that

$$\text{if } [t]_{\mathsf{C}} A \in w \text{ and } (w,v) \in (R_{\mathsf{E}})^n, \text{ then } [t]_C A \in v \ . \tag{6.7}$$

Base case $n = 1$**.** Assume $(w,v) \in R_{\mathsf{E}}$: i.e., $w/i \subseteq v$ for some i. As $[t]_{\mathsf{C}} A \in w$, $[\pi_i \mathsf{tail}(t)]_i\, [t]_{\mathsf{C}} A \in w$ by maximal consistency, and hence $[t]_{\mathsf{C}} A \in w/i \subseteq v$. The argument for the **induction step** is similar.

Now assume $(w, v) \in R_{\mathsf{C}} = \bigcup_{n=1}^{\infty}(R_{\mathsf{E}})^n$ and $A \in \mathcal{E}_{\mathsf{C}}(w,t)$. By definition of $\mathcal{E}_{\mathsf{C}}$, we have $[t]_{\mathsf{C}}A \in w$. As shown above, $[t]_{\mathsf{C}}A \in v$. Thus, $A \in \mathcal{E}_{\mathsf{C}}(v,t)$. □

Remark 6.16. Let R'_{C} denote the binary relation on W defined by

$$(w, v) \in R'_{\mathsf{C}} \quad \text{if and only if} \quad w/\mathsf{C} \subseteq v \ .$$

An argument similar to the one just used for monotonicity shows that $R_{\mathsf{C}} \subseteq R'_{\mathsf{C}}$. However, for $h > 1$ the converse does not hold for any homogeneous C-axiomatically appropriate constant specification $\mathcal{CS}$, which we demonstrate by adapting an example from [MH95]. For a fixed propositional variable P, let

$$\Phi := \{[s_n]_{\mathsf{E}} \dots [s_1]_{\mathsf{E}}P \ : \ n \geq 1,\, s_1, \dots, s_n \in \mathrm{Tm}_{\mathsf{E}}\} \\ \cup \{\neg\,[t]_{\mathsf{C}}P \ : \ t \in \mathrm{Tm}_{\mathsf{C}}\} \ .$$

This set is $\mathcal{CS}$-consistent for any $P \in \mathrm{Prop}$.

To prove this, let $\Phi' \subseteq \Phi$ be finite and let m denote the largest nonnegative integer such that $[s_m]_{\mathsf{E}} \dots [s_1]_{\mathsf{E}}P \in \Phi'$ for some $s_1, \dots, s_m \in \mathrm{Tm}_{\mathsf{E}}$ (in particular, $m = 0$ if no such terms exist). Define the model $\mathcal{N} := \big(\mathbb{N}, R^{\mathcal{N}}, \mathcal{E}^{\mathcal{N}}, \nu^{\mathcal{N}}\big)$ by

- $R_i^{\mathcal{N}} := \{(n, n+1) \in \mathbb{N}^2 \ : \ n \mod h = i\} \cup \{(n,n) \ : \ n \in \mathbb{N}\}$;
- $\mathcal{E}^{\mathcal{N}}(n, s) := \mathrm{Fm}_{\mathsf{LP}_h^{\mathsf{C}}}$ for all $n \in \mathbb{N}$ and all terms $s \in \mathrm{Tm}$;
- $\nu^{\mathcal{N}}(P_j) := \{1, 2, \dots, m+1\}$ for all $P_j \in \mathrm{Prop}$.

Clearly, $\mathcal{N}$ meets any constant specification; in particular, it meets the given $\mathcal{CS}$. For $h > 1$, it can also be easily verified that $\mathcal{N}, 1 \Vdash \Phi'$; therefore, Φ' is $\mathcal{CS}$-consistent.

Since Φ is $\mathcal{CS}$-consistent, there exists a maximal $\mathcal{CS}$-consistent set $w \supseteq \Phi$. Let us show that the set $\Psi := \{\neg P\} \cup (w/\mathsf{C})$ is also $\mathcal{CS}$-consistent. Indeed, if it were not the case, there would exist formulae $[t_1]_{\mathsf{C}}B_1, \dots, [t_n]_{\mathsf{C}}B_n \in w$ such that

$$\vdash_{\mathcal{CS}} B_1 \to (B_2 \to \dots \to (B_n \to P) \dots) \ .$$

Then, by Corollary 6.6, there would exist a term $s \in \mathrm{Tm}_{\mathsf{C}}$ such that

$$\vdash_{\mathcal{CS}} [s]_{\mathsf{C}}(B_1 \to (B_2 \to \dots \to (B_n \to P) \dots)) \ .$$

But this would imply $[(\ldots(s \cdot t_1) \cdots t_{n-1}) \cdot t_n]_{\mathsf{C}} P \in w$—a contradiction with the consistency of w.

Since Ψ is also $\mathcal{CS}$-consistent, there exists a maximal $\mathcal{CS}$-consistent set $v \supseteq \Psi$. Clearly, $w/\mathsf{C} \subseteq v$: i.e., $(w, v) \in R'_{\mathsf{C}}$. But $(w, v) \notin R_{\mathsf{C}}$ because this would imply $P \in v$, which would contradict the consistency of v. It follows that $R_{\mathsf{C}} \subsetneq R'_{\mathsf{C}}$. In accordance with the similiarities to the situation for the iteration operator in dynamic logic [HKJ00], we will refer to this as the *non-standard behavior* of the canonical model.

Similarly, we can define R'_{E} by $(w, v) \in R'_{\mathsf{E}}$ if and only if $w/\mathsf{E} \subseteq v$. However, $R'_{\mathsf{E}} = R_{\mathsf{E}}$ for any C-axiomatically appropriate constant specification $\mathcal{CS}$. Indeed, it is easy to show that $R_{\mathsf{E}} \subseteq R'_{\mathsf{E}}$. For the converse direction, assume $(w, v) \notin R_{\mathsf{E}}$, then $(w, v) \notin R_i$ for any $1 \leq i \leq h$. So there are formulae $A_1, \ldots, A_h$ such that $[t_i]_i A_i \in w$ for some $t_i \in \mathrm{Tm}_i$, but $A_i \notin v$. Now let $[c_i]_{\mathsf{C}}(A_i \to A_1 \vee \ldots \vee A_h) \in \mathcal{CS}$ for constants $c_1, \ldots, c_h$. Then $[\downarrow_i c_i \cdot t_i]_i (A_1 \vee \ldots \vee A_h) \in w$ for all $1 \leq i \leq h$, so $[\langle \downarrow_1 c_1 \cdot t_1, \ldots, \downarrow_h c_h \cdot t_h \rangle]_{\mathsf{E}}(A_1 \vee \ldots \vee A_h) \in w$. However, $A_i \notin v$ for any $1 \leq i \leq h$; therefore, by the maximal consistency of v, $A_1 \vee \ldots \vee A_h \notin v$ either. Hence, $w/\mathsf{E} \not\subseteq v$, so $(w, v) \notin R'_{\mathsf{E}}$.

Lemma 6.17 (Truth Lemma)**.** *Let $\mathcal{CS}$ be a constant specification and $\mathcal{M}$ be the canonical epistemic model meeting $\mathcal{CS}$. For all formulae A and all worlds $w \in W$,*

$$A \in w \textit{ if and only if } \mathcal{M}, w \Vdash A \ .$$

Proof. The proof is by induction on the structure of A. The cases for propositional variables and propositional connectives are immediate by definition of $\Vdash$ and by the maximal consistency of w. We check the remaining cases:

Case A is $[t]_i B$. Assume $A \in w$. Then $B \in w/i$ and $B \in \mathcal{E}_i(w, t)$. Consider any v such that $(w, v) \in R_i$. Since $w/i \subseteq v$, it follows that $B \in v$, and thus, by induction hypothesis, $\mathcal{M}, v \Vdash B$. It immediately follows that $\mathcal{M}, w \Vdash A$.

For the converse, assume $\mathcal{M}, w \Vdash [t]_i B$. By definition of $\Vdash$, we get $B \in \mathcal{E}_i(w, t)$, from which $[t]_i B \in w$ immediately follows by definition of $\mathcal{E}_i$.

Case A is $[t]_{\mathsf{E}} B$. Assume $A \in w$ and consider any v such that $(w, v) \in R_{\mathsf{E}}$. Then $(w, v) \in R_i$ for some $1 \leq i \leq h$: i.e., $w/i \subseteq v$. By definition of $\mathcal{E}_{\mathsf{E}}$, we have $B \in \mathcal{E}_{\mathsf{E}}(w, t)$. By the maximal consistency of w, it follows

that $[\pi_i t]_i B \in w$, and thus $B \in w/i \subseteq v$. Since by induction hypothesis, $\mathcal{M}, v \Vdash B$, we can conclude that $\mathcal{M}, w \Vdash A$. The argument for the converse repeats the one from the previous case.
Case A is $[t]_{\mathsf{C}} B$. Assume $A \in w$ and consider any v such that $(w, v) \in R_{\mathsf{C}}$: i.e., $(w, v) \in (R_{\mathsf{E}})^n$ for some $n \geq 1$. As in the previous cases, $B \in \mathcal{E}_{\mathsf{C}}(w, t)$ by definition of $\mathcal{E}_{\mathsf{C}}$. It follows from (6.7) in the proof of Lemma 6.15 that $A \in v$, and thus, by C-reflexivity and maximal consistency, also $B \in v$. Hence, by induction hypothesis, $\mathcal{M}, v \Vdash B$. Now $\mathcal{M}, w \Vdash A$ immediately follows. The argument for the converse repeats the one from the previous cases. ☐

Note that, unlike the converse directions in the proof above, the corresponding proofs in the modal case are far from trivial and require additional work (see e.g. [MH95]). The last case, in particular, usually requires more sophisticated methods that would guarantee the finiteness of the model. This simplification of proofs in justification logics is yet another benefit of using terms instead of modalities.

Theorem 6.18 (Completeness). $\mathsf{LP}^{\mathsf{C}}_h(\mathcal{CS})$ *is sound and complete with respect to the class of epistemic models meeting* $\mathcal{CS}$*: i.e., for all formulae* $A \in \mathrm{Fm}_{\mathsf{LP}^{\mathsf{C}}_h}$,

$$\vdash_{\mathcal{CS}} A \textit{ if and only if } \Vdash_{\mathcal{CS}} A \ .$$

Proof. Soundness was already shown in Lemma 6.11. For completeness, let $\mathcal{M}$ be the canonical model meeting $\mathcal{CS}$ and assume $\nvdash_{\mathcal{CS}} A$. Then $\{\neg A\}$ is $\mathcal{CS}$-consistent and hence is contained in some maximal $\mathcal{CS}$-consistent set $w \in W$. So, by Lemma 6.17, $\mathcal{M}, w \Vdash \neg A$, and hence, by Lemma 6.15, $\nVdash_{\mathcal{CS}} A$. ☐

6.4. Finite Model Property and Decidability

In the case of LP, the finite model property can be demonstrated by restricting the class of epistemic models to the so-called M-models, introduced by Mkrtychev in [Mkr97]. We will now adapt M-models to our logic and prove the finite model property for it.

Definition 6.19. An *M-model* is a singleton epistemic model.

Theorem 6.20 (Completeness w.r.t. M-models). $\mathsf{LP}^{\mathsf{C}}_h(\mathcal{CS})$ *is also sound and complete with respect to the class of M-models meeting* $\mathcal{CS}$*.*

Proof. Soundness follows immediately from Lemma 6.11. Now assume $\nvdash_{\mathcal{CS}} A$, then $\{\neg A\}$ is $\mathcal{CS}$-consistent, and hence $\mathcal{M}, w_0 \Vdash \neg A$ for some world $w_0 \in W$ in the canonical epistemic model $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ meeting $\mathcal{CS}$.

Let $\mathcal{M}' = (W', R', \mathcal{E}', \nu')$ be the restriction of $\mathcal{M}$ to $\{w_0\}$: i.e.,

- $W' := \{w_0\}$,
- $R'_i := \{(w_0, w_0)\}$ for all i,
- $\mathcal{E}' := \mathcal{E} \upharpoonright (W' \times \mathrm{Tm})$,
- $\nu'(P_n) := \nu(P_n) \cap W'$.

Since $\mathcal{M}'$ is clearly an M-model meeting $\mathcal{CS}$, it only remains to demonstrate that $\mathcal{M}', w_0 \Vdash B$ if and only if $\mathcal{M}, w_0 \Vdash B$ for all formulae B. We proceed by induction on the structure of B. The cases where either B is a propositional variable or its primary connective is propositional are trivial. Therefore, we only show the case of $B = [t]_{\circledast} C$. First, observe that

$$\mathcal{M}, w_0 \Vdash [t]_{\circledast} C \text{ if and only if } C \in \mathcal{E}'_{\circledast}(w_0, t) \ . \tag{6.8}$$

Indeed, by Lemma 6.17, $\mathcal{M}, w_0 \Vdash [t]_{\circledast} C$ if and only if $[t]_{\circledast} C \in w_0$, which, by definition of the canonical epistemic model, is equivalent to $C \in \mathcal{E}_{\circledast}(w_0, t) = \mathcal{E}'_{\circledast}(w_0, t)$.

If $\mathcal{M}, w_0 \Vdash [t]_{\circledast} C$, then $\mathcal{M}, w_0 \Vdash C$ since $R_{\circledast}$ is reflexive. By induction hypothesis, $\mathcal{M}', w_0 \Vdash C$. By (6.8) we have $C \in \mathcal{E}'_{\circledast}(w_0, t)$, and thus $\mathcal{M}', w_0 \Vdash [t]_{\circledast} C$.

If $\mathcal{M}, w_0 \nVdash [t]_{\circledast} C$, then by (6.8) we have $C \notin \mathcal{E}'_{\circledast}(w_0, t)$, so $\mathcal{M}', w_0 \nVdash [t]_{\circledast} C$. □

Corollary 6.21 (Finite model property). $\mathsf{LP}^{\mathsf{C}}_h(\mathcal{CS})$ *enjoys the finite model property with respect to epistemic models.*

Using the techniques from Chapter A, we then get decidability as stated in Theorem A.30

Theorem 6.22. $\mathsf{LP}^{\mathsf{C}}_h(\mathcal{CS})$ *with a decidable schematic* $\mathcal{CS}$ *is decidable.*

Remark 6.23. Note however that, in the case of $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS})$, the finite model property does not imply that common knowledge can be deduced from sufficiently many approximants, unlike in the modal case. This is an immediate consequence of the set

$$\Phi := \{[s_n]_{\mathsf{E}} \ldots [s_1]_{\mathsf{E}} P \; : \; n \geq 1,\, s_1, \ldots, s_n \in \mathrm{Tm}_{\mathsf{E}}\} \\ \cup \{\neg\, [t]_{\mathsf{C}} P \; : \; t \in \mathrm{Tm}_{\mathsf{C}}\}$$

being consistent, as shown in Remark 6.16. In modal logic, a set analogous to Φ can only be satisfied in infinite models, whereas in our case, due to the evidence function completely taking over the role of the accessibility relations, there is a singleton M-model that satisfies Φ.

6.5. Conservativity

We extend the two-agent version LP^2 of the Logic of Proofs [Yav08] to an arbitrary h in the natural way and rename it in accordance with our naming scheme:

Definition 6.24. The language of LP_h is obtained from that of $\mathsf{LP}_h^{\mathsf{C}}$ by restricting the set of operations to $\cdot_i$, $+_i$, and $!_i$ and by dropping all terms from Tm_{E} and Tm_{C}. The axioms are restricted to application, sum, reflexivity, and inspection for each i. The definition of constant specification is changed accordingly.

We show that $\mathsf{LP}_h^{\mathsf{C}}$ is conservative over LP_h by adapting the technique from [Fit08b], for which evidence terms are essential.

Definition 6.25. The mapping $.^{\times} : \mathrm{Fm}_{\mathsf{LP}_h^{\mathsf{C}}} \to \mathrm{Fm}_{\mathsf{LP}_h}$ is defined as follows:

1. $P_n^{\times} := P_n$ for propositional variables $P_n \in \mathrm{Prop}$;

2. $.^{\times}$ commutes with propositional connectives;

3. $$([t]_{\circledast} A)^{\times} := \begin{cases} A^{\times} & \text{if } t \text{ contains a subterm} \\ & \quad s \in \mathrm{Tm}_{\mathsf{E}} \cup \mathrm{Tm}_{\mathsf{C}}, \\ [t]_{\circledast} A^{\times} & \text{otherwise.} \end{cases}$$

Theorem 6.26. *Let $\mathcal{CS}$ be a constant specification for $\mathsf{LP}_h^{\mathsf{C}}$. For an arbitrary formula $A \in \mathrm{Fm}_{\mathsf{LP}_h}$,*

$$\text{if} \quad \mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash A, \qquad \text{then} \quad \mathsf{LP}_h(\mathcal{CS}^\times) \vdash A \ ,$$

where $\mathcal{CS}^\times := \{[c]_i E^\times \ : \ [c]_i E \in \mathcal{CS}\}$.

Proof. Since $A^\times = A$ for any $A \in \mathrm{Fm}_{\mathsf{LP}_h}$, it suffices to demonstrate that for any formula $D \in \mathrm{Fm}_{\mathsf{LP}_h^{\mathsf{C}}}$, if $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash D$, then $\mathsf{LP}_h(\mathcal{CS}^\times) \vdash D^\times$, which can be done by induction on the derivation of D.
Case when D is a propositional tautology. Then so is $D^\times$.
Case when $D = [t]_i B \to B$ is an instance of the reflexivity axiom. Then $D^\times$ is either the propositional tautology $B^\times \to B^\times$ or $[t]_i B^\times \to B^\times$, an instance of the reflexivity axiom of LP_h.
Case when $D = [t]_i B \to [!t]_i \, [t]_i B$ is an instance of the inspection axiom. Then $D^\times$ is either the propositional tautology $B^\times \to B^\times$ or $[t]_i B^\times \to [!t]_i \, [t]_i B^\times$, an instance of the inspection axiom of LP_h.
Case when $D = [t]_*(B \to C) \to ([s]_* B \to [t \cdot s]_* C)$ is an instance of the application axiom. We distinguish the following possibilities:

1. Both t and s contain a subterm from $\mathrm{Tm}_{\mathsf{E}} \cup \mathrm{Tm}_{\mathsf{C}}$. In this subcase, $D^\times$ has the form $(B^\times \to C^\times) \to (B^\times \to C^\times)$, which is a propositional tautology and, thus, an axiom of LP_h.

2. Neither t nor s contains a subterm from $\mathrm{Tm}_{\mathsf{E}} \cup \mathrm{Tm}_{\mathsf{C}}$. Then $D^\times$ is an instance of the application axiom of LP_h.

3. Term t contains a subterm from $\mathrm{Tm}_{\mathsf{E}} \cup \mathrm{Tm}_{\mathsf{C}}$ while s does not. Then $D^\times$ has the form $(B^\times \to C^\times) \to ([s]_i B^\times \to C^\times)$, which can be derived in $\mathsf{LP}_h(\mathcal{CS}^\times)$ from the reflexivity axiom $[s]_i B^\times \to B^\times$ by propositional reasoning. In this subcase, translation $\times$ does not map an axiom of $\mathsf{LP}_h^{\mathsf{C}}$ to an axiom of LP_h.

4. Term s contains a subterm from $\mathrm{Tm}_{\mathsf{E}} \cup \mathrm{Tm}_{\mathsf{C}}$ while t does not. Then $D^\times$ is $[t]_i(B^\times \to C^\times) \to (B^\times \to C^\times)$, an instance of the reflexivity axiom of LP_h.

Case when $D = [t]_* B \vee [s]_* B \to [t + s]_* B$ is an instance of the sum axiom. We distinguish the following possibilities:

1. Both t and s contain a subterm from $\mathrm{Tm}_{\mathsf{E}} \cup \mathrm{Tm}_{\mathsf{C}}$. In this subcase, $D^{\times}$ has the form $B^{\times} \vee B^{\times} \to B^{\times}$, which is a propositional tautology and, thus, an axiom of LP_h.

2. Neither t nor s contains a subterm from $\mathrm{Tm}_{\mathsf{E}} \cup \mathrm{Tm}_{\mathsf{C}}$. Then $D^{\times}$ is an instance of the sum axiom of LP_h.

3. Term t contains a subterm from $\mathrm{Tm}_{\mathsf{E}} \cup \mathrm{Tm}_{\mathsf{C}}$ while s does not. Then $D^{\times}$ has the form $B^{\times} \vee [s]_i B^{\times} \to B^{\times}$, which can be derived in $\mathsf{LP}_h(\mathcal{CS}^{\times})$ from the reflexivity axiom $[s]_i B^{\times} \to B^{\times}$ by propositional reasoning. This is another subcase when translation $\times$ does not map an axiom of $\mathsf{LP}_h^{\mathsf{C}}$ to an axiom of LP_h.

4. Term s contains a subterm from $\mathrm{Tm}_{\mathsf{E}} \cup \mathrm{Tm}_{\mathsf{C}}$ while t does not. Then $D^{\times}$ has the form $[t]_i B^{\times} \vee B^{\times} \to B^{\times}$, which can be derived in $\mathsf{LP}_h(\mathcal{CS}^{\times})$ from the reflexivity axiom $[t]_i B^{\times} \to B^{\times}$ by propositional reasoning. This is another subcase when translation $\times$ does not map an axiom of $\mathsf{LP}_h^{\mathsf{C}}$ to an axiom of LP_h.

Case when $D = [t_1]_1 B \wedge \ldots \wedge [t_h]_h B \to [\langle t_1, \ldots, t_h \rangle]_{\mathsf{E}} B$ is an instance of the tupling axiom. We distinguish the following possibilities:

1. At least one of the t_i's contains a subterm from $\mathrm{Tm}_{\mathsf{E}} \cup \mathrm{Tm}_{\mathsf{C}}$. Then $D^{\times}$ has the form $C_1 \wedge \ldots \wedge C_h \to B^{\times}$ with at least one $C_i = B^{\times}$ and is, therefore, a propositional tautology.

2. None of the t_i's contains a subterm from $\mathrm{Tm}_{\mathsf{E}} \cup \mathrm{Tm}_{\mathsf{C}}$. Then $D^{\times}$ has the form $[t_1]_1 B^{\times} \wedge \ldots \wedge [t_h]_h B^{\times} \to B^{\times}$, which can be derived in $\mathsf{LP}_h(\mathcal{CS}^{\times})$ from the reflexivity axiom. This is another subcase when translation $\times$ does not map an axiom of $\mathsf{LP}_h^{\mathsf{C}}$ to an axiom of LP_h.

Case when D is an instance of the projection axiom $[t]_{\mathsf{E}} B \to [\pi_i t]_i B$ or of the co-closure axiom: i.e., $[t]_{\mathsf{C}} B \to [\mathsf{head}(t)]_{\mathsf{E}} B$ or $[t]_{\mathsf{C}} B \to [\mathsf{tail}(t)]_{\mathsf{E}}\, [t]_{\mathsf{C}} B$. Then $D^{\times}$ is the propositional tautology $B^{\times} \to B^{\times}$.
Case when $D = [t]_{\mathsf{C}}(B \to [r]_{\mathsf{E}} B) \to ([s]_{\mathsf{E}} B \to [\mathsf{ind}(t,s)]_{\mathsf{C}} B)$ is an instance of the induction axiom. Then $D^{\times}$ is the propositional tautology $(B^{\times} \to B^{\times}) \to (B^{\times} \to B^{\times})$.
Case when D is derived by modus ponens is trivial.

Case when D is $[c]_{\circledast}B \in \mathcal{CS}$. Then $D^{\times}$ is either $B^{\times}$ or $[c]_i B^{\times}$. In the former case, $B^{\times}$ is derivable in $\mathsf{LP}_h(\mathcal{CS}^{\times})$, as shown above, because B is an axiom of $\mathsf{LP}_h^{\mathsf{C}}$; in the latter case, $[c]_i B^{\times} \in \mathcal{CS}^{\times}$. □

Remark 6.27. Note that $\mathcal{CS}^{\times}$ need not, in general, be a constant specification for LP_h because, as noted above, for an axiom D of $\mathsf{LP}_h^{\mathsf{C}}$, its image $D^{\times}$ is not always an axiom of LP_h. To ensure that $\mathcal{CS}^{\times}$ is a proper constant specification, all formulae of the forms

$$(A \to B) \to ([s]_i A \to B)\ , \qquad A \vee [s]_i A \to A\ ,$$
$$[t_1]_1 A \wedge \ldots \wedge [t_h]_h A \to A\ , \qquad [t]_i A \vee A \to A$$

have to be made axioms of LP_h. Another option is to use Fitting's concept of *embedding* one justification logic into another, which involves replacing constants in D with more complicated terms in $D^{\times}$ (see [Fit08b] for details).

6.6. Forgetful Projection and a Word on Realization

Most justification logics are introduced as explicit counterparts to particular modal logics in the strict sense described in Chapter. 5. Although the realization theorem for $\mathsf{LP}_h^{\mathsf{C}}$ remains an open problem, in this section we prove that each theorem of our logic $\mathsf{LP}_h^{\mathsf{C}}$ states a valid modal fact if all the terms are replaced with the corresponding modalities, which is one direction of the realization theorem. We also discuss approaches to the more difficult opposite direction.

Let us denote the set of all formulae in the language of $\mathsf{S4}_h^{\mathsf{C}}$ by $\mathrm{Fm}_{\mathsf{S4}_h^{\mathsf{C}}}$.

Definition 6.28 (Forgetful projection)**.** The mapping $\circ\colon \mathrm{Fm}_{\mathsf{LP}_h^{\mathsf{C}}} \to \mathrm{Fm}_{\mathsf{S4}_h^{\mathsf{C}}}$ is defined as follows:

1. $P_j^{\circ} := P_j$ for propositional variables $P_j \in \mathrm{Prop}$;
2. $\circ$ commutes with propositional connectives;
3. $([t]_i A)^{\circ} := \Box_i A^{\circ}$;
4. $([t]_{\mathsf{E}} A)^{\circ} := \mathsf{E} A^{\circ}$;

5. $([t]_{\mathsf{C}}A)^\circ := \mathsf{C}A^\circ$.

Lemma 6.29. *Let $\mathcal{CS}$ be a constant specification. For any formula $A \in \mathrm{Fm}_{\mathsf{LP}^{\mathsf{C}}_h}$, if $\mathsf{LP}^{\mathsf{C}}_h(\mathcal{CS}) \vdash A$, then $\mathsf{S4}^{\mathsf{C}}_h \vdash A^\circ$.*

Proof. The proof is by an easy induction on the derivation of A. □

Let us briefly recall the definition of a realization.

Definition 6.30 (Realization)**.** A realization is a mapping

$$r \colon \mathrm{Fm}_{\mathsf{S4}^{\mathsf{C}}_h} \to \mathrm{Fm}_{\mathsf{LP}^{\mathsf{C}}_h}$$

such that $(r(A))^\circ = A$. We usually write A^r instead of $r(A)$.

We can think of a realization as a function that replaces occurrences of modal operators (including E and C) with evidence terms of the corresponding type. The problem of realization for a given homogeneous C-axiomatically appropriate constant specification $\mathcal{CS}$ can be formulated as follows:

> Is there a realization r such that $\mathsf{LP}^{\mathsf{C}}_h(\mathcal{CS}) \vdash A^r$ for any given theorem A of $\mathsf{S4}^{\mathsf{C}}_h$?

A positive answer to this question would constitute the more difficult direction of the realization theorem, which is often demonstrated by means of induction on a cut-free sequent proof of the modal formula. In Chapter 3, Section 3.3, we have given a brief survey of sequent proof systems for modal logics with common knowledge. We will now shortly discuss theses systems with respect to their potential to be used to prove the realization theorem.

The cut-free system $\mathrm{K}^{\omega}_h(\mathsf{C})$ presented in [AJ05] and [BS09] is based on an infinitary ω-rule of the form

$$\frac{\mathsf{E}^m A, \Gamma \quad \text{for all } m \geq 1}{\mathsf{C}A, \Gamma} \qquad (\omega).$$

However, realizing such a rule presents a serious challenge because it requires achieving uniformity among the realizations of the approximants $\mathsf{E}^m A$.

Finitizing this ω-rule via the finite model property, Jäger et al. obtain the finitary cut-free system $\mathrm{K}^{<\omega}_h(\mathsf{C})$ in [JKS07]. Unfortunately, the "somewhat unusual" structural properties of the resulting system (see discussion in [JKS07]) make it hard to use it for realization.

In [Fit05; Fit10], Fitting presents a non-constructive, semantic realization method. However, this method cannot be applied directly because of the non-standard behavior of the canonical model (see Remark 6.16). See Chapter C for an outline of the problems occurring and possible approaches to overcome them.

Perhaps the infinitary system presented in [BKS10b], which is finitely branching but admits infinite branches, can help in proving the realization theorem for $\mathsf{LP}_h^{\mathsf{C}}$, however, more research on the structural properties of this system would be necessary. For example, it is not clear, how you would realize the very simple proof given in Figure 6.1. For now this remains work in progress.

6.7. Coordinated Attack

To illustrate our logic, we will now analyze the coordinated attack problem along the lines of [Fag+95], where additional references can be found. Let us briefly recall this classical problem. Suppose two divisions of an army, located in different places, are about to attack their enemy. They have some means of communication, but these may be unreliable, and the only way to secure a victory is to attack simultaneously. How should generals G and H who command the two divisions coordinate their attacks? Of course, general G could send a message m_1^G with the time of attack to general H. Let us use the proposition *del* to denote the fact that the message with the time of attack has been delivered. If the generals trust the authenticity of the message, say because of a signature, the message itself can be taken as evidence that it has been delivered. So general H, upon receiving the message, knows the time of attack: i.e., $\left[m_1^G\right]_H$ *del*. However, since communication is unreliable, G considers it possible that his message has not been delivered. But if general H sends an acknowledgment m_2^H, he in turn cannot be sure whether the acknowledgment has reached G, which prompts yet another acknowledgment m_3^G by general G, and so on.

In fact, common knowledge of *del* is a necessary condition for the attack. Indeed, it is reasonable to assume it to be common knowledge between the generals that they should only attack simultaneously or not

$$\dfrac{\dfrac{P,\neg P}{\mathsf{E}(P\vee\neg P)}(\mathsf{E}),(\vee) \quad \dfrac{\dfrac{\dfrac{P,\neg P}{\mathsf{E}(P\vee\neg P)}(\mathsf{E}),(\vee)\quad \dfrac{\dfrac{\dfrac{P,\neg P}{\mathsf{E}(P\vee\neg P)}(\mathsf{E}),(\vee)\quad\dfrac{\dfrac{\vdots}{\mathsf{C}(P\vee\neg P)}}{\mathsf{EC}(P\vee\neg P)}(\mathsf{E})}{\mathsf{C}(P\vee\neg P)}(\mathsf{C}),(\wedge)}{\mathsf{EC}(P\vee\neg P)}(\mathsf{E})}{\mathsf{C}(P\vee\neg P)}(\mathsf{C}),(\wedge)}{\mathsf{EC}(P\vee\neg P)}(\mathsf{E})}{\mathsf{C}(P\vee\neg P)}(\mathsf{C}),(\wedge)$$

Figure 6.1.: A sample proof of $\mathsf{C}(P\vee\neg P)$

attack at all, i.e., that they attack only if both know that they attack: $[t]_{\mathsf{C}}(att \to [s]_{\mathsf{E}}\, att)$ for some terms s and t. Thus, by the induction axiom, we get $[s]_{\mathsf{E}} att \to [\mathsf{ind}(t,s)]_{\mathsf{C}}\, att$. Another reasonable assumption is that it is common knowledge that neither general attacks unless the message with the time of attack has been delivered: $[r]_{\mathsf{C}}(att \to del)$ for some term r. Using the application axiom, we obtain $[s]_{\mathsf{E}} att \to [r \cdot \mathsf{ind}(t,s)]_{\mathsf{C}}\, del$. Now, from $[t]_{\mathsf{C}}(att \to [s]_{\mathsf{E}}\, att)$ again, we obtain $att \to [s]_{\mathsf{E}}\, att$ and thus by propositional reasoning

$$att \to [r \cdot \mathsf{ind}(t,s)]_{\mathsf{C}}\, del \enspace .$$

We now show that common knowledge of del cannot be achieved and that consequently no attack will take place, no matter how many messages and acknowledgments $m_1^G, m_2^H, m_3^G, \ldots$ are sent by the generals, even if all the messages are successfully delivered.

In the classical modeling without evidence, the reason is that the sender of the last message always considers the possibility that his last message, say m_{2k}^H, has not been delivered. To give a flavor of the argument carried out in detail in [Fag+95], we provide a countermodel where m_2^H is the last message, it has been delivered, but H is unsure of that: i.e.,

$$\left[m_1^G\right]_H del, \text{ and } \left[m_2^H\right]_G \left[m_1^G\right]_H del, \qquad \text{but } \neg\, [s]_H \left[m_2^H\right]_G \left[m_1^G\right]_H del$$

for all terms s. Consider any model $\mathcal{M}$ where $W := \{0,1,2,3\}$, $\nu(del) := \{0,1,2\}$, R_G is the reflexive closure of $\{(1,2)\}$, R_H is the reflexive closure of $\{(0,1),(2,3)\}$. The only requirements on the evidence function $\mathcal{E}$ are to satisfy $del \in \mathcal{E}_H\left(0, m_1^G\right)$ and $\left[m_1^G\right]_H del \in \mathcal{E}_G\left(0, m_2^H\right)$. Whatever $\mathcal{E}_{\mathsf{C}}$ is, we have $\mathcal{M}, 0 \nVdash [s]_H \left[m_2^H\right]_G \left[m_1^G\right]_H del$ and $\mathcal{M}, 0 \nVdash [t]_{\mathsf{C}}\, del$ for any s and t because $\mathcal{M}, 3 \nVdash del$.

Let us investigate a different scenario. In our models with evidence terms, there is an alternative possibility for the lack of knowledge: insufficient evidence. For example, G may receive the acknowledgment m_2^H but may not consider it to be evidence for $\left[m_1^G\right]_H del$ because the signature of H is missing. We now demonstrate that common knowledge of the time of attack cannot emerge, basing the argument solely on the lack of common knowledge evidence, in contrast to the classical approach.

Consider the M-model $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ obtained as follows: $W := \{w\}$, $R_i := \{(w, w)\}$, $\nu(del) := \{w\}$, and $\mathcal{E}$ is the minimal evidence function such that $del \in \mathcal{E}_H\left(w, m_1^G\right)$ and $\left[m_1^G\right]_H del \in \mathcal{E}_G\left(w, m_2^H\right)$. In this model, $M, w \not\Vdash [t]_{\mathsf{C}}\, del$ for any evidence term t because $del \notin \mathcal{E}_{\mathsf{C}}(w, t)$ for any t. To prove the latter statement, it is sufficient to note that for any term t, by Lemma 6.29,

$$\nvdash \left[m_1^G\right]_H del \wedge \left[m_2^H\right]_G \left[m_1^G\right]_H del \to [t]_{\mathsf{C}}\, del \tag{6.9}$$

because

$$\mathsf{S4}_h^{\mathsf{C}} \nvdash \Box_H\, del \wedge \Box_G \Box_H\, del \to \mathsf{C}\, del \ ,$$

which is easy to demonstrate. Let $\mathcal{M}^{\mathrm{can}}$ be the canonical epistemic model meeting the empty constant specification and let $\mathcal{E}^{\mathrm{can}}$ be its evidence function. Since the negation of the formula from (6.9) must be satisfiable, for each t there is a world w_t from $\mathcal{M}^{\mathrm{can}}$ such that $del \in \mathcal{E}_H^{\mathrm{can}}\left(w_t, m_1^G\right)$ and $\left[m_1^G\right]_H del \in \mathcal{E}_G^{\mathrm{can}}\left(w_t, m_2^H\right)$, but by the Truth Lemma 6.17, $del \notin \mathcal{E}_{\mathsf{C}}^{\mathrm{can}}(w_t, t)$. Since $\mathcal{E}^{\mathrm{can}} \restriction (\{w_t\} \times \mathrm{Tm})$ satisfies all the closure conditions, the minimality of $\mathcal{E}$ implies that $\mathcal{E}_{\mathsf{C}}(w, s) \subseteq \mathcal{E}_{\mathsf{C}}^{\mathrm{can}}(w_t, s)$ for any term s. In particular, $del \notin \mathcal{E}_{\mathsf{C}}(w, t)$ for any term t.

6.8. Discussion

In this chapter, we have provided a system of evidence terms for describing common knowledge, which can be used instead of modal logic representation. One benefit of this new representation is that several proofs that are quite hard in the modal case, e.g., those of completeness and conservativity, are made easier in our logic. There are other merits to this system as well.

In the single-agent case, as is pointed out in [Art08], an explicit codification of knowledge by evidence (in Artemov's case, of the individual knowledge of the agent) enables knowledge to be analyzed and recorded. Recording and subsequent retrieving of evidence can be viewed as a form of single-agent communication, with which any mathematician is familiar. A proof of a theorem, if not recorded immediately, may require as much effort to be restored later as finding it required originally. This role of evidence terms in knowledge transfer is reminiscent of what is called *explicit knowledge* in Knowledge Management[2] and is contrasted with *tacit knowledge*. As described in [Non91], "Explicit knowledge is formal and systematic. For this reason, it can be easily communicated and shared, in product specifications or a scientific formula or a computer program." In this sense, evidence terms in the single-agent case serve as a kind of explicit knowledge. Indeed, if an agent can find a proof he/she wrote down a year ago, it will restore his/her knowledge of the statement of the theorem.

The situation with common knowledge evidence is more complicated. An evidence of common knowledge of some fact A, even when transmitted to all agents and received by them[3], does not generally create common knowledge of A for the same reasons that were discussed in the previous section. In fact, there exist general results about the impossibility of achieving common knowledge via certain modes of communication, e.g., in asynchronous systems [Fag+95]. Clearly, an introduction of evidence terms cannot and should not change this general phenomenon.

However, one can think of modes of communication that ensure that a transmission of a common knowledge evidence term to all the agents in the group does create common knowledge among the agents. An example of such a mode is public announcements (see also Chapter B), a well-known method of creating common knowledge. Thus, one of the benefits of our system of terms is a finite encoding of common knowledge, which is largely infinitary in nature. This finite encoding enables to transmit evidence, which, under certain modes of communication, creates common knowledge among the agents. Of course, common knowledge can also be created by a public announcement of the fact itself rather than of evidence in support of the fact. There is an important difference, however. When, in his seminal 1989 work [Pla07a], Plaza analyzed one of the standard stories used to explain the concept of common knowledge, the Muddy Children Puzzle, in order to explain how common

[2] The term "explicit knowledge" sounds so natural that it has been used in different areas with completely different meanings. For instance, in epistemic logic, explicit knowledge is a type of knowledge that is not logically omniscient, as opposed to implicit knowledge [Fag+95].

[3] Unreliable communication does not prevent knowledge from being explicit. Thus, in the context of explicit vs. tacit knowledge, we only discuss the usefulness of evidence terms that have been received by the agent(s).

knowledge is created by a public announcement, he had to assume that the announcements are truthful and the agents are trustful. Indeed, an announced fact cannot become common knowledge, or any kind of knowledge, if the fact is false. And clearly, if the agents do not trust the announcement, their knowledge would only change provided they can verify the announced facts.

Verifiability of announcements is exactly what we achieve by introducing evidence terms into the language. An agent who receives a justification for A needs neither to assume that A is true nor to trust the speaker because the agent can simply verify the received information. A similar idea of supplying messages with justifications can be used to describe a distributed system that authorizes the disbursement of sensitive data, such as medical records, while maintaining a specified privacy policy [Bla+11]. Interestingly, like in our analysis of the coordinated attack, the authors also propose to use the sender's signature as evidence for the information about his/her intentions or policies.

Verifiability of evidence turns out to be sufficient for creating common knowledge. Indeed, Yavorskaya [Yav08] considered a situation where agents can verify each other's evidence, formally:

$$[t]_i A \to \left[!_i^j t\right]_j [t]_i A \text{ for } i \neq j \ . \tag{6.10}$$

The $!_i^j$-operation implicitly presumes communication since i's evidence t has to be somehow available to agent j. It is not hard to show that an addition of this operation to our logic leads to a situation where any individual knowledge also automatically creates common knowledge of the same fact: Let L denote $\mathsf{LP}_h^{\mathsf{C}}$ extended with (6.10) as additional axioms. Extension here means, that also the constant specification is extended in order to include the new axioms and the language is adapted accordingly in order to include the new operations $!_i^j$.

Lemma 6.31. *For each agent i, term $t \in \mathcal{E}_i$ and formula $A \in \mathrm{Fm}_{\mathsf{LP}_h^{\mathsf{C}}}$ there is a term $t \in \mathcal{E}_{\mathsf{C}}$ such that*

$$\mathsf{L} \vdash [s]_i A \to [t(s)]_{\mathsf{C}} A.$$

Proof. First of all note, that constructive necessitation still holds for the extended logic L, as we only added axioms. The proof can be trivially adapted for the new system, it is literally the same.

We have

$$\mathsf{L} \vdash [s]_i A \to \left[!_i^j t\right]_j [t]_i A$$

for all $j \neq i$ by our extension and

$$\mathsf{L} \vdash [s]_i A \to [!t]_i [t]_i A$$

by the inspection axiom.

Thus by propositional reasoning and the tupling axiom

$$\mathsf{L} \vdash [s]_i A \to \left[\left\langle !_i^1 s, \ldots, !_i^{i-1} s, !s, !_i^{i+1} s, \ldots, !_i^h s\right\rangle\right]_{\mathsf{E}} [s]_i A$$

As mentioned above, constructive necessitation holds for the extended logic and thus there is a term $r \in \mathcal{E}_{\mathsf{C}}$ such that

$$\mathsf{L} \vdash [r]_{\mathsf{C}}([s]_i A \to \left[\left\langle !_i^1 s, \ldots, !_i^{i-1} s, !s, !_i^{i+1} s, \ldots, !_i^h s\right\rangle\right]_{\mathsf{E}} [s]_i A)$$

Using propositional reasoning and the induction axiom we obtain

$$\mathsf{L} \vdash [s]_i A \to \left[\mathsf{ind}(r, \left\langle !_i^1 s, \ldots, !_i^{i-1} s, !s, !_i^{i+1} s, \ldots, !_i^h s\right\rangle)\right]_{\mathsf{C}} [s]_i A$$

There is a constant $c \in \mathcal{CS}$ witnessing the reflexivity axiom, i.e.

$$\mathsf{L} \vdash [c]_{\mathsf{C}}([s]_i A \to A)$$

and so, using the propositional reasoning and the application axiom we finally obtain

$$\mathsf{L} \vdash [s]_i A \to \left[c \cdot \mathsf{ind}(r, \left\langle !_i^1 s, \ldots, !_i^{i-1} s, !s, !_i^{i+1} s, \ldots, !_i^h s\right\rangle)\right]_{\mathsf{C}} A$$

Thus, we can set

$$t(x) := c \cdot \mathsf{ind}(r, \left\langle !_i^1 x, \ldots, !_i^{i-1} x, !x, !_i^{i+1} x, \ldots, !_i^h x\right\rangle) \ .$$

□

However, the $!_i^j$-operation depends on a mode of communication that must be reliable and immediate in order to work, which restricts the applicability of such a logic; for instance, it precludes an analysis of asynchronous systems.

Note that also the second form of evidence transfer proposed by Yavorskaya [Yav08] leads to the same result: Let L^* denote $\mathsf{LP}_h^{\mathsf{C}}$ extended with the following axioms

$$[t]_i A \to \left[\uparrow_i^j t\right]_j A \text{ for } i \neq j$$

where extension is understood as above.

Corollary 6.32. *For each agent i, term $t \in \mathcal{E}_i$ and formula $A \in \mathrm{Fm}_{\mathsf{LP}_h^{\mathsf{C}}}$ there is a term $t \in \mathcal{E}_{\mathsf{C}}$ such that*

$$\mathsf{L}^* \vdash [s]_i A \to [t(s)]_{\mathsf{C}} A.$$

Proof. It can be easily shown that the operation $!_i^j$ is definable in L^*. (Actually, also $\uparrow_i^j$ can be shown to be easily definable in L, but this is not of importance to this corollary).

We have

$$\mathsf{L}^* \vdash [t]_i A \to [!t]_i \, [t]_i A$$

by the inspection axiom and thus

$$\mathsf{L}^* \vdash [t]_i A \to \left[\uparrow_i^j !t\right]_j [t]_i A \text{ for } i \neq j.$$

by the newly added conversion axioms, i.e. we can set

$$!_i^j x := \uparrow_i^j !$$

Now we can apply the proof of Lemma 6.31 and immediately obtain the desired result. ☐

In summary, the kind of knowledge that can be induced via justification transmission is generally the same as in the case of statement transmission and depends primarily on the mode of communication, on its reliability.

So another benefit of introducing evidence terms is their verifiability, including cases when evidence terms are communicated between agents. Yet another benefit, this time on the meta-logical level, is an ability to analyze common knowledge and the process of its creation. Similar to Artemov's analysis of the famous Gettier examples in [Art08], the system of evidence terms for common knowledge can also be used to uncover hidden assumptions. Further, as shown in the previous section, it can yield new scenarios for well-known epistemic puzzles.

Our contribution is technical in the sense that we aim to study neither the nature of common knowledge nor ways of transmitting data to achieve it. Our goal is to provide tools for analyzing the fine structure of common knowledge, tools that can be used, irrespective of the mode of communication between the agents, even when the communication itself remains on the meta-logical level as in the standard rendition of the Muddy Children Puzzle, e.g., in [Fag+95].

7. More Justification Logics with Common Knowledge

MORE, *adj.* The comparative degree of too much.

Ambrose Bierce, The Devil's Dictonary [Bie11]

In the previous chapter we presented a justification logic with common knowledge with LP as its fixed base logic. In this chapter we generalize this logic to a selection of base logics and, furthermore, extend our language in order to talk about groups of agents. This is done in a similar manner as the modal logic $\mathsf{S4}_h^{\mathsf{C}}$ in Chapter 3, Section 3.5 was generalized to the modal logics $\mathsf{L}_h^{\mathsf{C}}$ such as e.g. $\mathsf{KD4}_h^{\mathsf{C}}$ or $\mathsf{K5}_h^{\mathsf{C}}$.

Most of the proofs presented here are analogous to the proofs presented in the previous chapter. We will thus often abbreviate the proofs and only mention the most notable changes where necessary.

7.1. Syntax

As before, we consider a system of h agents. Let us also recall the following notion from Chapter 3: By *a group of agents* $\mathsf{G} = \langle i_1, \dots, i_k \rangle$ we mean a (non-empty) tuple of $i_j \in \{1, \dots, h\}$ with $i_1 < i_2 < \dots < i_k$. We will use set notation $i_j \in \mathsf{G}$ and $\mathsf{G} \subseteq \mathsf{H}$ to state that i_j occurs in G and all $i_j \in \mathsf{G}$ occur in H, respectively. Throughout this chapter[1], i always denotes an element of $\{1, \dots, h\}$, $*$ always denotes an element of $\{1, \dots, h\} \cup \{\mathsf{C}_\mathsf{G} \mid \mathsf{G} \text{ a group of agents}\}$, and $\circledast$ always denotes an element of $\{1, \dots, h\} \cup \{\mathsf{E}_\mathsf{G}, \mathsf{C}_\mathsf{G} \mid \mathsf{G} \text{ a group of agents}\}$.

[1]Note that these conventions have been slightly changed in order to include groups of agents.

Let $\mathrm{Cons}_{\circledast} := \{c_1^{\circledast}, c_2^{\circledast}, \ldots\}$ and $\mathrm{Var}_{\circledast} := \{x_1^{\circledast}, x_2^{\circledast}, \ldots\}$ be countable sets of *proof constants* and *proof variables* respectively for each type of knowledge $\circledast$.

The operations on evidence terms are relativized with respects to groups of agents and furthermore a restriction operation is introduced to pass common knowledge from a group of agents to a subgroup. See also Section 3.5 in Chapter 3 for a discussion of the necessity of the corresponding axioms. The sets $\mathrm{Tm}_1, \ldots, \mathrm{Tm}_h$, $\mathrm{Tm}_{\mathsf{E}_\mathsf{G}}$, and $\mathrm{Tm}_{\mathsf{C}_\mathsf{G}}$ (for each group of agents G) of *evidence terms for individual agents* and for *mutual* and *common knowledge* for groups of agents respectively are inductively defined as follows:

1. $\mathrm{Cons}_{\circledast} \subseteq \mathrm{Tm}_{\circledast}$;
2. $\mathrm{Var}_{\circledast} \subseteq \mathrm{Tm}_{\circledast}$;
3. $t +_* s \in \mathrm{Tm}_*$ and $t \cdot_* s \in \mathrm{Tm}_*$ for any $t, s \in \mathrm{Tm}_*$;
4. $\langle t_{i_1}, \ldots, t_{i_k} \rangle_{\mathsf{G}} \in \mathrm{Tm}_{\mathsf{E}_\mathsf{G}}$ for any group of agents $\mathsf{G} = \{i_1, \ldots, i_k\}$ and $t_{i_j} \in \mathrm{Tm}_{i_j}$;
5. $\pi_i^{\mathsf{G}} t \in \mathrm{Tm}_i$ for any group of agents G with $i \in \mathsf{G}$ and $t \in \mathrm{Tm}_{\mathsf{E}_\mathsf{G}}$;
6. $\downarrow_{\mathsf{H}}^{\mathsf{G}} t \in \mathrm{Tm}_{\mathsf{C}_\mathsf{H}}$ for any groups of agents $\mathsf{H} \subseteq \mathsf{G}$ and $t \in \mathrm{Tm}_{\mathsf{C}_\mathsf{G}}$;
7. $\mathsf{head}_{\mathsf{G}}(t) \in \mathrm{Tm}_{\mathsf{E}_\mathsf{G}}$ and $\mathsf{tail}_{\mathsf{G}}(t) \in \mathrm{Tm}_{\mathsf{E}_\mathsf{G}}$ for any group of agents G and $t \in \mathrm{Tm}_{\mathsf{C}_\mathsf{G}}$;
8. $\mathsf{ind}_{\mathsf{G}}(t, s) \in \mathrm{Tm}_{\mathsf{C}_\mathsf{G}}$ for any group of agents G, any $t \in \mathrm{Tm}_{\mathsf{C}_\mathsf{G}}$ and any $s \in \mathrm{Tm}_{\mathsf{E}_\mathsf{G}}$.

By $\mathrm{Tm}_{\circledast}^{!}$ we denote the sets of terms obtained by allowing the following additional clause

9. $!_i t \in \mathrm{Tm}_i$ for any $t \in \mathrm{Tm}_i$,

by $\mathrm{Tm}_{\circledast}^{?}$ we denote the sets of terms obtained by allowing the following additional clause

10. $?_i t \in \mathrm{Tm}_i$ for any $t \in \mathrm{Tm}_i$,

and by $\mathrm{Tm}^{!?}_{\circledast}$ the set of terms obtained by allowing both additional clauses above.

$\mathrm{Tm} := \mathrm{Tm}_1 \cup \ldots \cup \mathrm{Tm}_h \cup \mathrm{Tm}_{\mathsf{E}} \cup \mathrm{Tm}_{\mathsf{C}}$ denotes the set of all evidence terms (accordingly with superscripts !, ? or both). The indices of the operations !, ?, +, and $\cdot$ will usually be omitted if they can be inferred from the context.

Let $\mathrm{Prop} := \{P_1, P_2, \ldots\}$ be a countable set of *propositional variables. Formulae* are denoted by A, B, C, etc. and defined by the following grammar:

$$A ::= P_j \mid \bot \mid (A \to A) \mid [t]_{\circledast} A \ ,$$

where $t \in \mathrm{Tm}_{\circledast}$. The set of all formulae is denoted by $\mathrm{Fm}_{\mathsf{L}^{\mathsf{C}}_h}$. The sets of formulae $\mathrm{Fm}^{!}_{\mathsf{L}^{\mathsf{C}}_h}$, $\mathrm{Fm}^{?}_{\mathsf{L}^{\mathsf{C}}_h}$ and $\mathrm{Fm}^{!?}_{\mathsf{L}^{\mathsf{C}}_h}$ are defined equally with $\mathrm{Tm}^{!}$, $\mathrm{Tm}^{?}$, and $\mathrm{Tm}^{!?}$, respectively, in place of Tm. We will usually omit superscripts ! and ? whenever safe, and we will also assume that whenever a statement concerning several of the logics defined below is made, the language is always chosen appropriately (see table 7.2 to see precisely which logic uses which language). Furthermore, we adopt the following convention: whenever a formula $[t]_{\circledast} A$ is used, it is assumed to be well-formed, i.e., it is implicitly assumed that term $t \in \mathrm{Tm}_{\circledast}$. This enables us to omit the explicit typification of terms.

The axioms of $\mathsf{J}^{\mathsf{C}}_h$ are

1. all propositional tautologies (taut)
2. $[t]_*(A \to B) \to ([s]_* A \to [t \cdot s]_* B)$ (application)
3. $[t]_* A \vee [s]_* A \to [t + s]_* A$ (sum)
4. $[t_{i_1}]_{i_1} A \wedge \ldots \wedge [t_{i_k}]_{i_k} A \to \big[\langle t_{i_1}, \ldots, t_{i_k}\rangle_{\mathsf{G}}\big]_{\mathsf{E}_{\mathsf{G}}} A$
   for any group of agents $\mathsf{G} = \{i_1, \ldots, i_k\}$ (tupling)
5. $[t]_{\mathsf{E}_{\mathsf{G}}} A \to \big[\pi^{\mathsf{G}}_i t\big]_i A$
   for any group of agents G with $i \in \mathsf{G}$ (projection)
6. $[t]_{\mathsf{C}_{\mathsf{G}}} A \to \big[{\downarrow}^{\mathsf{G}}_{\mathsf{H}} t\big]_{\mathsf{C}_{\mathsf{H}}} A$
   for any groups of agents $\mathsf{H} \subseteq \mathsf{G}$ (restriction)
7. $[t]_{\mathsf{C}_{\mathsf{G}}} A \to [\mathsf{head}_{\mathsf{G}}(t)]_{\mathsf{E}_{\mathsf{G}}} A$ and $[t]_{\mathsf{C}_{\mathsf{G}}} A \to [\mathsf{tail}_{\mathsf{G}}(t)]_{\mathsf{E}_{\mathsf{G}}} [t]_{\mathsf{C}_{\mathsf{G}}} A$
   for any group of agents G (co-closure)

8. $[t]_{\mathsf{C_G}}(A \to [r]_{\mathsf{E_G}}A) \to ([s]_{\mathsf{E_G}}A \to [\mathsf{ind_G}(t,s)]_{\mathsf{C_G}}A)$ (induction)

We furthermore also consider the extensions given by table 7.2 using the additional axioms from table 7.1, where all axiom schemas are taken from the appropriate language.

A *constant specification* $\mathcal{CS}$ for the logic $\mathsf{L}_h^{\mathsf{C}}$ is any subset

$$\mathcal{CS} \subseteq \bigcup_{\circledast \in \{1,\dots,h,E,C\}} \left\{[c]_\circledast A \ : \ c \in \mathrm{Cons}_\circledast \text{ and } A \text{ is an axiom of } \mathsf{L}_h^{\mathsf{C}}\right\}.$$

A constant specification $\mathcal{CS}$ for the logic $\mathsf{L}_h^{\mathsf{C}}$ is called C-*axiomatically appropriate* if for each axiom A of $\mathsf{L}_h^{\mathsf{C}}$, there is a proof constant $c \in \mathrm{Cons}_{\mathsf{C_H}}$ such that $[c]_{\mathsf{C_H}}A \in \mathcal{CS}$ where $\mathsf{H} = \{1,\dots,h\}$.

A constant specification $\mathcal{CS}$ is called *homogeneous*, if

$$\mathcal{CS} \subseteq \{[c]_\circledast A \ : \ c \in \mathrm{Cons}_\circledast \text{ and } A \text{ is an axiom}\}$$

for some fixed $\circledast$, i.e., if for all $[c]_\circledast A \in \mathcal{CS}$, the constants c are of the same type.

Let $\mathcal{CS}$ be a constant specification for the logic $\mathsf{L}_h^{\mathsf{C}}$. The deductive system $\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})$ is the Hilbert system given by the axioms of $\mathsf{L}_h^{\mathsf{C}}$ as given above and rules modus ponens and axiom necessitation:

$$\frac{A \quad A \to B}{B}\ (\mathrm{MP})\ , \qquad \frac{}{[c]_\circledast A}\ (\mathrm{AN})\ , \text{ where } [c]_\circledast A \in \mathcal{CS}.$$

By $\mathsf{L}_h^{\mathsf{C}}$ we denote the system $\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})$ with

$$\mathcal{CS} = \left\{[c]_{\mathsf{C_H}}A \ : \ c \in \mathrm{Cons}_{\mathsf{C_H}} \text{ and } A \text{ is an axiom of } \mathsf{L}_h^{\mathsf{C}}\right\} \tag{7.1}$$

with $\mathsf{H} = \{1,\dots,h\}$.

For an arbitrary $\mathcal{CS}$, we write $\Delta \vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} A$ to state that A is derivable from Δ in $\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})$. We use Δ, A to mean $\Delta \cup \{A\}$.

Note that the logics presented here are modelled on the logics from Chapter 3, Section 3.5 where we discussed, why we need the group restriction axiom and necessitation for the group of all agents. As usual for justification logics, we only have necessitation for axioms and we will show in Corollary 7.8 that we have constructive necessitation, the justification counterpart of "full" necessitation. Furthermore, in contrast

(jd)	$[t]_i\bot \to \bot$	(seriality)
(jt)	$[t]_i A \to A$	(reflexivity)
(j4)	$[t]_i A \to [!t]_i\,[t]_i A$	(positive inspection)
(j5)	$\neg\,[t]_i A \to [?t]_i(\neg\,[t]_i A)$	(negative inspection)

Table 7.1.: Additional Axioms

Logic $\mathsf{L}_h^{\mathsf{C}}$	$\mathsf{J}_h^{\mathsf{C}}$	$\mathsf{JD}_h^{\mathsf{C}}$	$\mathsf{JD4}_h^{\mathsf{C}}$	$\mathsf{JD5}_h^{\mathsf{C}}$	$\mathsf{JD45}_h^{\mathsf{C}}$	$\mathsf{JT}_h^{\mathsf{C}}$	$\mathsf{JT4}_h^{\mathsf{C}}$	$\mathsf{JT5}_h^{\mathsf{C}}$	$\mathsf{JT45}_h^{\mathsf{C}}$	$\mathsf{J4}_h^{\mathsf{C}}$	$\mathsf{J45}_h^{\mathsf{C}}$	$\mathsf{J5}_h^{\mathsf{C}}$
$\mathrm{Tm}^{\mathsf{L}_h^{\mathsf{C}}}$	Tm	Tm	$\mathrm{Tm}^!$	$\mathrm{Tm}^?$	$\mathrm{Tm}^{!?}$	Tm	$\mathrm{Tm}^!$	$\mathrm{Tm}^?$	$\mathrm{Tm}^{!?}$	$\mathrm{Tm}^!$	$\mathrm{Tm}^{!?}$	$\mathrm{Tm}^?$
(jd)		✓	✓	✓	✓							
(jt)						✓	✓	✓	✓			
(j4)			✓		✓		✓		✓	✓	✓	
(j5)				✓	✓			✓	✓		✓	✓

Table 7.2.: Extensions

to the situation in Chapter 5, we need no iterated axiom necessitation rule like (AN!) even if the base logic does not contain positive inspection, as common knowledge possesses this ability regardless of the base logic, see Lemma 7.4.

7.2. Basic Properties

The following lemmas are immediate adaptions of the lemmas in Section 6.2 from Chapter 6 taking into account the presence of groups of agents and different reasoning capabilities of the base logic that can be lifted to common knowledge.

Lemma 7.1. *For any logic* $\mathsf{L}_h^{\mathsf{C}}$*, any group of agents* G*, any constant specification* $\mathcal{CS}$ *and any formulae* A *and* B*:*

1. *for any* $t, s \in \mathrm{Tm}_{\mathsf{E_G}}^{\mathsf{L}_h^{\mathsf{C}}}$*, there is a term* $t \cdot_{\mathsf{E_G}} s \in \mathrm{Tm}_{\mathsf{E_G}}^{\mathsf{L}_h^{\mathsf{C}}}$ *such that*

$$\vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} [t]_{\mathsf{E_G}}(A \to B) \to ([s]_{\mathsf{E_G}}A \to [t \cdot_{\mathsf{E_G}} s]_{\mathsf{E_G}}B);$$

($\mathsf{E_G}$-*application*)

2. *for any* $t, s \in \mathrm{Tm}_{\mathsf{E_G}}^{\mathsf{L}_h^{\mathsf{C}}}$*, there is a term* $t +_{\mathsf{E_G}} s \in \mathrm{Tm}_{\mathsf{E_G}}^{\mathsf{L}_h^{\mathsf{C}}}$ *such that*

$$\vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} [t]_{\mathsf{E_G}}A \vee [s]_{\mathsf{E_G}}A \to [t +_{\mathsf{E_G}} s]_{\mathsf{E_G}}A;$$

($\mathsf{E_G}$-*sum*)

3. *for any* $t \in \mathrm{Tm}_{\mathsf{C_G}}^{\mathsf{L}_h^{\mathsf{C}}}$ *and any* $i \in \mathsf{G}$*, there is a term* $\downarrow_i^{\mathsf{G}} t \in \mathrm{Tm}_i^{\mathsf{L}_h^{\mathsf{C}}}$ *such that*

$$\vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} [t]_{\mathsf{C_G}}A \to \left[\downarrow_i^{\mathsf{G}} t\right]_i A;$$

(i-*down-conversion*)

4. *for any* $t \in \mathrm{Tm}_{\mathsf{C_G}}^{\mathsf{L}_h^{\mathsf{C}}}$ *and any* $i \in \mathsf{G}$*, there is a term* $\uparrow_i^{\mathsf{G}} t \in \mathrm{Tm}_i^{\mathsf{L}_h^{\mathsf{C}}}$ *such that*

$$\vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} [t]_{\mathsf{C_G}}A \to \left[\uparrow_i^{\mathsf{G}} t\right]_i [t]_{\mathsf{C_G}}A;$$

(i-*up-conversion*)

Proof. Assume $\mathsf{G} = \{i_1, \ldots, i_k\}$.

1. Set $t \cdot_{\mathsf{E_G}} s := \left\langle \pi^{\mathsf{G}}_{i_1} t \cdot_{i_1} \pi^{\mathsf{G}}_{i_1} s, \ldots, \pi^{\mathsf{G}}_{i_k} t \cdot_{i_k} \pi^{\mathsf{G}}_{i_k} s \right\rangle_{\mathsf{G}}$.
2. Set $t +_{\mathsf{E_G}} s := \left\langle \pi^{\mathsf{G}}_{i_1} t +_{i_1} \pi^{\mathsf{G}}_{i_1} s, \ldots, \pi^{\mathsf{G}}_{i_k} t +_{i_k} \pi^{\mathsf{G}}_{i_k} s \right\rangle_{\mathsf{G}}$.
3. Set $\downarrow^{\mathsf{G}}_i t := \pi^{\mathsf{G}}_i \mathsf{head}_{\mathsf{G}}(t)$.
4. Set $\uparrow^{\mathsf{G}}_i t := \pi^{\mathsf{G}}_i \mathsf{tail}_{\mathsf{G}}(t)$. □

Lemma 7.2. *For any logic $\mathsf{L}^{\mathsf{C}}_h$ containing (jd), any group of agents* G *and any constant specification $\mathcal{CS}$:*

1. $\vdash_{\mathsf{L}^{\mathsf{C}}_h(\mathcal{CS})} [t]_{\mathsf{E_G}} \bot \to \bot$ *for all* $t \in \mathrm{Tm}^{\mathsf{L}^{\mathsf{C}}_h}_{\mathsf{E_G}}$; ($\mathsf{E_G}$-*seriality*)
2. $\vdash_{\mathsf{L}^{\mathsf{C}}_h(\mathcal{CS})} [t]_{\mathsf{C_G}} \bot \to \bot$ *for all* $t \in \mathrm{Tm}^{\mathsf{L}^{\mathsf{C}}_h}_{\mathsf{C_G}}$. ($\mathsf{C_G}$-*seriality*)

Proof. 1. Immediate by the projection and seriality axioms.

2. Immediate by 3 from Lemma 7.1. and the seriality axiom. □

Lemma 7.3. *For any logic $\mathsf{L}^{\mathsf{C}}_h$ containing (jt), any group of agents* G*, any constant specification $\mathcal{CS}$ and any formulae A:*

1. $\vdash_{\mathsf{L}^{\mathsf{C}}_h(\mathcal{CS})} [t]_{\mathsf{E_G}} A \to A$ *for all* $t \in \mathrm{Tm}^{\mathsf{L}^{\mathsf{C}}_h}_{\mathsf{E_G}}$; ($\mathsf{E_G}$-*reflexivity*)
2. $\vdash_{\mathsf{L}^{\mathsf{C}}_h(\mathcal{CS})} [t]_{\mathsf{C_G}} A \to A$ *for all* $t \in \mathrm{Tm}^{\mathsf{L}^{\mathsf{C}}_h}_{\mathsf{C_G}}$. ($\mathsf{C_G}$-*reflexivity*)

Proof. 1. Immediate by the projection and reflexivity axioms.

2. Immediate by 3 from Lemma 7.1. and the reflexivity axiom. □

Unlike the previous lemmas, the next lemmas require that a constant specification $\mathcal{CS}$ be C-axiomatically appropriate.

Lemma 7.4. *Let $\mathsf{L}^{\mathsf{C}}_h$ be any of the logics defined above,* G *a group of agents, $\mathcal{CS}$ a* C*-axiomatically appropriate constant specification and A be a formula.*

1. *For any* $t \in \mathrm{Tm}^{\mathsf{L}^{\mathsf{C}}_h}_{\mathsf{C_G}}$*, there is a term* $!_{\mathsf{C_G}} t \in \mathrm{Tm}^{\mathsf{L}^{\mathsf{C}}_h}_{\mathsf{C_G}}$ *such that* $\vdash_{\mathsf{L}^{\mathsf{C}}_h(\mathcal{CS})} [t]_{\mathsf{C_G}} A \to [!_{\mathsf{C_G}} t]_{\mathsf{C_G}} [t]_{\mathsf{C_G}} A$. ($\mathsf{C_G}$-*inspection*)
2. *For any* $t \in \mathrm{Tm}^{\mathsf{L}^{\mathsf{C}}_h}_{\mathsf{C_G}}$*, there is a term* $\Leftarrow_{\mathsf{C_G}} t \in \mathrm{Tm}^{\mathsf{L}^{\mathsf{C}}_h}_{\mathsf{C_G}}$ *such that* $\vdash_{\mathsf{L}^{\mathsf{C}}_h(\mathcal{CS})} [t]_{\mathsf{C_G}} A \to [\Leftarrow_{\mathsf{C_G}} t]_{\mathsf{C_G}} [\mathsf{head}_{\mathsf{G}}(t)]_{\mathsf{E_G}} A$. ($\mathsf{C_G}$-*shift*)

Proof. Let $\mathsf{H} = \{1, \ldots, h\}$.

1. Set
$$!_{\mathsf{C_G}} t := \mathsf{ind_G}(\downarrow^{\mathsf{H}}_{\mathsf{G}} c, \mathsf{head_G}(t)) \ ,$$
where $[c]_{\mathsf{C_H}}([t]_{\mathsf{C_G}} A \to [\mathsf{tail_G}(t)]_{\mathsf{E_G}} [t]_{\mathsf{C_G}} A) \in \mathcal{CS}$.

2. Set
$$\Leftarrow_{\mathsf{C_G}} t := \downarrow^{\mathsf{H}}_{\mathsf{G}} c \cdot_{\mathsf{C_G}} (!_{\mathsf{C_G}} t) \ ,$$
where $[c]_{\mathsf{C_H}}([t]_{\mathsf{C_G}} A \to [\mathsf{head_G}(t)]_{\mathsf{E_G}} A) \in \mathcal{CS}$. □

As in the previous chapter, the following two theorems are standard in justification logics. Their proofs can be taken almost word for word from [Art01] and are, therefore, omitted here.

Lemma 7.5 (Deduction Theorem)**.** *Let $\mathsf{L}^{\mathsf{C}}_h$ be any of the logics defined above and $\mathcal{CS}$ be a constant specification and $\Delta \cup \{A, B\} \subseteq \mathrm{Fm}_{\mathsf{L}^{\mathsf{C}}_h}$. Then $\Delta, A \vdash_{\mathsf{L}^{\mathsf{C}}_h(\mathcal{CS})} B$ if and only if $\Delta \vdash_{\mathsf{L}^{\mathsf{C}}_h(\mathcal{CS})} A \to B$.*

Lemma 7.6 (Substitution)**.** *Let $\mathsf{L}^{\mathsf{C}}_h$ be any of the logics defined above. For any constant specification $\mathcal{CS}$, any propositional variable P, any $\Delta \cup \{A, B\} \subseteq \mathrm{Fm}_{\mathsf{L}^{\mathsf{C}}_h}$, any $x \in \mathrm{Var}_{\circledast}$, and any $t \in \mathrm{Tm}_{\circledast}$,*

$$\text{if } \Delta \vdash_{\mathsf{L}^{\mathsf{C}}_h(\mathcal{CS})} A, \text{ then } \Delta(x/t, P/B) \vdash_{\mathsf{L}^{\mathsf{C}}_h(\mathcal{CS}(x/t,P/B))} A(x/t, P/B) \ ,$$

where $A(x/t, P/B)$ denotes the formula obtained by simultaneously replacing all occurrences of x in A with t and all occurrences of P in A with B, accordingly for $\Delta(x/t, P/B)$ and $\mathcal{CS}(x/t, P/B)$.

The following lemma states that our logic can internalize its own proofs, which is an important property of justification logics. The proof is very similar to the corresponding proof in the previous chapter. However, groups of agents have to be taken into account and as is to be expected, this is witnessed by the frequent usage of the group restriction operator.

Lemma 7.7 (C-lifting)**.** *Let $\mathsf{L}^{\mathsf{C}}_h$ be any of the logics defined above, $\mathcal{CS}$ a homogeneous C-axiomatically appropriate constant specification and $\mathsf{H} = \{1, \ldots, h\}$. If*

$$[s_1]_{\mathsf{C_H}} B_1, \ldots, [s_n]_{\mathsf{C_H}} B_n, C_1, \ldots, C_m \vdash_{\mathsf{L}^{\mathsf{C}}_h(\mathcal{CS})} A \ ,$$

then for each $\circledast$, *there is a term* $t_\circledast(x_1, \dots, x_n, y_1, \dots, y_m) \in \mathrm{Tm}_\circledast$ *such that*

$$[s_1]_{\mathsf{C_H}} B_1, \dots, [s_n]_{\mathsf{C_H}} B_n, [y_1]_\circledast C_1, \dots, [y_m]_\circledast C_m \vdash_{\mathsf{L}^{\mathsf{C}}_h(\mathcal{CS})} [t_\circledast(s_1, \dots, s_n, y_1, \dots, y_m)]_\circledast A$$

for fresh variables $y_1, \dots, y_m \in \mathrm{Tm}_\circledast$.

Proof. We proceed by induction on the derivation of A.

If A is an axiom, then there is a constant $c \in \mathrm{Tm}_{\mathsf{C_H}}$ such that $[c]_{\mathsf{C_H}} A \in \mathcal{CS}$ because $\mathcal{CS}$ is C-axiomatically appropriate. Then take

$$t_{\mathsf{C_G}} := \downarrow^{\mathsf{H}}_{\mathsf{G}} c, \qquad t_i := \downarrow^{\mathsf{H}}_{i} c, \qquad t_{\mathsf{E_G}} := \mathsf{head}_{\mathsf{G}}(\downarrow^{\mathsf{H}}_{\mathsf{G}} c)$$

and use axiom necessitation and restriction, axiom necessitation and i-conversion, or axiom necessitation, restriction and the co-closure axiom respectively.

For $A = [s_j]_{\mathsf{C}} B_j$, $1 \leq j \leq n$, take

$$t_{\mathsf{C_G}} := \downarrow^{\mathsf{H}}_{\mathsf{G}} !_{\mathsf{C_H}} s_j, \qquad t_i := \downarrow^{\mathsf{H}}_{i} !_{\mathsf{C_H}} s_j, \qquad t_{\mathsf{E_G}} := \mathsf{tail}_{\mathsf{G}}(\downarrow^{\mathsf{H}}_{\mathsf{G}} s_j)$$

and use C-inspection and restriction, $\mathsf{C_H}$-inspection and i-conversion, or the co-closure axiom and restriction respectively.

For $A = C_j$, $1 \leq j \leq m$, take $t_\circledast := y_j \in \mathrm{Var}_\circledast$ for a fresh variable y_j.

For A derived by modus ponens from $D \to A$ and D, by induction hypothesis there are terms $r_\circledast, s_\circledast \in \mathrm{Tm}_\circledast$ such that $[r_\circledast]_\circledast(D \to A)$ and $[s_\circledast]_\circledast D$ are provable. Take $t_\circledast := r_\circledast \cdot_\circledast s_\circledast$ and use $\circledast$-application, which is an axiom for $\circledast = i$ and for $\circledast = \mathsf{C}$ or follows from Lemma 7.1 for $\circledast = \mathsf{E}$.

For $A = [c]_{\mathsf{C_H}} E \in \mathcal{CS}$ derived by axiom necessitation, take

$$t_{\mathsf{C_G}} := \downarrow^{\mathsf{H}}_{\mathsf{G}} !_{\mathsf{C_H}} c, \qquad t_i := \downarrow^{\mathsf{H}}_{i} !_{\mathsf{C_H}} c, \qquad t_{\mathsf{E_G}} := \mathsf{tail}_{\mathsf{G}}(\downarrow^{\mathsf{H}}_{\mathsf{G}} c)$$

and, as previously, use $\mathsf{C_G}$-inspection and restriction, $\mathsf{C_H}$-inspection and i-conversion, or the co-closure axiom and restriction respectively. Note that in this last case A has to have the form specified above due to the homogeneity of the constant specification. □

As usual, we obtain the following immediate corollary.

Corollary 7.8 (Constructive necessitation). *Let $\mathsf{L}_h^{\mathsf{C}}$ be any of the logics defined above, $\mathcal{CS}$ a homogeneous C-axiomatically appropriate constant specification. For any formula A, if $\vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} A$, then for each $\circledast$, there is a ground term (i.e. a term containing no variables) $t \in \mathrm{Tm}_\circledast$ such that $\vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} [t]_\circledast A$.*

Using constructive necessitation, we can now show that common knowledge inherits the negative inspection property from its base logic given that also reflexivity is present. As for Lemma 7.4, a direct proof could be spelled out. However, such a proof would be very lengthy and cumbersome.

Lemma 7.9. *For any logic $\mathsf{L}_h^{\mathsf{C}}$ containing (jt) and (j5) and let G be a group of agents, $\mathcal{CS}$ a C-axiomatically appropriate and A be a formula. For any $t \in \mathrm{Tm}_{\mathsf{C_G}}^{\mathsf{L}_h^{\mathsf{C}}}$, there is a term $?_{\mathsf{C_G}} t \in \mathrm{Tm}_{\mathsf{C_G}}^{\mathsf{L}_h^{\mathsf{C}}}$ such that $\vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} \neg [t]_{\mathsf{C_G}} A \to [?_{\mathsf{C_G}} t]_{\mathsf{C_G}} \neg [t]_{\mathsf{C_G}} A$. ($\mathsf{C_G}$-negative inspection)*

Proof. Let $i \in \mathsf{G}$. By 4 from Lemma 7.1 we have

$$[t]_{\mathsf{C_G}} A \to \left[\uparrow_i^{\mathsf{G}} t\right]_i [t]_{\mathsf{C_G}} A$$

and the contraposition of this gives

$$\neg \left[\uparrow_i^{\mathsf{G}} t\right]_i [t]_{\mathsf{C_G}} A \to \neg [t]_{\mathsf{C_G}} A.$$

Using constructive necessitation (Corollary 7.8) we find a term $r_i \in \mathrm{Tm}_i$ such that

$$[r_i]_i (\neg \left[\uparrow_i^{\mathsf{G}} t\right]_i [t]_{\mathsf{C_G}} A \to \neg [t]_{\mathsf{C_G}} A).$$

Using the application axiom, we obtain

$$\left[?_i \uparrow_i^{\mathsf{G}} t\right]_i \neg \left[\uparrow_i^{\mathsf{G}} t\right]_i [t]_{\mathsf{C_G}} A \to \left[r_i \cdot ?_i \uparrow_i^{\mathsf{G}} t\right]_i \neg [t]_{\mathsf{C_G}} A.$$

Again, taking the contrapositive we get

$$\neg \left[r_i \cdot ?_i \uparrow_i^{\mathsf{G}} t\right]_i \neg [t]_{\mathsf{C_G}} A \to \neg \left[?_i \uparrow_i^{\mathsf{G}} t\right]_i \neg \left[\uparrow_i^{\mathsf{G}} t\right]_i [t]_{\mathsf{C_G}} A. \qquad (7.2)$$

It is easy to see that

$$\neg \left[?_i \uparrow_i^{\mathsf{G}} t\right]_i \neg \left[\uparrow_i^{\mathsf{G}} t\right]_i [t]_{\mathsf{C_G}} A \to \left[\uparrow_i^{\mathsf{G}} t\right]_i [t]_{\mathsf{C_G}} A \qquad (7.3)$$

as this is simply the contrapositive of an instance of the negative inspection axiom (j5).

Also, the following is an instance of the reflexivity axiom (jt)

$$\left[\uparrow_i^{\mathsf{G}} t\right]_i [t]_{\mathsf{C_G}} A \to [t]_{\mathsf{C_G}} A. \tag{7.4}$$

Combining (7.2), (7.3) and (7.4) we obtain

$$\neg \left[r_i \cdot ?_i \uparrow_i^{\mathsf{G}} t\right]_i \neg [t]_{\mathsf{C_G}} A \to [t]_{\mathsf{C_G}} A.$$

The contrapositive of which gives us

$$\neg [t]_{\mathsf{C_G}} A \to \left[r_i \cdot ?_i \uparrow_i^{\mathsf{G}} t\right]_i \neg [t]_{\mathsf{C_G}} A.$$

As $i \in \mathsf{G} = \{i_1, \ldots, i_k\}$ was arbitrary, using propositional reasoning and the tupling axiom, we get

$$\neg [t]_{\mathsf{C_G}} A \to \left[\left\langle r_{i_1} \cdot ?_{i_1} \uparrow_{i_1}^{\mathsf{G}} t, \ldots, r_{i_k} \cdot ?_{i_k} \uparrow_{i_k}^{\mathsf{G}} t\right\rangle_{\mathsf{G}}\right]_{\mathsf{E_G}} \neg [t]_{\mathsf{C_G}} A. \tag{7.5}$$

Let

$$p(t) := \left\langle r_{i_1} \cdot ?_{i_1} \uparrow_{i_1}^{\mathsf{G}} t, \ldots, r_{i_k} \cdot ?_{i_k} \uparrow_{i_k}^{\mathsf{G}} t\right\rangle_{\mathsf{G}}.$$

Again, applying constructive necessitation (Corollary 7.8) gives us a term $s \in \mathrm{Tm}_{\mathsf{C_G}}$ such that

$$[s]_{\mathsf{C_G}} (\neg [t]_{\mathsf{C_G}} A \to [p(t)]_{\mathsf{E_G}} \neg [t]_{\mathsf{C_G}} A).$$

Now we can use the induction axiom to obtain

$$[p(t)]_{\mathsf{E_G}} \neg [t]_{\mathsf{C_G}} A \to [\mathsf{ind}_{\mathsf{G}}(s, p(t))]_{\mathsf{C_G}} \neg [t]_{\mathsf{C_G}} A. \tag{7.6}$$

Finally, using (7.5) and (7.6) we get

$$\neg [t]_{\mathsf{C_G}} A \to [\mathsf{ind}_{\mathsf{G}}(s, p(t))]_{\mathsf{C_G}} \neg [t]_{\mathsf{C_G}} A.$$

Therefore, we can set $?_{\mathsf{C_G}} t := \mathsf{ind}_{\mathsf{G}}(s, p(t))$. □

Remark 7.10. Note that reflexivity is necessary in the previous proof. If the relations for all agents are Euclidean, the transitive closure of the union is not necessarily Euclidean, as can be shown by an easy counter-example, see Figure 7.1 and, as is to be expected, we will show in the next section that these frame conditions correspond to the axioms mentioned.

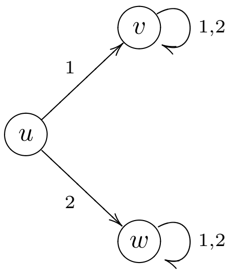

Figure 7.1.: A simple two-agent counterexample showing that the transitive closure of the union of Euclidean accessibility relations is not necessarily Euclidean.

As in the previous chapter, we also obtain internalized induction rules corresponding to (I-R2) and (I-R1) from Chapter 3.

Lemma 7.11 (Internalized induction rule 2). *Let $\mathsf{L}_h^{\mathsf{C}}$ be any of the logics defined above, $\mathcal{CS}$ a homogeneous* C*-axiomatically appropriate constant specification. For any formula A, if $\vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} A \to [r]_{\mathsf{E}_\mathsf{G}} A$, there is a term $t \in \mathrm{Tm}_{\mathsf{C}_\mathsf{G}}$ such that $\vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} [s]_{\mathsf{E}_\mathsf{G}} A \to [\mathsf{ind}_\mathsf{G}(t,s)]_{\mathsf{C}_\mathsf{G}} A$.*

Proof. By constructive necessitation, there exists a term $t \in \mathrm{Tm}_{\mathsf{C}_\mathsf{G}}$ such that $\vdash_{\mathcal{CS}} [t]_{\mathsf{C}_\mathsf{G}}(A \to [r]_{\mathsf{E}_\mathsf{G}} A)$. It remains to use the induction axiom and propositional reasoning. □

Lemma 7.12 (Internalized induction rule 1). *Let $\mathsf{L}_h^{\mathsf{C}}$ be any of the logics defined above, $\mathcal{CS}$ a homogeneous* C*-axiomatically appropriate constant specification and* G *any group of agents. For any formulae A and B, if we have*

$$\vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} B \to [s]_{\mathsf{E}_\mathsf{G}}(A \wedge B) \enspace ,$$

then there exist a term $t \in \mathrm{Tm}_{\mathsf{C}_\mathsf{G}}$ and a constant $c \in \mathrm{Tm}_{\mathsf{C}_\mathsf{H}}$ such that

$$\vdash_{\mathcal{CS}} B \to \left[\downarrow^{\mathsf{H}}_{\mathsf{G}} c \cdot_{\mathsf{C}_\mathsf{G}} \mathsf{ind}_\mathsf{G}(t,s)\right]_{\mathsf{C}_\mathsf{G}} A \enspace ,$$

where $[c]_{\mathsf{C}_\mathsf{H}}(A \wedge B \to A) \in \mathcal{CS}$ and $\mathsf{H} = \{1, \ldots, h\}$.

Proof. Assume

$$\vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} B \to [s]_{\mathsf{E}_\mathsf{G}}(A \land B) \ . \tag{7.7}$$

From this we immediately get $\vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} A \land B \to [s]_{\mathsf{E}_\mathsf{G}}(A \land B)$. Thus, by Lemma 7.11, there is a $t \in \mathrm{Tm}_{\mathsf{C}_\mathsf{G}}$ with

$$\vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} [s]_{\mathsf{E}_\mathsf{G}}(A \land B) \to [\mathsf{ind}_\mathsf{G}(t,s)]_{\mathsf{C}_\mathsf{G}}(A \land B) \ . \tag{7.8}$$

Since $\mathcal{CS}$ is C-axiomatically appropriate, there is a constant $c \in \mathrm{Tm}_\mathsf{C}$ such that

$$\vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} [c]_{\mathsf{C}_\mathsf{H}}(A \land B \to A) \tag{7.9}$$

and thus

$$\vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} \left[\downarrow_\mathsf{G}^\mathsf{H} c\right]_{\mathsf{C}_\mathsf{G}}(A \land B \to A) \ . \tag{7.10}$$

Making use of C_G-application, we find by (7.8) and (7.10) that

$$\vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} [s]_{\mathsf{E}_\mathsf{G}}(A \land B) \to \left[\downarrow_\mathsf{G}^\mathsf{H} c \cdot_{\mathsf{C}_\mathsf{G}} \mathsf{ind}_\mathsf{G}(t,s)\right]_{\mathsf{C}_\mathsf{G}}(A) \ . \tag{7.11}$$

This, together with (7.7), finally yields

$$\vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} B \to \left[\downarrow_\mathsf{G}^\mathsf{H} c \cdot_{\mathsf{C}_\mathsf{G}} \mathsf{ind}_\mathsf{G}(t,s)\right]_{\mathsf{C}_\mathsf{G}}(A) \ .$$

□

7.3. Soundness and Completeness

A major change with respect to the previous chapter concerns models. We generalize the notion of epistemic models to deal with groups of agents in the obvious way. We will call such models *weak*. However, if the base logic contains negative inspection, we need a more restrictive notion of models, called *strong models* in order to obtain completeness. Such models have the strong evidence property stating that admissible evidence of a statement already implies knowledge of that statement. Soundness and completeness with respect to the class of these models is then obtained in the usual way using induction on the derivation and a canonical model construction.

Definition 7.13. A *weak (epistemic) model for logic* L_h^C *meeting a constant specification* $\mathcal{CS}$ is a structure $\mathcal{M} = (W, R, \mathcal{E}, \nu)$, where (W, R, ν)

(jt)	for all $w \in W$ we have $(w, w) \in R_i$;	(reflexivity)
(j4)	for all $w, v, u \in W$, if $(w, v) \in R_i$ and $(v, u) \in R_i$ then $(w, u) \in R_i$;	(transitivity)

Table 7.3.: Frame conditions corresponding to axioms

(jd)	$\bot \notin \mathcal{E}_i(w, t)$ for all $w \in W$ and $t \in \mathrm{Tm}_i$	(consistency)
(j4)	$\mathcal{E}_i(w, t) \subseteq \mathcal{E}_i(v, t)$ whenever $(w, v) \in R_i$;	(monotonicity)
	if $A \in \mathcal{E}_i(w, t)$, then $[t]_i A \in \mathcal{E}_i(w, !t)$;	(inspection)
(j5)	if $A \notin \mathcal{E}_i(w, t)$, then $\neg [t]_i A \in \mathcal{E}_i(w, ?t)$;	(negative inspection)

Table 7.4.: Closure conditions corresponding to axioms

is a Kripke model for the corresponding modal logic with a *set of possible worlds* $W \neq \varnothing$, with a function $R\colon \{1,\dots,h\} \to \mathcal{P}(W \times W)$ that assigns a *accessibility relation* on W satisfying the frame conditions corresponding to the axioms of the logic $\mathsf{L}_h^{\mathsf{C}}$ (as given by table 7.3) to each agent $i \in \{1,\dots,h\}$, and with a *truth valuation* $\nu\colon \mathrm{Prop} \to \mathcal{P}(W)$. We always write R_i instead of $R(i)$ and define the accessibility relations for mutual and common knowledge in the standard way: $R_{\mathsf{E}_\mathsf{G}} := \bigcup_{i\in\mathsf{G}} R_i$ and $R_{\mathsf{C}_\mathsf{G}} := \bigcup_{n=1}^{\infty}(R_{\mathsf{E}_\mathsf{G}})^n$, i.e. $R_{\mathsf{C}_\mathsf{G}}$ is the transitive closure of $R_{\mathsf{E}_\mathsf{G}}$.

An *evidence function* $\mathcal{E}\colon W \times \mathrm{Tm}^{\mathsf{L}_h^{\mathsf{C}}} \to \mathcal{P}\left(\mathrm{Fm}_{\mathsf{L}_h^{\mathsf{C}}}\right)$ determines the formulae evidenced by a term at a world. We define

$$\mathcal{E}_{\circledast} := \mathcal{E} \restriction (W \times \mathrm{Tm}_{\circledast}^{\mathsf{L}_h^{\mathsf{C}}}) \ .$$

Note that whenever $A \in \mathcal{E}_{\circledast}(w,t)$, it follows that $t \in \mathrm{Tm}_{\circledast}^{\mathsf{L}_h^{\mathsf{C}}}$. The evidence function $\mathcal{E}$ must satisfy the following closure conditions: for any worlds $w, v \in W$ and groups of agents $\mathsf{H} \subseteq \mathsf{G} = \{i_1,\dots,i_k\}$,

1. if $[c]_{\circledast} A \in \mathcal{CS}$, then $A \in \mathcal{E}_{\circledast}(w,c)$; (constant specification)
2. if $(A \to B) \in \mathcal{E}_*(w,t)$ and $A \in \mathcal{E}_*(w,s)$, then $B \in \mathcal{E}_*(w, t\cdot s)$; (application)
3. $\mathcal{E}_*(w,s) \cup \mathcal{E}_*(w,t) \subseteq \mathcal{E}_*(w, s+t)$; (sum)
4. $\mathcal{E}_{\mathsf{C}_\mathsf{G}}(w,t) \subseteq \mathcal{E}_{\mathsf{C}_\mathsf{G}}(v,t)$ whenever $(w,v) \in R_{\mathsf{C}_\mathsf{G}}$; ($\mathsf{C}$ – monotonicity)
5. if $A \in \mathcal{E}_i(w,t_i)$ for all $i \in \mathsf{G}$, then $A \in \mathcal{E}_{\mathsf{E}}(w, \langle t_{i_1},\dots,t_{i_k}\rangle_{\mathsf{G}})$; (tupling)
6. if $A \in \mathcal{E}_{\mathsf{E}_\mathsf{G}}(w,t)$ and $i \in \mathsf{G}$, then $A \in \mathcal{E}_i(w, \pi_i^{\mathsf{G}} t)$; (projection)
7. if $A \in \mathcal{E}_{\mathsf{C}_\mathsf{G}}(w,t)$, then $A \in \mathcal{E}_{\mathsf{C}_\mathsf{H}}(w, \downarrow_{\mathsf{H}}^{\mathsf{G}} t)$. (restriction)
8. if $A \in \mathcal{E}_{\mathsf{C}_\mathsf{G}}(w,t)$, then $A \in \mathcal{E}_{\mathsf{E}_\mathsf{G}}(w, \mathsf{head}_\mathsf{G}(t))$ and $[t]_{\mathsf{C}_\mathsf{G}} A \in \mathcal{E}_{\mathsf{E}_\mathsf{G}}(w, \mathsf{tail}_\mathsf{G}(t))$; (co-closure)
9. if $A \in \mathcal{E}_{\mathsf{E}_\mathsf{G}}(w,s)$ and $(A \to [r]_{\mathsf{E}_\mathsf{G}} A) \in \mathcal{E}_{\mathsf{C}_\mathsf{G}}(w,t)$, then $A \in \mathcal{E}_{\mathsf{C}_\mathsf{G}}(w, \mathsf{ind}_\mathsf{G}(t,s))$. (induction)
10. the closure conditions corresponding to the axioms (j4) and (j5) as given by table 7.4, if any of them is present in the logic $\mathsf{L}_h^{\mathsf{C}}$.

When the model is clear from the context, we will directly refer to $R_1, \ldots, R_h$, $R_{\mathsf{E_G}}$, $R_{\mathsf{C_G}}$, $\mathcal{E}_1, \ldots, \mathcal{E}_h$, $\mathcal{E}_{\mathsf{E_G}}$, $\mathcal{E}_{\mathsf{C_G}}$, W, and ν.

Definition 7.14. A ternary relation $\mathcal{M}, w \Vdash A$ for *formula A being satisfied at a world $w \in W$ in a model $\mathcal{M} = (W, R, \mathcal{E}, \nu)$* is defined by induction on the structure of the formula A:

1. $\mathcal{M}, w \Vdash P$ if and only if $w \in \nu(P)$;
2. $\Vdash$ behaves classically with respect to the propositional connectives;
3. $\mathcal{M}, w \Vdash [t]_{\circledast} A$ if and only if 1) $A \in \mathcal{E}_{\circledast}(w, t)$ and 2) $\mathcal{M}, v \Vdash A$ for all $v \in W$ with $(w, v) \in R_{\circledast}$.

Definition 7.15. A *strong (epistemic) model for logic $\mathsf{L}_h^{\mathsf{C}}$ meeting a constant specification $\mathcal{CS}$* is a weak epistemic model for logic $\mathsf{L}_h^{\mathsf{C}}$ meeting a constant specification $\mathcal{CS}$ that has the following additional property

- if $A \in \mathcal{E}_i(w, t)$, then $\mathcal{M}, w \Vdash [t]_i A$ (strong evidence)

We write $\mathcal{M} \Vdash A$ if $\mathcal{M}, w \Vdash A$ for all $w \in W$. We write $\Vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} A$ and say that formula A is *valid with respect to $\mathcal{CS}$* if $\mathcal{M} \Vdash A$ for all strong models $\mathcal{M}$ for logic $\mathsf{L}_h^{\mathsf{C}}$ meeting $\mathcal{CS}$.

Lemma 7.16 (Soundness)**.** *Provable formulae are valid, i.e., $\vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} A$ implies $\Vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} A$.*

Proof. Let $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ be an model for $\mathsf{L}_h^{\mathsf{C}}$ meeting $\mathcal{CS}$ and let $w \in W$. We show soundness by induction on the derivation of A. The cases for propositional tautologies, for the application, sum, reflexivity, and positive inspection axioms, and for modus ponens rule are the same as for the single-agent case in [Fit05] and are, therefore, omitted. We show the remaining cases:

(seriality) As we have $\bot \notin \mathcal{E}_i(w, t)$ for all worlds $w \in W$ and terms $t \in \mathrm{Tm}_i$, we immediately get $\mathcal{M}, w \nVdash [t]_i \bot$ and thus $\mathcal{M}, w \Vdash [t]_i \bot \to \bot$.

(negative inspection) Assume $\mathcal{M}, w \Vdash \neg [t]_i A$. So, by the strong evidence condition, $A \notin \mathcal{E}_i(w, t)$. By the negative inspection closure condition we get $\neg [t]_i A \in \mathcal{E}_i(w, ?_i t)$ and thus, again by the strong evidence condition, $\mathcal{M}, w \Vdash [?_i]_i \neg [t]_i A$.

Note that this is the only place where the strong evidence condition is necessary, thus whenever our logic does not contain (j5), we get also the soundness and completeness results for weak models.

(tupling) Assume $\mathcal{M}, w \Vdash [t_i]_i A$ for all $i \in \mathsf{G} = \{i_1, \ldots, i_k\}$. Then for all $i \in \mathsf{G}$, we have 1) $\mathcal{M}, v \Vdash A$ for all $v \in W$ with $(w, v) \in R_i$ and 2) $A \in \mathcal{E}_i(w, t_i)$. So, by the tupling closure condition, $A \in \mathcal{E}_{\mathsf{E_G}}(w, \langle t_{i_1}, \ldots, t_{i_k}\rangle_{\mathsf{G}})$ from 2). Since by definition $R_{\mathsf{E_G}} = \bigcup_{i \in \mathsf{G}} R_i$, it follows from 1) that $\mathcal{M}, v \Vdash A$ for all $v \in W$ with $(w, v) \in R_{\mathsf{E_G}}$. Hence, $\mathcal{M}, w \Vdash \left[\langle t_{i_1}, \ldots, t_{i_k}\rangle_{\mathsf{G}}\right]_{\mathsf{E_G}} A$.

(projection) Assume, $\mathcal{M}, w \Vdash [t]_{\mathsf{E_G}} A$ and $i \in \mathsf{G}$. Then 1) $\mathcal{M}, v \Vdash A$ for all $v \in W$ with $(w, v) \in R_{\mathsf{E_G}}$ and 2) $A \in \mathcal{E}_{\mathsf{E_G}}(w, t)$. By the projection closure condition, it follows from 2) that $A \in \mathcal{E}_i(w, \pi_i^{\mathsf{G}} t)$. In addition, since $R_{\mathsf{E_G}} = \bigcup_{i \in \mathsf{G}} R_i$, we get $\mathcal{M}, v \Vdash A$ for all $v \in W$ with $(w, v) \in R_i$ by 1). Thus, $\mathcal{M}, w \Vdash \left[\pi_i^{\mathsf{G}} t\right]_i A$.

(restriction) Assume $\mathcal{M}, w \Vdash [t]_{\mathsf{C_G}} A$ and $\mathsf{H} \subseteq \mathsf{G}$. Then 1) $\mathcal{M}, v \Vdash A$ for all $v \in W$ with $(w, v) \in R_{\mathsf{C_G}}$ and 2) $A \in \mathcal{E}_{\mathsf{C_G}}(w, t)$. By the restriction closure condition, it follows from 2) that $A \in \mathcal{E}_{\mathsf{C_H}}(w, \downarrow_{\mathsf{H}}^{\mathsf{G}} t)$. Furthermore, since $R_{\mathsf{C_H}} \subseteq R_{\mathsf{C_G}}$ we get $\mathcal{M}, v \Vdash A$ for all $v \in W$ with $(w, v) \in R_{\mathsf{C_H}}$ by 1). Thus, $\mathcal{M}, w \Vdash \left[\downarrow_{\mathsf{H}}^{\mathsf{G}} t\right]_{\mathsf{C_H}} A$.

(co-closure) Assume $\mathcal{M}, w \Vdash [t]_{\mathsf{C_G}} A$. Then 1) $\mathcal{M}, v \Vdash A$ for all $v \in W$ with $(w, v) \in R_{\mathsf{C_G}}$ and 2) $A \in \mathcal{E}_{\mathsf{C_G}}(w, t)$. It follows from 1) that for all $v' \in W$ with $(w, v') \in R_{\mathsf{E_G}}$, we have $\mathcal{M}, v' \Vdash A$ since $R_{\mathsf{E_G}} \subseteq R_{\mathsf{C_G}}$; also, due to the C-monotonicity closure condition, $\mathcal{M}, v' \Vdash [t]_{\mathsf{C_G}} A$ since $R_{\mathsf{E_G}} \circ R_{\mathsf{C_G}} \subseteq R_{\mathsf{C_G}}$. From 2), by the co-closure closure condition, $A \in \mathcal{E}_{\mathsf{E_G}}(w, \mathsf{head_G}(t))$ and $[t]_{\mathsf{C_G}} A \in \mathcal{E}_{\mathsf{E_G}}(w, \mathsf{tail_G}(t))$. Hence, $\mathcal{M}, w \Vdash [\mathsf{head_G}(t)]_{\mathsf{E_G}} A$ and $\mathcal{M}, w \Vdash [\mathsf{tail_G}(t)]_{\mathsf{E_G}} [t]_{\mathsf{C_G}} A$.

(induction) Assume $\mathcal{M}, w \Vdash [t]_{\mathsf{C_G}}(A \to [r]_{\mathsf{E_G}} A)$ and $\mathcal{M}, w \Vdash [s]_{\mathsf{E_G}} A$. The induction closure condition gives us $A \in \mathcal{E}_{\mathsf{C_G}}(w, \mathsf{ind_G}(t, s))$. To show $\mathcal{M}, v \Vdash A$ for all $v \in W$ with $(w, v) \in R_{\mathsf{C_G}}$, we prove that $\mathcal{M}, v \Vdash A$ for all $v \in W$ with $(w, v) \in (R_{\mathsf{E_G}})^n$ by induction on the positive integer n.

The **base case** $n = 1$ immediately follows from $\mathcal{M}, w \Vdash [s]_{\mathsf{E_G}} A$.

Induction step. Let $(w, v') \in (R_{\mathsf{E_G}})^n$ and $(v', v) \in R_{\mathsf{E_G}}$ for some $v, v' \in W$. As $\mathcal{M}, w \Vdash [t]_{\mathsf{C_G}}(A \to [r]_{\mathsf{E_G}} A)$, we have $\mathcal{M}, v' \Vdash A \to$

$[r]_{\mathsf{E_G}} A$ and by induction hypothesis, $\mathcal{M}, v' \Vdash A$, thus, $\mathcal{M}, v' \Vdash [r]_{\mathsf{E_G}} A$, which yields $\mathcal{M}, v \Vdash A$.

Finally, we conclude that $\mathcal{M}, w \Vdash [\mathsf{ind_G}(t, s)]_{\mathsf{C_G}} A$.

(axiom necessitation) Let A be an axiom of $\mathsf{L}_h^{\mathsf{C}}$ and c be a proof constant such that $[c]_{\circledast} A \in \mathcal{CS}$. Since A is an axiom of $\mathsf{L}_h^{\mathsf{C}}$, $\mathcal{M}, w \Vdash A$ for all $w \in W$, as shown above. Also, as $\mathcal{M}$ is a model for $\mathsf{L}_h^{\mathsf{C}}$ meeting $\mathcal{CS}$, we have $A \in \mathcal{E}_{\circledast}(w, c)$ for all $w \in W$ by the constant specification closure condition. Thus, $\mathcal{M}, w \Vdash [c]_{\circledast} A$ for all $w \in W$. □

Definition 7.17. Let $\mathcal{CS}$ be a constant specification. A set Φ of formulae is called $\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})$-*consistent* if $\Phi \nvdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} \phi$ for some formula ϕ. A set Φ is called *maximal* $\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})$-*consistent* if it is $\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})$-consistent and has no $\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})$-consistent proper extensions.

Whenever safe, we do not mention the constant specification and only talk about consistent and maximal consistent sets. It can be easily shown that maximal consistent sets contain all axioms of $\mathsf{L}_h^{\mathsf{C}}$ and are closed under modus ponens.

Definition 7.18. For a set Φ of formulae, we define

$$\Phi/\circledast := \{A \ : \ \text{there is a } t \in \mathrm{Tm}_{\circledast}^{\mathsf{L}_h^{\mathsf{C}}} \text{ such that } [t]_{\circledast} A \in \Phi\} \ .$$

Definition 7.19. Let $\mathcal{CS}$ be a constant specification. The *canonical (epistemic) model* $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ *for* $\mathsf{L}_h^{\mathsf{C}}$ *meeting* $\mathcal{CS}$ is defined as follows:

1. $W := \{w \subseteq \mathrm{Fm}_{\mathsf{L}_h^{\mathsf{C}}} \ : \ w \text{ is a maximal } \mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})\text{-consistent set}\}$;
2. $R_i := \{(w, v) \in W \times W \ : \ w/i \subseteq v\}$;
3. $\mathcal{E}_{\circledast}(w, t) := \{A \in \mathrm{Fm}_{\mathsf{L}_h^{\mathsf{C}}} \ : \ [t]_{\circledast} A \in w\}$;
4. $\nu(P_n) := \{w \in W \ : \ P_n \in w\}$.

Lemma 7.20. *Let $\mathcal{CS}$ be a constant specification for $\mathsf{L}_h^{\mathsf{C}}$. The canonical model for $\mathsf{L}_h^{\mathsf{C}}$ meeting $\mathcal{CS}$ is a weak model for $\mathsf{L}_h^{\mathsf{C}}$ meeting $\mathcal{CS}$.*

Proof. The proof of the frame conditions for each R_i, as well as the argument for the constant specification, application, sum, inspection closure, and monotonicity conditions, is the same as in the single-agent case (see [Fit05]). We show the remaining closure conditions on the evidence function:

(consistency) Is immediate by the fact that our worlds are maximal consistent sets and the definition of the evidence function. (Otherwise there would be $\bot \in \mathcal{E}_i(w,t)$ for some $w \in W$ and $t \in \mathrm{Tm}_i$) and thus also $[t]_i\bot \in w$ which by (jd) and maximal consistency would yield $\bot \in w$ contradicting the consistency of w.)

(negative inspection) Assume $A \notin \mathcal{E}_i(w,t)$. By definition of $\mathcal{E}_i$, we have $[t]_iA \notin w$ and so $\neg\,[t]_iA \in w$ by maximal consistency. Therefore, by the negative inspection axiom and maximal consistency $[?t]_i\neg\,[t]A \in w$. Thus, $\neg\,[t]_iA \in \mathcal{E}_i(w,?t)$.

(tupling) Assume $A \in \mathcal{E}_i(w,t_i)$ for all $i \in \mathsf{G} = \{i_1,\ldots,i_k\}$. By definition of $\mathcal{E}_i$, we have $[t_i]_iA \in w$ for all $i \in \mathsf{G}$. Therefore, by the tupling axiom and maximal consistency, $\left[\langle t_{i_1},\ldots,t_{i_k}\rangle_{\mathsf{G}}\right]_{\mathsf{E_G}}A \in w$. Thus, $A \in \mathcal{E}_{\mathsf{E_G}}(w,\langle t_{i_1},\ldots,t_{i_k}\rangle_{\mathsf{G}})$.

(projection) Assume $A \in \mathcal{E}_{\mathsf{E_G}}(w,t)$ and $i \in \mathsf{G}$. Thus, we have $[t]_{\mathsf{E_G}}A \in w$. Then, by the projection axiom and maximal consistency, $\left[\pi_i^{\mathsf{G}}t\right]_iA \in w$, and thus $A \in \mathcal{E}_i(w,\pi_i^{\mathsf{G}}t)$.

(restriction) Assume $A \in \mathcal{E}_{\mathsf{E_G}}(w,t)$ and $\mathsf{H} \subseteq \mathsf{G}$. By definition of $\mathcal{E}_{\mathsf{E_G}}$, we have $[t]_{\mathsf{E_G}}A \in w$. Therefore, by the restriction axiom and maximal consistency, $\left[\downarrow_{\mathsf{H}}^{\mathsf{G}}t\right]_{\mathsf{E_H}}A \in w$. Thus, $A \in \mathcal{E}_{\mathsf{E_H}}(w,\downarrow_{\mathsf{H}}^{\mathsf{G}}t)$.

(co-closure) Assume $A \in \mathcal{E}_{\mathsf{C_G}}(w,t)$. Thus, $[t]_{\mathsf{C_G}}A \in w$, and, by the co-closure axioms and maximal consistency, $[\mathsf{head}_{\mathsf{G}}(t)]_{\mathsf{E_G}}A \in w$ and $[\mathsf{tail}_{\mathsf{G}}(t)]_{\mathsf{E_G}}\,[t]_{\mathsf{C_G}}A \in w$. Hence, $A \in \mathcal{E}_{\mathsf{E_G}}(w,\mathsf{head}_{\mathsf{G}}(t))$ and $[t]_{\mathsf{C_G}}A \in \mathcal{E}_{\mathsf{E_G}}(w,\mathsf{tail}_{\mathsf{G}}(t))$.

(induction) Assume $A \in \mathcal{E}_{\mathsf{E_G}}(w,s)$ and $(A \to [r]_{\mathsf{E_G}}A) \in \mathcal{E}_{\mathsf{C_G}}(w,t)$. Then we have $[s]_{\mathsf{E_G}}A \in w$ and $[t]_{\mathsf{C_G}}(A \to [r]_{\mathsf{E_G}}A) \in w$. From the induction axiom, it follows by maximal consistency that $[\mathsf{ind}_{\mathsf{G}}(t,s)]_{\mathsf{C_G}}A \in w$. Therefore, $A \in \mathcal{E}_{\mathsf{C_G}}(w,\mathsf{ind}_{\mathsf{G}}(t,s))$.

(C-monotonicity) It is sufficient to prove by induction on the positive integer n that

$$\text{if } [t]_{\mathsf{C_G}}A \in w \text{ and } (w,v) \in (R_{\mathsf{E_G}})^n, \text{ then } [t]_{\mathsf{C_G}}A \in v \ . \quad (7.12)$$

Base case $n = 1$**.** Assume $(w,v) \in R_{\mathsf{E_G}}$, i.e., $w/i \subseteq v$ for some $i \in \mathsf{G}$. As $[t]_{\mathsf{C_G}}A \in w$, $\left[\pi_i^{\mathsf{G}}\mathsf{tail_G}(t)\right]_i [t]_{\mathsf{C_G}}A \in w$ by maximal consistency, and hence $[t]_{\mathsf{C_G}}A \in w/i \subseteq v$. The argument for the **induction step** is similar.

Now assume $(w,v) \in R_{\mathsf{C_G}} = \bigcup_{n=1}^{\infty}(R_{\mathsf{E_G}})^n$ and $A \in \mathcal{E}_{\mathsf{C_G}}(w,t)$, i.e., $[t]_{\mathsf{C_G}}A \in w$. As shown above, $[t]_{\mathsf{C_G}}A \in v$. Thus, $A \in \mathcal{E}_{\mathsf{C_G}}(v,t)$. □

Lemma 7.21 (Truth Lemma)**.** *Let $\mathcal{CS}$ be a constant specification and $\mathcal{M}$ be the canonical model for $\mathsf{L}_h^{\mathsf{C}}$ meeting $\mathcal{CS}$. For all formulae A and all worlds $w \in W$,*

$$A \in w \textit{ if and only if } \mathcal{M}, w \Vdash A \ .$$

Proof. The proof is by induction on the structure of A. The cases for propositional variables and propositional connectives are immediate by the definition of $\Vdash$ and by the maximal consistency of w. We check the remaining cases:

Case A is $[t]_iB$. Assume $A \in w$. Then $B \in w/i$ and $B \in \mathcal{E}_i(w,t)$. Consider any v such that $(w,v) \in R_i$. Since $w/i \subseteq v$, it follows that $B \in v$, and thus, by induction hypothesis, $\mathcal{M}, v \Vdash B$. And $\mathcal{M}, w \Vdash A$ immediately follows from this.

For the converse, assume $\mathcal{M}, w \Vdash [t]_iB$. By definition of $\Vdash$ we get $B \in \mathcal{E}_i(w,t)$, from which $[t]_iB \in w$ immediately follows by definition of $\mathcal{E}_i$.

Case A is $[t]_{\mathsf{E_G}}B$. Assume $A \in w$ and consider any v such that $(w,v) \in R_{\mathsf{E_G}}$. Then $(w,v) \in R_i$ for some $i \in \mathsf{G}$, i.e., $w/i \subseteq v$. By definition of $\mathcal{E}_{\mathsf{E_G}}$, we get $B \in \mathcal{E}_{\mathsf{E_G}}(w,t)$. By maximal consistency of w, it follows that $\left[\pi_i^{\mathsf{G}}t\right]_iB \in w$, and thus $B \in w/i \subseteq v$. Since, by induction hypothesis, $\mathcal{M}, v \Vdash B$, we conclude that $\mathcal{M}, w \Vdash A$. The argument for the converse repeats the one from the previous case.

Case A is $[t]_{\mathsf{C_G}}B$. Assume $A \in w$ and consider any v such that $(w,v) \in R_{\mathsf{C_G}}$, i.e., $(w,v) \in (R_{\mathsf{E_G}})^n$ for some $n \geq 1$. As in the previous cases,

$B \in \mathcal{E}_{\mathsf{C_G}}(w,t)$ by definition of $\mathcal{E}_{\mathsf{C_G}}$. As $(w,v) \in (R_{\mathsf{E_G}})^n$, there is a $v' \in W$ such that $(w,v') \in (R_{\mathsf{E_G}})^{n-1}$ and $(v',v) \in R_i$ for some $i \in \mathsf{G}$. By (7.12) we find $A \in v'$, and thus, by the co-closure and projection axioms and maximal consistency, also $\left[\pi_i^{\mathsf{G}}\mathsf{head}_{\mathsf{G}}(t)\right]_i B \in v'$ and so $B \in v$. Hence, by the induction hypothesis $\mathcal{M}, v \Vdash B$. Now $\mathcal{M}, w \Vdash A$ immediately follows. The argument for the converse repeats the one from the previous cases. □

Note that the converse directions in the proof above are far from trivial in the modal case, see e.g. [MH95]. The last case, in particular, usually requires more sophisticated methods that guarantee the finiteness of the model.

Corollary 7.22. *Let $\mathcal{CS}$ be a constant specification for $\mathsf{L}_h^{\mathsf{C}}$. The canonical model $\mathcal{M}$ for $\mathsf{L}_h^{\mathsf{C}}$ meeting $\mathcal{CS}$ is a strong model.*

Proof. Suppose $A \in \mathcal{E}_i(w,t)$. By definition, $[t]_i A \in w$, and so by the Truth Lemma 7.21 $\mathcal{M}, w \Vdash [t]_i A$. □

Remark 7.23. By the same argument we also get

$$\text{if } A \in \mathcal{E}_{\mathsf{E_G}}(w,t), \quad \text{then } \mathcal{M}, w \Vdash [t]_{\mathsf{E_G}} A,$$
$$\text{if } A \in \mathcal{E}_{\mathsf{C_G}}(w,t), \quad \text{then } \mathcal{M}, w \Vdash [t]_{\mathsf{C_G}} A,$$

i.e., the canonical model is also "strong with respect to the relations $R_{\mathsf{E_G}}$ and $R_{\mathsf{C_G}}$".

Theorem 7.24 (Completeness)**.** *Let $\mathcal{CS}$ be a constant specification for $\mathsf{L}_h^{\mathsf{C}}$. Then $\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})$ is sound and complete with respect to the class of strong models for $\mathsf{L}_h^{\mathsf{C}}$ meeting $\mathcal{CS}$, i.e., for all formulae $A \in \mathrm{Fm}_{\mathsf{L}_h^{\mathsf{C}}}$,*

$$\vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} A \text{ if and only if } \Vdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} A \enspace .$$

Proof. Soundness has already been shown in Lemma 7.16. For completeness, let $\mathcal{M}$ be the canonical model for $\mathsf{L}_h^{\mathsf{C}}$ meeting $\mathcal{CS}$ and assume $\nvdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} A$. Then $\{\neg A\}$ is $\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})$-consistent and hence is contained in some maximal $\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})$-consistent set $w \in W$. So, by Lemma 7.21, $\mathcal{M}, w \Vdash \neg A$, and hence, by Lemma 7.20, $\nVdash_{\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})} A$. □

Furthermore, for certain logics we also get two more completeness results with respect to different classes of models.

Corollary 7.25. *If $\mathsf{L}_h^{\mathsf{C}}$ contains (j5), then $\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})$ is sound and complete with respect to the class of strong models for $\mathsf{L}_h^{\mathsf{C}}$ meeting $\mathcal{CS}$ with the additional properties*

1. *if $A \notin \mathcal{E}_i(w,t)$ and $(w,v) \in R_i$, then $A \notin \mathcal{E}_i(v,t)$* (*Anti-Monotonicity*)
2. *for all $w, v, u \in W$, if $(w,v) \in R_i$ and $(w,u) \in R_i$, then $(v,u) \in R_i$* (*Euclideanness*)

Proof. It is sufficient to show that the canonical model for $\mathsf{L}_h^{\mathsf{C}}$ has the desired properties.

1. Assume $A \notin \mathcal{E}_i(w,t)$ and $(w,v) \in R_i$. By negative inspection closure we obtain $\neg [t]_i \in \mathcal{E}(w, ?t)$. By the strong evidence property we get $\mathcal{M}, w \Vdash [?t]_i \neg [t]_i A$ and thus $\mathcal{M}, v \Vdash \neg [t]_i A$ and so $\mathcal{M}, v \nVdash [t]_i A$. Thus, again by the strong evidence property, $A \notin \mathcal{E}_i(w,t)$.
2. Let $(w,v) \in R_i$ and $(w,u) \in R_i$. For any $[t]_i A \in v$ we have $\mathcal{M}, v \Vdash [t]_i A$ by the Truth Lemma 7.21 and so, by strong evidence $A \in \mathcal{E}_i(v,t)$. By anti-monotonicity from the previous part we obtain $A \in \mathcal{E}_i(w,t)$. Again, by strong evidence we get $\mathcal{M}, w \Vdash [t]_i A$ and thus $\mathcal{M}, u \Vdash A$. So, we have established $w/i \subseteq u$ and we are finished. □

Corollary 7.26. *Let $\mathcal{CS}$ be a C-axiomatically appropriate constant specification and $\mathsf{L}_h^{\mathsf{C}}$ contain the axiom (jd). Then $\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})$ is sound and complete with respect to the class of strong Models for $\mathsf{L}_h^{\mathsf{C}}$ meeting $\mathcal{CS}$ lacking the consistency closure condition but with the additional property*

- *for all $w \in W$ there is a $v \in W$ such that $(w,v) \in R_i$.* (*seriality*)

Proof. For soundness we need to show $\mathcal{M}, w \nVdash [t]_i \bot$. By seriality there must be a world $v \in W$ with $(w,v) \in R_i$ and there, trivially $\mathcal{M}, v \nVdash \bot$ and we immediately get the desired result.

For completeness it is sufficient to show that the canonical model has the desired frame property. Let $w \in W$. We need to show that there is a $v \in W$ such that $(w,v) \in R_i$. It is sufficient to show that w/i is consistent, because then w/i is contained in some maximal consistent set v and by definition of R_i we get $(w,v) \in R_i$. So, assume (towards a contradiction) that w/i is not consistent. This means, there must be formulae $[s_1]_i A_1, \ldots, [s_n]_i A_n \in w$ such that $A_1, \ldots, A_n \vdash \bot$. By Lemma 7.7 we get

$$[x_1]_i A_1, \ldots, [x_n]_i A_n \vdash [t(x_1, \ldots, x_n)]_i \bot$$

for fresh variables x_i and some term t. Replacing the x_i by s_i yields

$$[s_1]_i A_1, \ldots, [s_n]_i A_n \vdash [t'(s_1, \ldots, s_n)]_i \bot$$

where t' is the same as t except that some justification constants are (possibly) replaced by other justification constants. But, as we are dealing with logics containing (jd), this immediately gives

$$[s_1]_i A_1, \ldots, [s_n]_i A_n \vdash \bot$$

contradicting the consistency of w. □

7.4. Finite Model Property and Decidability

The decidability proof from the previous chapter and Chapter A can also be adapted to the general case in this chapter with the exception of the case for logics containing negative inspection, as discussed in Section A.7.

Definition 7.27 (M-models)**.** An *M-model* is a singleton model.

We will denote an M-model by $\mathcal{M} = (\mathcal{E}, \nu)$ instead of $(\{w\}, R, \mathcal{E}, \nu)$ and we will write

- $\nu(P_j) = \texttt{true}$ if and only if $w \in \nu(P_j)$,
- $A \in \mathcal{E}_{\circledast}(t)$ if and only if $A \in \mathcal{E}_{\circledast}(w, t)$.

Definition 7.28 (Validity in M-models)**.** Validity in an M-model $\mathcal{M} = (\mathcal{E}, \nu)$ for $\mathsf{L}_h^{\mathsf{C}}$ is inductively defined as follows:

- $\mathcal{M} \Vdash P_j$ if and only if $\nu(P_j) = \texttt{true}$,
- $\Vdash$ behaves classically with respect to propositional connectives,
- – if $\mathsf{L}_h^{\mathsf{C}}$ does not contain (jt), then

$$\mathcal{M} \Vdash [t]_{\circledast} A \text{ if and only if } A \in \mathcal{E}_{\circledast}(t) \ ,$$

– if $\mathsf{L}_h^{\mathsf{C}}$ does contain (jt), then

$$\mathcal{M} \Vdash [t]_{\circledast} A \text{ if and only if } \mathcal{M} \Vdash A \text{ and } A \in \mathcal{E}_{\circledast}(t) \ .$$

Theorem 7.29 (Completeness with respect to M-models)**.** *Let $\mathcal{CS}$ be a constant specification. $\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})$ is also sound and complete with respect to the class of M-models meeting $\mathcal{CS}$.*

Proof. Soundness follows immediately from Lemma 7.16. Now assume that $\nvdash_{\mathcal{CS}} A$, then $\{\neg A\}$ is $\mathcal{CS}$-consistent, and hence $\mathcal{M}, w \Vdash \neg A$ for some world $w_0 \in W$ in the canonical model $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ meeting $\mathcal{CS}$.

Let $\mathcal{M}' = (\mathcal{E}', \nu')$ be the restriction of $\mathcal{M}$ to $\{w_0\}$, i.e., for $W' := \{w_0\}$, let $\mathcal{E}' := \mathcal{E} \restriction (W' \times \mathrm{Tm})$ for any $\circledast$, and $\nu'(P_n) := \nu(P_n) \cap W'$.

Since $\mathcal{M}'$ is clearly an M-model meeting $\mathcal{CS}$, it remains to demonstrate that $\mathcal{M}' \Vdash B$ if and only if $\mathcal{M}, w_0 \Vdash B$ for all formulae B. We proceed by induction on the structure of B. The cases where either B is a propositional variable or its primary connective is propositional are trivial. Therefore, we only show the case of $B = [t]_{\circledast} C$. First, observe that

$$\mathcal{M}, w_0 \Vdash [t]_{\circledast} C \text{ if and only if } C \in \mathcal{E}'_{\circledast}(t) \ . \tag{7.13}$$

Indeed, by Lemma 7.21, $\mathcal{M}, w_0 \Vdash [t]_{\circledast} C$ if and only if $[t]_{\circledast} C \in w_0$, which, by definition of the canonical model, is equivalent to $C \in \mathcal{E}_{\circledast}(w_0, t) = \mathcal{E}'_{\circledast}(t)$.

We have to consider two cases.

- Let $\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})$ not contain (jt). If $\mathcal{M}, w_0 \Vdash [t]_{\circledast} C$, then $C \in \mathcal{E}'_{\circledast}(t)$ by (7.13) and thus $\mathcal{M}' \Vdash C$.

  If $\mathcal{M}, w_0 \nVdash [t]_{\circledast} C$ then by (7.13) we have $C \notin \mathcal{E}'_{\circledast}(t)$, so $\mathcal{M}' \nVdash [t]_{\circledast} C$.

- Let $\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})$ contain (jt). If $\mathcal{M}, w_0 \Vdash [t]_{\circledast} C$, then $\mathcal{M}, w_0 \Vdash C$ since $R_{\circledast}$ is reflexive.

  By induction hypothesis, $\mathcal{M}' \Vdash C$. By (7.13) we have $C \in \mathcal{E}'_{\circledast}(t)$, and thus $\mathcal{M}' \Vdash [t]_{\circledast} C$.

  If $\mathcal{M}, w_0 \nVdash [t]_{\circledast} C$, then by (7.13) we have $C \notin \mathcal{E}'_{\circledast}(t)$, so $\mathcal{M}' \nVdash [t]_{\circledast} C$. □

Corollary 7.30 (Finite model property). *Any of the previously defined logics $\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS})$ enjoy the finite model property with respect to epistemic models.*

Using techniques from Chapter A, we get decidability as stated in Theorem A.30. The decidability question for logics containing the (5) axiom remains open, see the discussion in Chapter A.

Theorem 7.31.

1. *Any justification logic*

$$\mathsf{L}_h^{\mathsf{C}} \in \{\mathsf{J}_h^{\mathsf{C}}(\mathcal{CS}), \mathsf{JT}_h^{\mathsf{C}}(\mathcal{CS}), \mathsf{J4}_h^{\mathsf{C}}(\mathcal{CS}), \mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS})\}$$

   with a decidable schematic $\mathcal{CS}$ is decidable.

2. *Any justification logic*

$$\mathsf{L}_h^{\mathsf{C}} \in \{\mathsf{JD}_h^{\mathsf{C}}(\mathcal{CS}), \mathsf{JD4}_h^{\mathsf{C}}(\mathcal{CS})\}$$

   with a decidable, schematic and axiomatically appropriate $\mathcal{CS}$ is decidable.

7.5. Conservativity

The conservativity proof from the previous chapter can also be adapated for logics containing the reflexivity axiom. However, the case for logics not containing the reflexivity axiom remains open, see the discussion at the end of this section.

Definition 7.32. The language of L_h is obtained from that of $\mathsf{L}_h^{\mathsf{C}}$ by restricting the set of operations to $\cdot_i$, $+_i$, and $!_i$ and by dropping all terms from $\mathrm{Tm}_{\mathsf{E_G}}$ and $\mathrm{Tm}_{\mathsf{C_G}}$ for all groups of agents G. The axioms are restricted to application, sum, and, if present in the base logic L, reflexivity, inspection, and negative inspection for each i. The definition of constant specification is changed accordingly.

We show that any $\mathsf{L}_h^{\mathsf{C}}$ containing the reflexivity axiom is conservative over its corresponding multi-agent logic L_h without common knowledge by adapting a technique from [Fit08b].

Definition 7.33. The mapping $.^{\times} : \mathrm{Fm}_{\mathsf{L}_h^{\mathsf{C}}} \to \mathrm{Fm}_{\mathsf{L}_h}$ is defined as follows:

1. $P^{\times} := P$ for propositional variables $P \in \mathrm{Prop}$;
2. $.^{\times}$ commutes with propositional connectives;
3. $$([t]_{\circledast} A)^{\times} := \begin{cases} A^{\times} & \text{if } t \text{ contains a subterm } s \in \mathrm{Tm}_{\mathsf{E}_\mathsf{G}} \\ & \text{or } s \in \mathrm{Tm}_{\mathsf{C}_\mathsf{G}} \text{ for any group} \\ & \text{of agents } \mathsf{G}, \\ [t]_{\circledast} A^{\times} & \text{otherwise.} \end{cases}$$

For a set $\Phi \subseteq \mathrm{Fm}_{\mathsf{L}_h^{\mathsf{C}}}$ we define $\Phi^{\times} := \{\varphi^{\times} \ : \ \varphi \in \Phi\}$.

Theorem 7.34. *Let $\mathsf{L}_h^{\mathsf{C}}$ be a logic containing the reflexivity axiom and $\mathcal{CS}$ a constant specification for $\mathsf{L}_h^{\mathsf{C}}$. For an arbitrary formula $A \in \mathrm{Fm}_{\mathsf{L}_h}$, if $\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS}) \vdash A$, then $\mathsf{L}_h(\mathcal{CS}^{\times}) \vdash A$.*

Proof. Since $A^{\times} = A$ for any $A \in \mathrm{Fm}_{\mathsf{L}_h}$, it suffices to demonstrate that for any formula $D \in \mathrm{Fm}_{\mathsf{L}_h^{\mathsf{C}}}$, if $\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS}) \vdash D$, then $\mathsf{L}_h(\mathcal{CS}^{\times}) \vdash D^{\times}$, which can be done by induction on the derivation of D.
Case when D is a propositional tautology, then so is $D^{\times}$.
Case when $D = [t]_*(B \to C) \to ([s]_* B \to [t \cdot s]_* C)$ is an instance of the application axiom. We distinguish the following possibilities:

1. Both t and s contain a subterm from $\mathrm{Tm}_{\mathsf{E}_\mathsf{G}} \cup \mathrm{Tm}_{\mathsf{C}_\mathsf{G}}$ for some group of agents G. Then $D^{\times}$ has the form $(B^{\times} \to C^{\times}) \to (B^{\times} \to C^{\times})$, which is a propositional tautology and, thus, an axiom of L_h.

2. Neither t nor s contains a subterm from $\mathrm{Tm}_{\mathsf{E}_\mathsf{G}} \cup \mathrm{Tm}_{\mathsf{C}_\mathsf{G}}$ for all groups of agents G. Then $D^{\times}$ is an instance of the application axiom of L_h.

3. Term t contains a subterm from $\mathrm{Tm}_{\mathsf{E}_\mathsf{G}} \cup \mathrm{Tm}_{\mathsf{C}_\mathsf{G}}$ for some group of agents G while s does not. Then $D^{\times}$ is $(B^{\times} \to C^{\times}) \to ([s]_i B^{\times} \to C^{\times})$, which can be derived in $\mathsf{L}_h(\mathcal{CS}^{\times})$ from the reflexivity axiom $[s]_i B^{\times} \to B^{\times}$ by propositional reasoning. In this case, translation $\times$ does not map an axiom of $\mathsf{L}_h^{\mathsf{C}}$ to an axiom of L_h.

4. Term s contains a subterm from $\mathrm{Tm}_{\mathsf{E}_\mathsf{G}} \cup \mathrm{Tm}_{\mathsf{C}_\mathsf{G}}$ for some group of agents G while t does not. Then $D^{\times}$ is $[t]_i(B^{\times} \to C^{\times}) \to (B^{\times} \to C^{\times})$, an instance of the reflexivity axiom of L_h.

Case when $D = [t]_* B \to [t+s]_* B$ is an instance of the sum axiom. Then $D^\times$ becomes $B^\times \to B^\times$, $[t]_i B^\times \to B^\times$, or $[t]_i B^\times \to [t+s]_i B^\times$, i.e., a propositional tautology, an instance of the reflexivity axiom of L_h, or an instance of the sum axiom of L_h respectively. The sum axiom $[s]_* B \to [t+s]_* B$ is treated in the same manner.
Case when $D = [t_{i_1}]_{i_1} B \wedge \ldots \wedge [t_{i_k}]_{i_k} B \to \left[\langle t_{i_1}, \ldots, t_{i_k} \rangle_{\mathsf{G}}\right]_{\mathsf{E_G}} B$ is an instance of the tupling axiom for some group of agents $\mathsf{G} = \{i_1, \ldots, i_k\}$. We distinguish the following possibilities:

1. At least one of the t_i's contains a subterm from $\mathrm{Tm}_{\mathsf{E_H}} \cup \mathrm{Tm}_{\mathsf{C_H}}$ for some group of agents H. Then $D^\times$ has the form $C_1 \wedge \ldots \wedge C_k \to B^\times$ with at least one $C_i = B^\times$ and is, therefore, a propositional tautology.

2. None of the t_i's contains a subterm from $\mathrm{Tm}_{\mathsf{E_H}} \cup \mathrm{Tm}_{\mathsf{C_H}}$ for all groups of agents H. Then $D^\times$ has the form $[t_{i_1}]_{i_1} B^\times \wedge \ldots \wedge [t_{i_k}]_{i_k} B^\times \to B^\times$, which can be derived in $\mathsf{L}_h(\mathcal{CS}^\times)$ from the reflexivity axiom. This is another case when translation $\times$ does not map an axiom of $\mathsf{L}_h^{\mathsf{C}}$ to an axiom of L_h.

Case when D is an instance of the projection axiom $[t]_{\mathsf{E_G}} B \to \left[\pi_i^{\mathsf{G}} t\right]_i B$, the restriction axiom $[t]_{\mathsf{C_G}} A \to \left[\downarrow_{\mathsf{H}}^{\mathsf{G}} t\right]_{\mathsf{C_G}} A$ or of the co-closure axioms, i.e., $[t]_{\mathsf{C_G}} B \to [\mathsf{head}_{\mathsf{G}}(t)]_{\mathsf{E_G}} B$ or $[t]_{\mathsf{C_G}} B \to [\mathsf{tail}_{\mathsf{G}}(t)]_{\mathsf{E_G}} [t]_{\mathsf{C_G}} B$ for some group of agents G with $i \in \mathsf{G}$ and $\mathsf{H} \subseteq \mathsf{G}$. Then $D^\times$ is the propositional tautology $B^\times \to B^\times$.
Case when $D = [t]_{\mathsf{C_G}}(B \to [r]_{\mathsf{E_G}} B) \to ([s]_{\mathsf{E_G}} B \to [\mathsf{ind}_{\mathsf{G}}(t,s)]_{\mathsf{C}} B$ is an instance of the induction axiom. Then $D^\times$ is $(B^\times \to B^\times) \to (B^\times \to B^\times)$, a propositional tautology.
Case when $D = [t]_i B \to B$ is an instance of the reflexivity axiom (jt). Then $D^\times$ is either $[t]_i B^\times \to B^\times$ or $B^\times \to B^\times$, i.e., an instance of the reflexivity axiom of L_h or a propositional tautology respectively.
Case when $D = [t]_i B \to [!t]_i\, [t]_i B$ is an instance of the inspection axiom (j4). Then $D^\times$ is either the propositional tautology $B^\times \to B^\times$ or $[t]_i B^\times \to [!t]_i\, [t]_i B^\times$, an instance of the inspection axiom of L_h.
Case when $D = \neg\, [t]_i B \to [?t]_i \neg\, [t]_i B$ is an instance of the negative inspection axiom (j5). The $D^\times$ is either the propositional tautology $\neg B^\times \to \neg B^\times$ or $\neg\, [t]_i B^\times \to [?t]_i \neg\, [t]_i B^\times$, an instance of the negative inspection axiom of L_h.
Case when D is derived by modus ponens is trivial.

Case when D is $[c]_{\circledast}B \in \mathcal{CS}$. Then $D^{\times}$ is either $B^{\times}$ or $[c]_i B^{\times}$. In the former case, B is an axiom of $\mathsf{L}_h^{\mathsf{C}}$, and hence $B^{\times}$ is derivable in $\mathsf{L}_h(\mathcal{CS}^{\times})$, as shown above; in the latter case, $[c]_i B^{\times} \in \mathcal{CS}^{\times}$. □

Note that as previously $\mathcal{CS}^{\times}$ need not, in general, be a constant specification for L_h as discussed in Remark 6.27.

The case for the logics without the reflexivity axiom remains open. Note that the semantical method from [Mil09] can be adapted to the logics presented here: Given a model for the logic without common knowledge, we can extend this model to a model for the logic with common knowledge using the fact that the closure conditions on the evidence functions define operators (see Chapter A). However, we would start with a "common knowledge-free" constant specification and the process of extension would not add anything to this constant specification. Thus, the conservativity result would be rather weak. Hence, it remains to find a mapping similar to the mapping $.^{\times}$ defined above in order to also extend the constant specification in a "natural" way.

Also, note that $\mathsf{JD4}$ is not conservative over JD as shown in [Mil09], as it can be easily shown that $\{x : P,\, y : (x : P \to \bot)\}$ is consistent in JD, but not in $\mathsf{JD4}$ (as in the latter you can derive $y {\cdot} !x : \bot$). In some sense, adding common knowledge is similar to adding positive inspection due to the fact that common knowledge is the *transitive* closure of the union of the agents' accessibility relations. A similar effect might occur for JD_h and $\mathsf{JD}_h^{\mathsf{C}}$. However, the typedness of the evidence terms seems to prevent this.

7.6. Forgetful Projection and a Word on Realization

As is to be expected, we can also easily generalize the results on forgetful projection from the previous chapter, but the opposite direciton, the realization theorem, remains open.

Let $\mathrm{Fm}_{\Box}$ denote the set of formulae in the language of the (modal) logics of common knowledge from Chapter 3, Section 3.5.

Definition 7.35 (Forgetful projection)**.** The mapping $\circ\colon \mathrm{Fm}_{\mathsf{L}_h^{\mathsf{C}}} \to \mathrm{Fm}_{\Box}$ is defined as follows:

Logic $\mathsf{L}_h^{\mathsf{C}}$	Corresponding modal logic $(\mathsf{L}_h^{\mathsf{C}})^\circ$
$\mathsf{J}_h^{\mathsf{C}}$	$\mathsf{K}_h^{\mathsf{C}}$
$\mathsf{JD}_h^{\mathsf{C}}$	$\mathsf{KD}_h^{\mathsf{C}}$
$\mathsf{JD4}_h^{\mathsf{C}}$	$\mathsf{KD4}_h^{\mathsf{C}}$
$\mathsf{JD5}_h^{\mathsf{C}}$	$\mathsf{KD5}_h^{\mathsf{C}}$
$\mathsf{JD45}_h^{\mathsf{C}}$	$\mathsf{KD45}_h^{\mathsf{C}}$
$\mathsf{JT}_h^{\mathsf{C}}$	$\mathsf{KT}_h^{\mathsf{C}}$
$\mathsf{JT4}_h^{\mathsf{C}}$	$\mathsf{KT4}_h^{\mathsf{C}}$
$\mathsf{JT5}_h^{\mathsf{C}}$	$\mathsf{KT5}_h^{\mathsf{C}}$
$\mathsf{JT45}_h^{\mathsf{C}}$	$\mathsf{KT45}_h^{\mathsf{C}}$
$\mathsf{J4}_h^{\mathsf{C}}$	$\mathsf{K4}_h^{\mathsf{C}}$
$\mathsf{J45}_h^{\mathsf{C}}$	$\mathsf{K45}_h^{\mathsf{C}}$
$\mathsf{J5}_h^{\mathsf{C}}$	$\mathsf{K5}_h^{\mathsf{C}}$

Table 7.5.: Forgetful projection

1. $P^\circ := P$ for propositional variables $P \in \mathrm{Prop}$;
2. $\circ$ commutes with propositional connectives;
3. $([t]_i A)^\circ := \Box_i A^\circ$;
4. $([t]_{\mathsf{E_G}} A)^\circ := \mathsf{E_G} A^\circ$;
5. $([t]_{\mathsf{C_G}} A)^\circ := \mathsf{C_G} A^\circ$.

For a given justification logic $\mathsf{L}_h^{\mathsf{C}}$ we will call $(\mathsf{L}_h^{\mathsf{C}})^\circ$ *the correspondig modal logic* as given by table 7.6 (see Chapter 6, Section 6.6 for a definition of these logics).

Lemma 7.36. *Let $\mathcal{CS}$ be any constant specification. For any formula $A \in \mathrm{Fm}_{\mathsf{L}_h^{\mathsf{C}}}$, if $\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS}) \vdash A$, then $(\mathsf{L}_h^{\mathsf{C}})^\circ \vdash A^\circ$.*

Proof. The proof is an easy induction on the derivation of A. ☐

Definition 7.37 (Realization)**.** A realization is a mapping

$$r\colon \mathrm{Fm}_\Box \to \mathrm{Fm}_{\mathsf{L}_h^{\mathsf{C}}}$$

such that $(r(A))^\circ = A$. We usually write A^r instead of $r(A)$.

As before, the problem of realization for a given homogeneous C-axiomatically appropriate constant specification $\mathcal{CS}$ can be stated as:

> Is there a realization r such that $\mathsf{L}_h^{\mathsf{C}}(\mathcal{CS}) \vdash A^r$ for any theorem A of the corresponding modal logic $(\mathsf{L}_h^{\mathsf{C}})^\circ$?

The problem of realization remains open. The comments from Chapter 6, Section 6.6, especially concerning realization, apply here as well.

8. Conclusions

> Let not sleep fall upon thy eyes till thou has thrice reviewed the transactions of the past day. Where have I turned aside from rectitude? What have I been doing? What have I left undone, which I ought to have done? Begin thus from the first act, and proceed; and, in conclusion, at the ill which thou hast done, be troubled, and rejoice for the good.
>
> *Samuel Johnson's translation of Pythagoras's Golden Verses 40–44,* The Rambler No. 8 (14 April 1750) *[Joh10]*

In Chapter 4, we presented a proof system S for common knowledge that uses the general idea that common knowledge as a greatest fixed point can be unfolded "infinitely often". This leads to finitely branching proof trees with infinitely long branches that must contain infinitely many applications of the unfolding of the same common knowledge formula. Similar systems were studied for the modal μ-calculus and linear time logic [NW96; Stu08; DHL06].

It is interesting to investigate the relationship between the system presented and the systems presented in the previous chapter, e.g. H_{Ax}, but these would require syntactic cut elimination in particular. Also, this could shed a new light on how common knowledge emerges. Another important observation is, that the system presented in [Weh10] seems to be a finitized version of the co-inductive system presentend here, using annotations in order to check possible repetitions of sequents when fixed points are unfolded. However, a formal relationship betweens these system has yet to be established. For this purpose, but also as a very general question, it is of course of particular interest, to learn more about the behavior and form of common knowledge formulae in

C-threads.

In Chapters 6 and 7, we presented several justification logics with common knowledge (in particular our "example" logic $\mathsf{LP}_h^{\mathsf{C}}$). The major open problem at the moment remains proving the realization theorem, one direction of which we have demonstrated. See Section 6.6 and Chapter C for a discussion of this problem and possible approaches. Further open problems include the conservativity of the logics without the reflexivity axiom over the corresponding multi-agent justification logics as well as the decidability of logics containing the negative introspection axiom.

Our analysis of the coordinated attack problem in the language of $\mathsf{LP}_h^{\mathsf{C}}$ shows that access to evidence creates more alternatives than the classical modal approach. In particular, the lack of knowledge can occur either because messages are not delivered, or because evidence of authenticity is missing.

We have mostly focused on the study of C-axiomatically appropriate constant specifications. For modeling distributed systems with different reasoning capabilities of agents, it is also interesting to consider i-axiomatic appropriate, E-axiomatic appropriate, and heterogeneous constant specifications, where only certain aspects of reasoning are common knowledge.

We established soundness and completeness with respect to epistemic models and singleton M-models. The question, whether other semantics for justification logics such as (arithmetical) provability semantics [Art95; Art01] and game semantics [Ren09b] can be adapted to $\mathsf{LP}_h^{\mathsf{C}}$, remains. Using the filtration techniques described in Chapter A, we have also been able to show the decidability of the logics presented, except for the logics that include negative inspection.[1] Further avenues of research might include, but are not limited to, the comparison of the complexity of the justification logics with common knowledge to the complexity of the corresponding modal logics with common knowledge.

A long-term goal of our research is to find justification counterparts of dynamic epistemic logics with common knowledge, see Chapter B. Clearly, both types of systems, explicit counterparts to common knowledge logics and to dynamic epistemic logics, will have to be studied on their own first, before being combined.

[1]See Section A.7 for a discussion of negative introspection.

Yet another direction of research might be justification logics that are released from their “modal chains”, i.e., justification logics that are not related to modal logics. Justification logics with common knowledge with a more fine grained control of reasoning capabilities as mentioned above might be a very simple example for such logics. Another possibility might be an exploration of the evidential dynamics of announcements, e.g. the announced formula itself might be a justification for its announcement. In this vicinity it might also be interesting to consider the relationship of Fitting models and neighborhood semantics, i.e., also the relationship of justification logics and classical modal logics.

It also seems worthwhile to have adequate model theoretic tools at hand, such as e.g. filtrations from Chapter A. Besides the usual modal logic suspects such as bounded morphisms, bisimulations, etc., this would also include techniques to deal with the evidence function in the spirit of those presented in [Fit09]. As a simple use case of such tools we could think of non-definability results in the spirit of [BRV02]. Furthermore, a better model theoretic understanding might also help in overcoming the problems with the semantical realization approach as described in Chapter C.

A. Filtrations

In Chapters 6 and 7 we stated the decidability of some justification logics with common knowledge. In this chapter, we investigate the techniques used to obtain these results as presented in [BKS12a] and use them to prove the decidability results for justification logics with common knowledge.

Filtrations are a tool in modal logic for obtaining from a given, usually infinite, model a smaller, usually finite, model by factoring the set of worlds with respect to a certain equivalence relation. As noted in [BRV02], filtrations were first introduced in [Seg71] and given their name in [LS77]. Given the close relationship between Fitting models and Kripke models, it is a natural task to adopt filtrations for justification logics. The crucial step is, of course, to take into account the evidence relation when identifying states.

Filtrations are often used to prove a finite model property and thereby establish decidability of a given modal logic, see e.g. [BRV02]. Decidability for the justification logics presented here was originally shown in [Kuz00; Kuz08; Mkr97]. We adapt the filtration technique from modal logic to obtain an alternative uniform proof of decidability for these justification logics. We then apply the newly developed technique to establish the decidability of the multi-agent justification logic with common knowledge presented in Chapters 6 and 7.

First, we will introduce some additional definitions relevant to this chapter. In Section A.2, we define filtrations for justification logics and prove their basic properties. We treat two specific examples of filtrations in Sections A.3 and A.4. In Section A.5, we use these two examples to prove the decidability of the justification logics defined in Chapter 5. This also leads us to investigate general properties necessary for the decidability of justification logics and enables us to prove the decidability of the multi-agent justification logics with common knowledge from Chapters 6 and 7 in Section A.6.

A.1. Preliminary Definitions

The set of subformulae $\mathrm{Sub}(A)$ of a given formula A is defined inductively as follows:

$$\begin{aligned}
\mathrm{Sub}(p_i) &:= \{p_i\},\\
\mathrm{Sub}(\neg A) &:= \{\neg A\} \cup \mathrm{Sub}(A),\\
\mathrm{Sub}(A_1 \to A_2) &:= \{A_1 \to A_2\} \cup \mathrm{Sub}(A_1) \cup \mathrm{Sub}(A_2),\\
\mathrm{Sub}(t : A) &:= \{t : A\} \cup \mathrm{Sub}(A).
\end{aligned}$$

A set of formulae $\Phi \subseteq \mathrm{Fm}$ is *closed under subformulae* if

$$\bigcup_{F \in \Phi} \mathrm{Sub}(F) \subseteq \Phi \quad .$$

Definition A.1 (t-evidence relation)**.** Suppose we drop the monotonicity condition (6)

$$\text{if } (t, A, w) \in \mathcal{E} \text{ and } wRv, \text{ then } (t, A, v) \in \mathcal{E} \tag{6}$$

from Definition 5.5. A relation satisfying the remaining closure conditions (for logics containing the (4) axiom) is called a *t-evidence relation*.

Definition A.2 (Evidence bases)**.** 1. An *evidence base* $\mathcal{B}$ is a subset of $\mathrm{Tm} \times \mathrm{Fm} \times W$.

2. An evidence relation $\mathcal{E}$ is *based on* $\mathcal{B}$, if $\mathcal{B} \subseteq \mathcal{E}$.

Remark A.3. The closure conditions in the definition of an admissible evidence function give rise to a monotone operator. The minimal evidence relation based on $\mathcal{B}$ is the least fixed point of that operator and thus always exists.

A.2. Filtrations

Given the close relationship of models for justification logics to Kripke models, it is not surprising that the two definitions of filtrations look very similar. The major difference is that we have to take the evidence relation into consideration. In modal logic, we identify worlds that behave the same way, whereas in justification logic we identify worlds that behave the same way *for the same reason*.

Definition A.4 (Filtration)**.** Let $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ be a model and Φ some set of formulae that is closed under subformulae. We define an equivalence relation $=_\Phi \subseteq W \times W$ by setting $w =_\Phi v$ if and only if for all $A \in \Phi$

$$M, w \Vdash A \text{ if and only if } M, v \Vdash A$$

and for all $t : B \in \Phi$

$$\mathcal{E}(t, B, w) \text{ if and only if } \mathcal{E}(t, B, v).$$

We denote the equivalence classes of Φ by $[w]_\Phi$. When Φ is clear from the context, we will often only write $[w]$ instead of $[w]_\Phi$.

A model $\mathcal{M}_\Phi = (W_\Phi, R_\Phi, \mathcal{E}_\Phi, \nu_\Phi)$ is called a *filtration of* $\mathcal{M}$ *through* Φ if it satisfies the following:

1. $W_\Phi = \{[w]_\Phi \mid w \in W\}$,
2. R_Φ satisfies

   (R1) for all $w, v \in W$ if $R(w, v)$, then $R_\Phi([w]_\Phi, [v]_\Phi)$,

   (R2) for all $[w]_\Phi, [v]_\Phi \in W_\Phi$, if $R_\Phi([w]_\Phi, [v]_\Phi)$, then for any $t : B \in \Phi$ we have

   $$\text{if } \mathcal{M}, w \Vdash t : B \text{ then } \mathcal{M}, v \Vdash B,$$

3. $\mathcal{E}_\Phi$ satisfies

   (E1) for all $w \in W$ and $t : B \in \Phi$ we have

   $$\text{if } \mathcal{M}, w \Vdash t : B \text{ then } (t, B, [w]_\Phi) \in \mathcal{E}_\Phi,$$

   (E2) for all $w \in W$ and $t : B \in \Phi$ we have

   $$\text{if } (t, B, [w]_\Phi) \in \mathcal{E}_\Phi \text{ then } (t, B, w) \in \mathcal{E},$$

4. ν_Φ satisfies for all atomic propositions $p \in \Phi$

   $$\nu_\Phi(p) = \{[w]_\Phi \mid w \in \nu(p)\} \ .$$

There are two major changes in the definition compared to the case for modal logic. The first change concerns the definition of the equivalence relation to identify worlds. Whereas a modal formula $\Box B$ can only fail due to the existence of an accessible world not satisfying B, a justification formula $t : B$ might fail in two ways: either B is not satisfied in an accessible world or t is not admissible evidence for B at the current world. So we have to refine our equivalence relation to only identify worlds that do not only satisfy the same formulae but also behave the same with respect to the evidence relation. The second change concerns the evidence relation of the filtration: it has to satisfy conditions similar to the Min- and Max-conditions (R1) and (R2) for the accessibility relation.

The crucial property of a filtration of a model through Φ is that the behavior of the model and the filtration is the same with respect to formulae in Φ:

Lemma A.5. *Let $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ be a model, Φ a set of formulae closed under subformulae, and $\mathcal{M}_\Phi = (W_\Phi, R_\Phi, \mathcal{E}_\Phi, \nu_\Phi)$ a filtration of $\mathcal{M}$ through Φ. Then for all worlds $w \in W$ and formulae $A \in \Phi$ we have*

$$\mathcal{M}_\Phi, [w]_\Phi \Vdash A \textit{ if and only if } \mathcal{M}, w \Vdash A.$$

Proof. The proof is by induction on the structure of A. The case for propositional variables is immediate by the definition of ν_Φ and the cases for the propositional connectives are immediate by the induction hypothesis. Let us now consider the case $A = t : B$.

First we show the direction from right to left. Assume $\mathcal{M}, w \Vdash t : B$. If $R_\Phi([w]_\Phi, [v]_\Phi)$, then by (R2) we have $\mathcal{M}, v \Vdash B$. By the induction hypothesis we get $\mathcal{M}_\Phi, [v]_\Phi \Vdash B$. Further from (E1) we get $(t, B, [w]_\Phi) \in \mathcal{E}_\Phi$ and thus $\mathcal{M}_\Phi, [w]_\Phi \Vdash t : B$.

For the other direction suppose $\mathcal{M}_\Phi, [w]_\Phi \Vdash t : B$, that is

$$\mathcal{M}_\Phi, [v]_\Phi \Vdash B \text{ for all } [v]_\Phi \text{ with } R_\Phi([w]_\Phi, [v]_\Phi) \tag{A.1}$$

$$(t, B, [w]_\Phi) \in \mathcal{E}_\Phi \tag{A.2}$$

If $R(w, v)$, then by (R1) also $R_\Phi([w]_\Phi, [v]_\Phi)$ and by (A.1) and the induction hypothesis we get $\mathcal{M}, v \Vdash B$. Furthermore from (A.2) and (E2) we get $\mathcal{E}(t, B, w)$ and we conclude $\mathcal{M}, w \Vdash t : B$. $\Box$

A filtration inherits some conditions on the accessibility relations. Furthermore, a filtration through a finite set has finitely many worlds.

Lemma A.6. *Let $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ be a model, Φ a set of formulae closed under subformulae, and $\mathcal{M}_\Phi = (W_\Phi, R_\Phi, \mathcal{E}_\Phi, \nu_\Phi)$ a filtration of $\mathcal{M}$ through Φ.*

1. *If R is serial, so is R_Φ.*
2. *If R is reflexive, so is R_Φ.*
3. *If Φ is finite, then so is W_Φ.*

Proof. The first two claims follow immediately from (R1). The last claim follows from the fact that each element $[w]_\Phi \in W_\Phi$ can be characterized by the set of formulae $A \in \Phi$ that hold in $[w]_\Phi$ as well as the set of formulae $t : B \in \Phi$ with $\mathcal{E}_\Phi(t, B, [w]_\Phi)$ and the fact that $\mathcal{P}(\Phi) \times \mathcal{P}(\Phi)$ has only finitely many elements. □

A.3. Non-transitive Case

As a first example we will define filtrations for logics not containing the (4) axiom.

Definition A.7. Let $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ be a model and Φ a set of formulae closed under subformulae. We consider the structure $\mathcal{M}^{\mathsf{nt}}_\Phi = (W^{\mathsf{nt}}_\Phi, R^{\mathsf{nt}}_\Phi, \mathcal{E}^{\mathsf{nt}}_\Phi, \nu^{\mathsf{nt}}_\Phi)$ that is given by

1. W^{nt}_Φ is the set of equivalence classes induced by $=_\Phi$,
2. $R^{\mathsf{nt}}_\Phi([w], [v])$ iff for all $t : B \in \Phi$ we have $\mathcal{M}, w \Vdash t : B$ implies $\mathcal{M}, v \Vdash B$,
3. $\mathcal{E}^{\mathsf{nt}}_\Phi$ is the minimal evidence relation based on $\mathcal{B}^{\mathsf{nt}}_\Phi$, where

$$\mathcal{B}^{\mathsf{nt}}_\Phi(t, B, [v]) \text{ if and only if } t : B \in \Phi \text{ and } \mathcal{E}(t, B, v)$$

4. ν^{nt}_Φ is given by

$$\nu^{\mathsf{nt}}_\Phi(p) = \begin{cases} \{[w] \mid w \in \nu(p)\} & \text{if } p \in \Phi, \\ \emptyset & \text{otherwise.} \end{cases}$$

Lemma A.8. *$\mathcal{M}^{\mathsf{nt}}_{\Phi}$ is a filtration of $\mathcal{M}$ through Φ.*

Proof. We have to check the following conditions.

(R1) Assume $R(w, v)$. If $\mathcal{M}, w \Vdash t : B$, then $\mathcal{M}, v \Vdash B$. Thus we conclude $R^{\mathsf{nt}}_{\Phi}([w], [v])$.

(R2) Let $t : B \in \Phi$ and $R^{\mathsf{nt}}_{\Phi}([w], [v])$. If $\mathcal{M}, w \Vdash t : B$, then we get $\mathcal{M}, v \Vdash B$ immediately from the definition of R^{nt}_{Φ}.

(E1) Assume $t : B \in \Phi$ and $\mathcal{M}, w \Vdash t : B$. We have $\mathcal{E}(t, B, w)$ and we immediately get $\mathcal{E}^{\mathsf{nt}}_{\Phi}(t, B, [w])$ by the definition of $\mathcal{E}^{\mathsf{nt}}_{\Phi}$.

(E2) We show for all $t : B$, not only for those contained in Φ, that for all $w' \in [w]$

$$\mathcal{E}^{\mathsf{nt}}_{\Phi}(t, B, [w]) \text{ implies } \mathcal{E}(t, B, w') \ .$$

We proceed by induction on the construction of $\mathcal{E}^{\mathsf{nt}}_{\Phi}$.

- If $\mathcal{E}^{\mathsf{nt}}_{\Phi}(t, B, [w])$ because $\mathcal{B}^{\mathsf{nt}}_{\Phi}(t, B, [w])$, then by definition of $\mathcal{B}^{\mathsf{nt}}_{\Phi}$ we have that $t : B \in \Phi$ and $\mathcal{E}(t, B, w'')$ for some $w'' \in [w]$. By $w' =_{\Phi} w''$ we conclude $\mathcal{E}(t, B, w')$.
- If $t = t_1 + t_2$ and $\mathcal{E}^{\mathsf{nt}}_{\Phi}(t, B, [w])$ because of $\mathcal{E}^{\mathsf{nt}}_{\Phi}(t_i, B, [w])$ (for some $i \in \{1, 2\}$), then by induction hypothesis we get that $\mathcal{E}(t_i, B, w')$ and thus also $\mathcal{E}(t_1 + t_2, B, w')$ by the closure conditions on $\mathcal{E}$.
- If $t = t_1 \cdot t_2$ and $\mathcal{E}^{\mathsf{nt}}_{\Phi}(t, B, [w])$ because there is an $A \in \mathrm{Fm}$ such that $\mathcal{E}^{\mathsf{nt}}_{\Phi}(t_1, A \to B, [w])$ and $\mathcal{E}^{\mathsf{nt}}_{\Phi}(t_2, A, [w])$, then by induction hypothesis $\mathcal{E}(t_1, A \to B, w')$ and $\mathcal{E}(t_2, A, w')$. So, by the closure conditions, we get $\mathcal{E}(t_1 \cdot t_2, B, w')$.
- The case for axiom necessitation is trivial, as we have

$$\mathcal{E}(\underbrace{! \cdots !!}_{n} c, \underbrace{! \cdots !!}_{n-1} c : \cdots : !c : c : A, v)$$

for any world $v \in W$. □

A.4. Transitive Case

The case for logics containing the (4) axiom is a bit more involved, as we now have to ensure that the accessibility relation of the filtration has to be transitive as well.

Definition A.9. Let $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ be a model and Φ a set of formulae closed under subformulae. We consider the structure $\mathcal{M}^{\mathsf{tr}}_{\Phi} = (W^{\mathsf{tr}}_{\Phi}, R^{\mathsf{tr}}_{\Phi}, \mathcal{E}^{\mathsf{tr}}_{\Phi}, \nu^{\mathsf{tr}}_{\Phi})$ that is given by

1. W^{tr}_{Φ} is the set of equivalence classes induced by $=_{\Phi}$,
2. $R^{\mathsf{tr}}_{\Phi}([w],[v])$ iff for all $t : B \in \Phi$ we have $\mathcal{M}, w \Vdash t : B$ implies $\mathcal{M}, v \Vdash B \wedge t : B$,
3. $\mathcal{E}^{\mathsf{tr}}_{\Phi}$ is the minimal t-evidence relation based on $\mathcal{B}^{\mathsf{tr}}_{\Phi}$, where
$$\mathcal{B}^{\mathsf{tr}}_{\Phi}(t, B, [v]) \text{ if and only if } t : B \in \Phi \text{ and } \mathcal{M}, v \Vdash t : B,$$
4. ν^{tr}_{Φ} is given by
$$\nu^{\mathsf{tr}}_{\Phi}(p) = \begin{cases} \{[w] \mid w \in \nu(p)\} & \text{if } p \in \Phi, \\ \emptyset & \text{otherwise.} \end{cases}$$

As a first step we have to show that $\mathcal{E}^{\mathsf{tr}}_{\Phi}$ as defined is not only a t-evidence relation but an actual evidence relation.

Lemma A.10. *$\mathcal{E}^{tr}_{\Phi}$ is an admissible evidence relation over $(W^{tr}_{\Phi}, R^{tr}_{\Phi})$.*

Proof. We have to show that condition (6) in Definition 5.5 holds, i.e., we have to show

$$\mathcal{E}^{\mathsf{tr}}_{\Phi}(t, B, [w]) \text{ and } R^{\mathsf{tr}}_{\Phi}([w],[v]) \text{ imply } \mathcal{E}^{\mathsf{tr}}_{\Phi}(t, B, [v]) \ .$$

So assume $\mathcal{E}^{\mathsf{tr}}_{\Phi}(t, B, [w])$ and $R^{\mathsf{tr}}_{\Phi}([w],[v])$. We now show $\mathcal{E}^{\mathsf{tr}}_{\Phi}(t, B, [v])$ by induction on the construction of $\mathcal{E}^{\mathsf{tr}}_{\Phi}$.

Let $\mathcal{E}^{\mathsf{tr}}_{\Phi}(t, B, [w])$ because of $\mathcal{B}^{\mathsf{tr}}_{\Phi}(t, B, [w])$. We have $t : B \in \Phi$ and $\mathcal{M}, w \Vdash t : B$ by definition of $\mathcal{B}^{\mathsf{tr}}_{\Phi}$. Since $R_{\Phi}([w],[v])$, it follows that $\mathcal{M}, v \Vdash B \wedge t : B$ and, in particular, $\mathcal{M}, v \Vdash t : B$. Thus, $\mathcal{B}^{\mathsf{tr}}_{\Phi}(t, B, [v])$ by definition of $\mathcal{B}^{\mathsf{tr}}_{\Phi}$, and clearly $\mathcal{E}^{\mathsf{tr}}_{\Phi}(t, B, [v])$.

Let us now distinguish the different possible closure conditions from Definition 5.5:

1. Assume we have $t = t_1 + t_2$ and $\mathcal{E}^{\mathsf{tr}}_{\Phi}(t, B, [w])$ because of $\mathcal{E}^{\mathsf{tr}}_{\Phi}(t_i, B, [w])$ for $i = 1$ or $i = 2$. Then by induction hypothesis $\mathcal{E}^{\mathsf{tr}}_{\Phi}(t_i, B, [v])$ and thus also $\mathcal{E}^{\mathsf{tr}}_{\Phi}(t, B, [v])$.

2. The case for $\cdot$ and ! follows immediately from the induction hypothesis in the same manner as the previous case.

3. The case for axiom necessitation (AN) trivially holds. □

The accessibility relation for the filtration is transitive.

Lemma A.11. *R^{tr}_{Φ} is transitive.*

Proof. Assume (a) $R^{\mathsf{tr}}_{\Phi}([w], [v])$ and (b) $R^{\mathsf{tr}}_{\Phi}([v], [u])$. Suppose $t : B \in \Phi$ and $\mathcal{M}, w \Vdash t : B$. By (a) we get $\mathcal{M}, v \Vdash t : B$. Then by (b) we obtain

$$\mathcal{M}, u \Vdash B \wedge t : B.$$

Hence we conclude $R^{\mathsf{tr}}_{\Phi}([w], [u])$. □

Lemma A.12. *$\mathcal{M}^{\mathsf{tr}}_{\Phi}$ is a filtration of $\mathcal{M}$ through Φ.*

Proof. We have to check the following conditions.

(R1) Assume $R(w, v)$. If $\mathcal{M}, w \Vdash t : B$, then $\mathcal{M}, v \Vdash B$ and $\mathcal{M}, w \Vdash !t : t : B$ which implies $\mathcal{M}, v \Vdash t : B$. Thus we conclude $R^{\mathsf{tr}}_{\Phi}([w], [v])$.

(R2) Let $t : B \in \Phi$ and $R_{\Phi}([w], [v])$. If $\mathcal{M}, w \Vdash t : B$, then we get $\mathcal{M}, v \Vdash B$ immediately from the definition of R^{tr}_{Φ}.

(E1) Assume $t : B \in \Phi$ and $\mathcal{M}, w \Vdash t : B$. We immediately get $\mathcal{E}^{\mathsf{tr}}_{\Phi}(t, B, [w])$ by the definition of $\mathcal{E}^{\mathsf{tr}}_{\Phi}$.

(E2) As in the proof of Lemma A.8 we can show for all $t : B$ and all $w' \in [w]$

$$\mathcal{E}^{\mathsf{tr}}_{\Phi}(t, B, [w]) \text{ implies } \mathcal{E}(t, B, w') \ .$$

□

A.5. Decidability

The theorems in this section originate from [Kuz08]. We will thus only give proof sketches for the sake of brevity.

Definition A.13 (Finitary model)**.** A model $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ is called *finitary* if

1. W is finite,
2. there exists a finite base $\mathcal{B}$ such that $\mathcal{E}$ is the minimal evidence relation based on $\mathcal{B}$, and
3. the set $\{(w, p) \in W \times \text{Prop} \mid w \in \nu(p)\}$ is finite.

Using filtrations we see that if a formula is satisfiable then it is satisfiable in a finitary model. Thus we have the following:

Lemma A.14 (Completeness w.r.t. finitary models)**.**

1. *Let* $\mathsf{L}_{\mathcal{CS}} \in \{\mathsf{J}_{\mathcal{CS}}, \mathsf{JT}_{\mathcal{CS}}, \mathsf{J4}_{\mathcal{CS}}, \mathsf{LP}_{\mathcal{CS}}\}$ *and* $\mathcal{CS}$ *be a constant specification for* L*. If a formula* A *is not derivable in* $\mathsf{L}_{\mathcal{CS}}$*, then there exists a finitary model* $\mathcal{M}$ *for* $\mathsf{L}_{\mathcal{CS}}$ *with* $\mathcal{M}, w \nVdash A$ *for some world* w *in* $\mathcal{M}$*.*
2. *Let* $\mathsf{L}_{\mathcal{CS}} \in \{\mathsf{JD}_{\mathcal{CS}}, \mathsf{JD4}_{\mathcal{CS}}\}$ *and* $\mathcal{CS}$ *be an axiomatically appropriate constant specification for* L*. If a formula* A *is not derivable in* $\mathsf{L}_{\mathcal{CS}}$*, then there exists a finitary model* $\mathcal{M}$ *for* $\mathsf{L}_{\mathcal{CS}}$ *with* $\mathcal{M}, w \nVdash A$ *for some world* w *in* $\mathcal{M}$*.*

Proof. Let $\mathcal{CS}$ be as required above. If A is not derivable in $\mathsf{L}_{\mathcal{CS}}$, then by Theorem 5.11 there exists a model $\mathcal{M}$ for $\mathsf{L}_{\mathcal{CS}}$ with $\mathcal{M}, v \nVdash A$ for some world v in $\mathcal{M}$. Now set $\Phi := \text{Sub}(A)$ and let $\mathcal{M}_\Phi$ denote either $\mathcal{M}_\Phi^{\mathsf{nt}}$ or $\mathcal{M}_\Phi^{\mathsf{tr}}$ from Definitions A.7 and A.9 respectively, depending on whether $\mathsf{L}_{\mathcal{CS}}$ contains the (4) axiom. It is easy to see that $\mathcal{M}_\Phi$ is a finitary model: by Lemma A.6 the set of worlds is finite and, by definition of $\mathcal{M}_\Phi$, the evidence relation is finitely based and the valuation function satisfies condition 3 from Definition A.13. Finally, since $\mathcal{M}_\Phi$ is a filtration of $\mathcal{M}$ through Φ by Lemma A.8 or by Lemma A.12, by Lemma A.5 we have $\mathcal{M}_\Phi, [v] \nVdash A$. □

Corollary A.15. *All statements of Lemma A.14 hold if an additional restriction is imposed that the domain of the model $\mathcal{M}$ be a finite subset of $\mathbb{N}$.*

Proof. The claim follows trivially from Lemma A.14 by renaming worlds to natural numbers. □

The following theorem is a simple instance of Post's theorem [Pos44]: A set is decidable if and only if both the set and its complement are recursively enumerable.

Theorem A.16. *A logic is decidable if it is recursively enumerable and is sound and complete with respect to a set $\mathcal{C}$ such that*

1. *$\mathcal{C}$ is a recursively enumerable set of finite models and*
2. *the relation $\mathcal{M}, w \Vdash A$ between models $\mathcal{M} \in \mathcal{C}$, worlds w in $\mathcal{M}$, and formulae A is decidable.*

Proof. We give a proof sketch, for full details cf. [Kuz08, Theorem 4.3.3]

Given a formula A, we can simultaneously enumerate theorems $B_0, B_1, \ldots$ of the logic and potential counter-models $\mathcal{M}_0, \mathcal{M}_1, \ldots \in \mathcal{C}$ and at each step check whether (a) $A = B_i$ or (b) $\mathcal{M}_i, w \nVdash A$ for some $w \in \mathcal{M}_i$. Eventually either (a) or (b) will hold for some i, thus indicating whether the logic proves A. □

Lemma A.17. *Let $\mathsf{L}_{\mathcal{CS}} \in \{\mathsf{J}_{\mathcal{CS}}, \mathsf{JD}_{\mathcal{CS}}, \mathsf{JD4}_{\mathcal{CS}}, \mathsf{JT}_{\mathcal{CS}}, \mathsf{J4}_{\mathcal{CS}}, \mathsf{LP}_{\mathcal{CS}}\}$. The set of finitary models for $\mathsf{L}_{\mathcal{CS}}$ with the domain being a finite subset of $\mathbb{N}$ is recursively enumerable.*

Proof. We give a proof sketch, for full details cf. [Kuz08, Lemma 4.4.6].

It is obvious that the set of such models for J can be recursively enumerated. Models of each of the other five logics must additionally satisfy certain conditions on the accessibility relation, some combination of transitivity, reflexivity, and seriality. Since each of these conditions can be effectively verified, the models of J that are unsuitable for a given logic can be effectively removed from the enumeration of models for $\mathsf{L}_{\mathcal{CS}}$. □

Lemma A.18. *Let $\mathcal{CS}$ be a decidable schematic constant specification and $\mathsf{L}_{\mathcal{CS}} \in \{\mathsf{J}_{\mathcal{CS}}, \mathsf{JD}_{\mathcal{CS}}, \mathsf{JD4}_{\mathcal{CS}}, \mathsf{JT}_{\mathcal{CS}}, \mathsf{J4}_{\mathcal{CS}}, \mathsf{LP}_{\mathcal{CS}}\}$. Let $\mathcal{M} = (W, R, \mathcal{E}, \nu)$*

be a finitary model for $\mathsf{L}_{\mathcal{CS}}$. Then the relation $\mathcal{M}, w \Vdash A$ between worlds $w \in W$ and formulae A is decidable.

Proof. We give a proof sketch, for full details cf. [Kuz08, Corollary 4.4.8].

We can show this by induction on the formula A, the cases for propositions and Boolean connectives being trivial.

The crucial step is to show that the relation $\mathcal{E}(t, B, w)$ between terms $t \in \mathrm{Tm}$, formulae $B \in \mathrm{Fm}$ and worlds $w \in W$ is decidable (see [Kuz08, Lemma 4.4.7]).

Let $\mathcal{B}$ be the base for the minimal evidence relation $\mathcal{E}$ of $\mathcal{M}$. Given a fixed term t, we will construct a sequence of sets $\mathcal{E}_t^i(w)$ inductively, which can be seen as a partial evidence function that lists all formulae for which t or one of its subterms are admissible evidence at world w.

In order to keep the sets finite and as we are given a schematic constant specification, we will use variables $X, Y, \ldots$ ranging over schemes of formulae and variables $P, Q, \ldots$ ranging over formulae. Also, we assume that our constant specification is given in terms of schemes, i.e.

$$\mathcal{CS} = \{c : X \mid c \text{ is a constant and } X \text{ is a scheme}\}.$$

The sets are defined as follows

$$\mathcal{E}_t^0(w) := \{(s, B) \mid \mathcal{B}(s, B, w) \text{ and } s \in \mathrm{Sub}(t)\} \\ \cup \{(c, X) \mid c : X \in \mathcal{CS} \text{ and } c \in \mathrm{Sub}(t)\}$$

Assume $\mathcal{E}_t^n(w)$ has been constructed, in order to obtain $\mathcal{E}_t^{n+1}(w)$ add the following

- $(s_1 \cdot s_2, Y_1\sigma)$ for any $(s_1, X_1 \to Y_1) \in \mathcal{E}_t^n(w)$ and $(s_2, X_2) \in \mathcal{E}_t^n(w)$ such that the most general unifier σ of X_1 and X_2 exists and $s_1 \cdot s_2 \in \mathrm{Sub}(t)$
- $(s_1 \cdot s_2, Q)$ for any $(s_1, P) \in \mathcal{E}_t^n(w)$ and $(s_2, X_2) \in \mathcal{E}_t^n(w)$ where Q is a fresh variable over formulas and $s_1 \cdot s_2 \in \mathrm{Sub}(t)$
- $(s_1 + s_2, X)$ for any (s_1, X) or $(s_2, X) \in \mathcal{E}_t^n(w)$ with $s_1 + s_2 \in \mathrm{Sub}(t)$
- depending on whether the logic $\mathsf{L}_{\mathcal{CS}}$ contains the (4) axiom, we distinguish the following two cases: If the logic does not contain the (4) axiom, we add

- $(\underbrace{!!\cdots!}_{n+1}c, \underbrace{!!\cdots!}_{n}c : \ldots :!c : c : X)$ for any $c : X \in \mathcal{CS}$ with $\underbrace{!!\cdots!}_{n+1}c \in \mathrm{Sub}(t)$

If the logic contains the (4) axiom, we add

- $(!s, s : X)$ for any $(s, X) \in \mathcal{E}_t^n(w)$ with $!s \in \mathrm{Sub}(t)$
- (s, X) for any $(s, X) \in \mathcal{E}_t^n(v)$ with $R(v, w)$ and $s \in \mathrm{Sub}(t)$

All the sets $\mathcal{E}_t^i(w)$ are finite. As W and $\mathrm{Sub}(t)$ are finite, there is an n easily computable from the size of W and the length of t such that $\mathcal{E}_t^n(w) = \mathcal{E}_t^i(w)$ for all $i \geq n$. Furthermore, we have $\mathcal{E}(t, B, w)$ if and only if B unifies with some X such that $(t, X) \in \mathcal{E}_t^n(w)$. Thus, the relation $\mathcal{E}(t, B, w)$ is decidable. □

Corollary A.19 (Decidability)**.**

1. *Any justification logic in* $\{\mathsf{J}_{\mathcal{CS}}, \mathsf{JT}_{\mathcal{CS}}, \mathsf{J4}_{\mathcal{CS}}, \mathsf{LP}_{\mathcal{CS}}\}$ *with a decidable schematic* $\mathcal{CS}$ *is decidable.*
2. *Any justification logic in* $\{\mathsf{JD}_{\mathcal{CS}}, \mathsf{JD4}_{\mathcal{CS}}\}$ *with a decidable, schematic, and axiomatically appropriate* $\mathcal{CS}$ *is decidable.*

Proof. All logics presented are obviously recursively enumerable. By Corollary A.15, Lemma A.17 and Lemma A.18 all logics presented satisfy the conditions of Theorem A.16 and are, therefore, decidable. □

A.6. The Case of Common Knowledge

While the finiteness of the sets of worlds is a key feature of filtrations, the finite bases of our examples are due to the specific setup of the models and are by no means a necessary property of filtrations. On the other hand, if we start with a logic $\mathsf{L}_{\mathcal{CS}}$, which we already know to be sound and complete with respect to a class of finite models, we can adapt the construction we used to finitely base the evidence function for the filtrations.

Definition A.20. Let $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ be a model and Φ some set of formulae that is closed under subformulae. The Φ-generated submodel $\mathcal{M} \restriction \Phi$ of $\mathcal{M}$ is defined as $(W, R, \mathcal{E} \restriction \Phi, \nu \restriction \Phi)$ where

1. $\mathcal{E} \restriction \Phi$ is the minimal evidence relation based on $\mathcal{B}_\Phi$ where

$$\mathcal{B}_\Phi(t, B, w) \text{ if and only if } t : B \in \Phi \text{ and } \mathcal{E}(t, B, w)$$

2. $\nu \restriction \Phi$ is given by

$$\nu \restriction \Phi(p) = \begin{cases} \{w \mid w \in \nu(p)\} & \text{if } p \in \Phi \\ \emptyset & \text{otherwise} \end{cases}$$

Lemma A.21. *Let $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ be a model, Φ a set of formulae closed under subformulae, and $\mathcal{M} \restriction \Phi$ the Φ-generated submodel of $\mathcal{M}$. Then for all worlds w in $\mathcal{M}$ and formulae $A \in \Phi$ we have*

$$\mathcal{M} \restriction \Phi, w \Vdash A \textit{ if and only if } \mathcal{M}, w \Vdash A.$$

Proof. The proof is by induction on A. The case for atomic propositions is immediate by the definition of $\nu \restriction \Phi$ and the cases for boolean connectives follow immediately by induction hypothesis. Let us consider the case when A is $t : B$.

So assume $\mathcal{M} \restriction \Phi, w \Vdash t : B$. We get $(t, B, w) \in \mathcal{E} \restriction \Phi$ and $\mathcal{M} \restriction \Phi, v \Vdash B$ for all $v \in W$ with $R(w, v)$. The latter gives us $\mathcal{M}, v \Vdash B$ by induction hypothesis whereas from the former we get $(t, B, w) \in \mathcal{E}$ as both $\mathcal{E}$ and $\mathcal{E} \restriction \Phi$ are based on $\mathcal{B}_\Phi$ and $\mathcal{E} \restriction \Phi$ is minimal with that property and hence $\mathcal{E} \restriction \Phi \subseteq \mathcal{E}$. So we have $\mathcal{M}, w \Vdash t : B$.

For the other direction assume $\mathcal{M}, w \Vdash t : B$. We have thus $\mathcal{E}(t, B, w)$ and $\mathcal{M}, v \Vdash B$ for all $v \in W$ with $R(w, v)$. Again, the latter gives us $\mathcal{M} \restriction \Phi, v \Vdash B$ by induction hypothesis and by the definition of $\mathcal{E} \restriction \Phi$ we immediately get $(t, B, w) \in \mathcal{E} \restriction \Phi$ from the former and thus $\mathcal{M} \restriction \Phi, w \Vdash t : B$. □

We can use this technique (adapted to the multi-agent case) to establish decidability for the justification logic with common knowledge $\mathsf{LP}^{\mathsf{C}}_h$ that was introduced in 6 and 7.

The crucial observation here is that the closure conditions for the evidence function, except for the negative intropsection condition, given in Definitions 6.9 and 7.13 define a monotone operator as before for the logics defined in Chapter 6 and 7.

For the logics $\mathsf{L}^{\mathsf{C}}_h \in \{\mathsf{J}^{\mathsf{C}}_h(\mathcal{CS}), \mathsf{JT}^{\mathsf{C}}_h(\mathcal{CS}), \mathsf{J4}^{\mathsf{C}}_h(\mathcal{CS}), \mathsf{LP}^{\mathsf{C}}_h(\mathcal{CS})\}$ we can easily adapt the Φ-generated submodels from Definition A.20 and

Lemma A.21 to the multi-agent case and turn these singleton models into finitary models. Obviously the class of these finitary, singleton models is recursively enumerable, and adapting Lemma A.18 to the multi-agent case shows that these logics satisfy the conditions of Theorem A.16 and decidability follows as in the previous section.

Lemma A.22. *Any justification logic*

$$\mathsf{L}^{\mathsf{C}}_h \in \{\mathsf{J}^{\mathsf{C}}_h(\mathcal{CS}), \mathsf{JT}^{\mathsf{C}}_h(\mathcal{CS}), \mathsf{J4}^{\mathsf{C}}_h(\mathcal{CS}), \mathsf{LP}^{\mathsf{C}}_h(\mathcal{CS})\}$$

with a decidable schematic $\mathcal{CS}$ *is decidable.*

For the logics $\mathsf{L}^{\mathsf{C}}_h \in \{\mathsf{JD}^{\mathsf{C}}_h(\mathcal{CS}), \mathsf{JD4}^{\mathsf{C}}_h(\mathcal{CS})\}$, the proof is a bit more involved. While we have also completeness with respect to the class of singleton models, the evidence relation for these models has to satisfy the consistency condition from Table 7.4 which prevents decidability of the evidence relation. However, we also have completeness with respect to the class of models with serial accessibility relations and without the consistency condition on the evidence relation, see 7.26. In order to establish decidability, we can thus use a filtration construction. We will present the transitive case, as the intransitive case can be easily obtained using the obvious modifications.

Remark A.23. Filtrations for the multi-agent justification logic with common knowledge $\mathsf{L}^{\mathsf{C}}_h$ are defined as filtrations for the single agent case in Definition A.4, except that now conditions (R1) and (R2) have to be satisfied by all accessibility relations $R_1, \ldots, R_h, R_{\mathsf{E}}$, and R_{C} as well as all evidence relations $\mathcal{E}_1, \ldots, \mathcal{E}_h, \mathcal{E}_{\mathsf{E}}$, and $\mathcal{E}_{\mathsf{C}}$ have to satisfy conditions (E1) and (E2). All other notions can also be easily adapted to the multi-agent case.

If $\mathcal{M}$ is a model, $w \in W$ a world in this model and $\mathcal{M}^\Phi$ a filtration through a set of formulae Φ we can show

$$\mathcal{M}, w \Vdash A \text{ if and only if } \mathcal{M}^\Phi, [w] \Vdash A \qquad \text{(A.3)}$$

for any $A \in \Phi$ just as in the proof of Lemma A.5 using the above mentioned conditions (R1), (R2), (E1), and (E2).

Definition A.24. Let $A \in \mathrm{Fm}_{\mathsf{LP}^{\mathsf{C}}_h}$ be a formula, let

$$\begin{aligned}
\Phi_1 &:= \{B, \neg B \mid B \in \mathrm{Sub}(A)\},\\
\Phi_2 &:= \{[\pi_i t]_i B, \neg\,[\pi_i t]_i B \mid \text{for each } [t]_{\mathsf{E}} B \in \Phi_1 \text{ and } i = 1 \ldots h\},\\
\Phi_3 &:= \{[\mathsf{head}(t)]_{\mathsf{E}} B, \neg\,[\mathsf{head}(t)]_{\mathsf{E}} B,\\
&\qquad [\pi_i \mathsf{head}(t)]_i B, \neg\,[\pi_i \mathsf{head}(t)]_i B,\\
&\qquad [\mathsf{tail}(t)]_{\mathsf{E}}\,[t]_{\mathsf{C}} B, \neg\,[\mathsf{tail}(t)]_{\mathsf{E}}\,[t]_{\mathsf{C}} B,\\
&\qquad [\pi_i \mathsf{tail}(t)]_i\,[t]_{\mathsf{C}} B, \neg\,[\pi_i \mathsf{tail}(t)]_i\,[t]_{\mathsf{C}} B\\
&\qquad\qquad \mid \text{for each } [t]_{\mathsf{C}} B \in \Phi_1 \text{ and } i = 1 \ldots h\},
\end{aligned}$$

and set $\Phi := \Phi_1 \cup \Phi_2 \cup \Phi_3$. Φ is called a *suitable set* for A.

Remark A.25.

1. Φ is closed under subformulae.
2. For each $B \in \Phi$ there is a $C \in \Phi$ such that for any constant specification $\mathcal{CS}$ we have
$$\mathsf{JD4}^{\mathsf{C}}_h(\mathcal{CS}) \vdash \neg B \leftrightarrow C.$$
3. If $[t]_{\mathsf{C}} B \in \Phi$, then $[\mathsf{head}(t)]_{\mathsf{E}} B \in \Phi$ and $[\mathsf{tail}(t)]_{\mathsf{E}}\,[t]_{\mathsf{C}} B \in \Phi$.
4. If $[t]_{\mathsf{E}} B \in \Phi$, then $[\pi_i t]_i B \in \Phi$ for each $i = 1 \ldots h$.

Let us fix a formula $A \in \mathrm{Fm}_{\mathsf{JD4}^{\mathsf{C}}_h}$ and a suitable set Φ for A for the remaining part of this section.

Definition A.26. Let $\mathcal{CS}$ be a C-axiomatically appropriate constant specification and let $\mathcal{M}$ be a model for $\mathsf{JD4}^{\mathsf{C}}_h$ meeting $\mathcal{CS}$ as described in Corollary 7.26, i.e., with serial (and transitive) accessibility relations but no consistency condition on the evidence function.

We define the $\mathcal{M}^{\Phi} = (W^{\Phi}, R^{\Phi}, \mathcal{E}^{\Phi}, \nu^{\Phi})$ in the following way:

- $W^{\Phi} := \{[w]_{\Phi} \mid w \in W\}$,
-
  - $R^{\Phi}_i([w],[v])$ if and only if for all $[t]_i B \in \Phi$ we have
  $$\mathcal{M}, w \Vdash [t]_i B \text{ implies } \mathcal{M}, v \Vdash B \wedge [t]_i B,$$
  - $R^{\Phi}_{\mathsf{E}} := \bigcup_{i=1}^{h} R^{\Phi}_i$,

– $R^\Phi_{\mathsf{C}} := \bigcup_{n=1}^{\infty} (R^\Phi_{\mathsf{E}})^n$,

- $\mathcal{E}^\Phi$ is the minimal t-evidence relation based on $\mathcal{B}^\Phi$ given by

$$\mathcal{B}^\Phi_{\circledast}(t, B, [w]) \text{ if and only if } [t]_{\circledast} B \in \Phi \text{ and } \mathcal{M}, w \Vdash [t]_{\circledast} B,$$

- $\nu^\Phi(P) := \begin{cases} \{[w] \mid w \in \nu(P)\} & \text{if } P \in \Phi, \\ \emptyset & \text{otherwise.} \end{cases}$

Remark A.27. $\mathcal{M}^\Phi$ is a model for $\mathsf{JD4}^{\mathsf{C}}_h$, i.e. R^Φ_i is transitive and serial for each $i = 1 \ldots h$ and $\mathcal{E}^\Phi$ is an evidence relation.

Proof. To show that R^Φ_i is transitive is a simple repetition of the proof in Lemma A.11.

Let us show that R^Φ_i is serial. Assume (towards a contradiction) that R^Φ_i is not serial. Then there must be a world $w \in W$ such that $[w]$ is isolated with respect to R^Φ_i. By definition of R^Φ_i, this means that for each $v \in W$ there is a formula $[t_v]_i B_v \in \Phi$ such that $\mathcal{M}, w \Vdash [t_v]_i B_v$ but $\mathcal{M}, v \nVdash B_v \wedge [t_v]_i B_v$. But this contradicts the seriality of R_i which implies there is a world $u \in W$ such that $R_i(w, u)$ and so in particular $\mathcal{M}, w \Vdash [t_u]_i B_u$ and $\mathcal{M}, u \Vdash B_u \wedge [t_u]_i B_u$.

Finally, we need to show that $\mathcal{E}$ is not only a t-evidence relation but an actual evidence relation, i.e.,

$$\text{if } \mathcal{E}^\Phi_*(t, B, [w]) \text{ and } R^\Phi_*([w], [v]) \text{ then } \mathcal{E}^\Phi_*(t, B, [v]). \tag{A.4}$$

In order to do so, we first show the following property for any formula $[t]_{\circledast} B$

$$\text{if } \mathcal{E}^\Phi_{\circledast}(t, B, [w]) \text{ then } \mathcal{M}, w \Vdash [t]_{\circledast} B. \tag{A.5}$$

We proceed by induction on the construction of $\mathcal{E}^\Phi$. The base case is trivial: if $\mathcal{B}^\Phi_{\circledast}(t, B, [w])$ because of $\mathcal{M}, w \Vdash [t]_{\circledast} B$, we already have the desired result. The induction step is equally simple, as the closure conditions of $\mathcal{E}^\Phi$ are directly modelled on the axioms of $\mathsf{JD4}^{\mathsf{C}}_h$ and thus the conclusion follows immediately.

Now we can show that (A.4) holds. Assume first $* = i \in \{1 \ldots h\}$. If $\mathcal{E}^\Phi_i(t, B, [w])$, then, by (A.5), we have $\mathcal{M}, w \Vdash [t]_i B$. From $R^\Phi_i([w], [v])$ and by the definition of R^Φ_i we get $\mathcal{M}, v \Vdash [t]_i B$. Thus we have $\mathcal{B}^\Phi_i(t, B, [w])$ and so also $\mathcal{E}^\Phi_i(t, B, [w])$.

Now assume, $* = \mathsf{C}$. Again, from $\mathcal{E}^{\Phi}_{\mathsf{C}}(t, B, [w])$ and (A.5) we get $\mathcal{M}, w \Vdash [t]_{\mathsf{C}} B$. As $R^{\Phi}_{\mathsf{C}}([w], [v])$, there are $v_1, \ldots, v_n \in W$ with $w = v_1$ and $v = v_n$ such that $[v_1] R^{\Phi}_{\mathsf{E}} [v_2] R^{\Phi}_{\mathsf{E}} \ldots R^{\Phi}_{\mathsf{E}} [v_n]$.

Using Lemma 6.2, we can get terms $t_1, \ldots, t_{n-1} \in \mathrm{Tm}_{\mathsf{C}}$ such that

$$\mathcal{M}, w \Vdash [t_1]_{\mathsf{E}}\, [t_2]_{\mathsf{E}} \ldots [t_{n-1}]_{\mathsf{E}}\, [t]_{\mathsf{C}} B$$

and we immediately obtain $\mathcal{M}, v \Vdash [t]_{\mathsf{C}} B$. Again we get $\mathcal{B}^{\Phi}_{\mathsf{C}}(t, B, [w])$ and so also $\mathcal{E}^{\Phi}_{\mathsf{C}}(t, B, [w])$. □

Note that Lemma 6.2 requires the C-axiomatical appropriateness of $\mathcal{CS}$.

Lemma A.28. *$\mathcal{M}^{\Phi}$ is a filtration of $\mathcal{M}$ through Φ.*

Proof. We have to show that conditions (R1), (R2), (E1), and (E2) hold for $R_1, \ldots, R_h, R_{\mathsf{E}}, R_{\mathsf{C}}$ and $\mathcal{E}_1, \ldots, \mathcal{E}_h, \mathcal{E}_{\mathsf{E}}, \mathcal{E}_{\mathsf{C}}$, respectively.

(**R1**) We consider the case $\circledast = i$ first. Assume $R_i(w, v)$ and let $[t]_i B \in \Phi$ with $\mathcal{M}, w \Vdash [t]_i B$. Then, by the positive inspection axiom, we also have $\mathcal{M}, w \Vdash [!t]_i\, [t]_i B$ and we get $\mathcal{M}, v \Vdash B \wedge [t]_i B$. So, we can conclude $R^{\Phi}_i([w], [v])$.

Now let us consider the case $\circledast = \mathsf{E}$. If $R_{\mathsf{E}}(w, v)$, then we have $R_i(w, v)$ for some i and by the previously shown (R1) for i we get $R^{\Phi}_i([w], [v])$ and so also $R^{\Phi}_{\mathsf{E}}([w], [v])$.

Finally, we let $\circledast = \mathsf{C}$. So assume $R_{\mathsf{C}}(w, v)$. This means, we have $(w, v) \in (R_{\mathsf{E}})^n$ for some natural number n and by the previously shown (R1) for E we obtain $([w], [v]) \in (R^{\Phi}_{\mathsf{E}})^n$ and thus $R^{\Phi}_{\mathsf{C}}([w], [v])$ by definition.

(**R2**) The case for $\circledast = i$ follows immediately by definition of R^{Φ}_i.

For $\circledast = \mathsf{E}$, let $R^{\Phi}_{\mathsf{E}}([w], [v])$, $[t]_{\mathsf{E}} B \in \Phi$ and $\mathcal{M}, w \Vdash [t]_{\mathsf{E}} B$. From $R^{\Phi}_{\mathsf{E}}([w], [v])$ we get that $R^{\Phi}_i([w], [v])$ for some i and as Φ is a suitable set, from $[t]_{\mathsf{E}} \Phi$ we get $[\pi_i t]_i B \in \Phi$. By the projection axiom we get $\mathcal{M}, w \Vdash [\pi_i t]_i B$ and by (R2) for i we finally obtain $\mathcal{M}, w \Vdash B$.

For the case $\circledast = \mathsf{C}$, let $R^{\Phi}_{\mathsf{C}}([w], [v])$, $[t]_{\mathsf{C}} B \in \Phi$ and $\mathcal{M}, w \Vdash [t]_{\mathsf{C}} B$. By definition of R^{Φ}_{C}, there are $w = v_1, v_2, \ldots, v_n = v \in W$ with $[v_1] R^{\Phi}_{\mathsf{E}} [v_2] R^{\Phi}_{\mathsf{E}} \ldots R^{\Phi}_{\mathsf{E}} [v_n]$. The suitability of Φ implies

$[\mathsf{tail}(t)]_{\mathsf{E}}\,[t]_{\mathsf{C}}B \in \Phi$ and by (R2) for E we get $\mathcal{M}, v_j \Vdash [t]_{\mathsf{C}}B$ as $\mathcal{M}, v_{j-1} \Vdash [\mathsf{tail}(t)]_{\mathsf{E}}\,[t]_{\mathsf{C}}B$ for each $j < n$. In particular $\mathcal{M}, v_{n-1} \Vdash [t]_{\mathsf{C}}B$ and thus also $\mathcal{M}, v_{n-1} \Vdash [\mathsf{head}(t)]_{\mathsf{E}}B$. The suitability of Φ and (R2) for E finally imply $\mathcal{M}, v_n \Vdash B$, i.e., $\mathcal{M}, v \Vdash B$

(**E1**) This follows immediately by the definition of $\mathcal{E}^{\Phi}_{\circledast}$. Let $[t]_{\circledast}B \in \Phi$ and $\mathcal{M}, w \Vdash [t]_{\circledast}B$. Then $\mathcal{B}^{\Phi}_{\circledast}(t, B, w)$ and so also $\mathcal{E}^{\Phi}_{\circledast}(t, B, w)$.

(**E2**) Let $\mathcal{E}^{\Phi}_{\circledast}(t, B, w)$ and $[t]_{\circledast}B \in \Phi$. By (A.5) we get $\mathcal{M}, w \Vdash [t]_{\circledast}B$ and so also $\mathcal{E}_{\circledast}(t, B, w)$. □

Now we can again employ the same techniques as in the previous section to obtain

Lemma A.29. *Any justification logic*

$$\mathsf{L}^{\mathsf{C}}_{h} \in \{\mathsf{JD}^{\mathsf{C}}_{h}(\mathcal{CS}), \mathsf{JD4}^{\mathsf{C}}_{h}(\mathcal{CS})\}$$

with a decidable, schematic and axiomatically appropriate $\mathcal{CS}$ is decidable.

Combining Lemma A.22 and A.29, we get the following results.

Theorem A.30.

1. *Any justification logic*

$$\mathsf{L}^{\mathsf{C}}_{h} \in \{\mathsf{J}^{\mathsf{C}}_{h}(\mathcal{CS}), \mathsf{JT}^{\mathsf{C}}_{h}(\mathcal{CS}), \mathsf{J4}^{\mathsf{C}}_{h}(\mathcal{CS}), \mathsf{LP}^{\mathsf{C}}_{h}(\mathcal{CS})\}$$

   with a decidable schematic $\mathcal{CS}$ is decidable.

2. *Any justification logic*

$$\mathsf{L}^{\mathsf{C}}_{h} \in \{\mathsf{JD}^{\mathsf{C}}_{h}(\mathcal{CS}), \mathsf{JD4}^{\mathsf{C}}_{h}(\mathcal{CS})\}$$

   with a decidable, schematic and axiomatically appropriate $\mathcal{CS}$ is decidable.

For the case of negative inspection (j5), see the discussion in the following section.

A.7. Notes

We have presented a uniform method of proving decidability for justification logics using a refinement of the finite model property. In order to achieve this property, we have adapted the modal techniques of filtration and generated submodels to justification logics. Apart from reproving the known decidability results for $\mathsf{J}_{\mathcal{CS}}$, $\mathsf{JD}_{\mathcal{CS}}$, $\mathsf{JT}_{\mathcal{CS}}$, $\mathsf{J4}_{\mathcal{CS}}$, $\mathsf{JD4}_{\mathcal{CS}}$, and $\mathsf{LP}_{\mathcal{CS}}$, this method has enabled us to establish the decidability of the justification logics with common knowledge introduced in Chapters 6 and 7.

The main difference from the modal case is the presence of an additional element in models called evidence relation. As evidence relations are in general infinite objects, the filtration has to be performed in such a way that apart from finitizing the set of worlds, also the evidence relation is finitely representable. This finite representation is achieved by using least fixed points of a certain monotone operator that can be read off the axioms of the logic. The existence of the least fixed point is guaranteed when the operator is monotone, which is the case for all the logics considered. Some logics, e.g. justification logics with negative inspection, however, give rise to non-monotone operators. Proving decidability for them requires more involved techniques, see [Stu11].

B. Public Announcements

We saw in Chapter 3, Section 3.6 how public announcements can lead to common knowledge. It is a natural idea to combine logics of common knowledge and public announcements (see [DHK07]) and it is a long-term goal of our research to present a justification counterpart of these logics. Of course, it is first necessary to study the justification counterparts of these systems by themselves before combining them. The material in this chapter is based on [BKS11b; Buc+10; BKS12b] where such systems are introduced. We will give a brief overview including the main theorems but omitting most of the proofs as they can be found in the referenced sources. From the given remarks it should also become clear, that the combination of these logics for now is work in progress, a better understanding of both the justification logic for common knowledge and the justification logic for public announcements is required.

Dynamic epistemic logic [DHK07] studies the relationship between communication, knowledge, and belief. It is based on the language of modal logic enriched with statements to express various forms of communication. A basic form of communication is provided by *public announcements* where a statement A is publicly communicated to all the agents. The logic of public announcements [Pla07b; GG97] uses a statement $[A]B$ to express that B holds after the public announcement of A.

In order to simplify our presentation, as in previous chapters, we will treat belief rather than knowledge and, hence, rely on Gerbrandy–Groeneveld's axiomatization of public announcements [GG97]. One of its postulates is

$$\Box(A \to [A]B) \leftrightarrow [A]\Box B \ , \tag{B.1}$$

which says, from left to right, that an agent who believes that B must be the case whenever a true fact A is announced will believe B after an actual announcement of A.

To illustrate how this principle works, let us briefly recall the following example from [BKS11b]. Elite-level frequent flyers can usually check in for their flight at the business counter by presenting their elite membership card, which can also be attached to their luggage to make public their elite status. This check-in rule is known to airline employees. In this situation, it follows by the implication (B.1) that when Ann presents her elite membership card to Bob at the business counter, he knows that he should check her in.

Modal public announcement logic tells us *how* beliefs change after public announcements but not *why*. We try to formalize possible answers to this *why* using the approach of justification logic. If we convert the left to right implication from (B.1) to a statement with explicit justifications, we obtain something like

$$s:(A \to [A]B) \to [A]t:B \ , \tag{B.2}$$

where s represents the airline's regulations regarding business-counter check-in procedures and t is the reason why Bob starts checking Ann in.

The question is how the terms s and t, which represent justifications, relate to each other; in particular, how to arrive at t given s. We use the above example to discuss different answers to this question. There are the following possibilities.

1. $t = s$. The regulations themselves tell Bob to check Ann in. This option is implemented in the logic $\mathsf{JPAL(K)}$, which we developed jointly with Bryan Renne and Joshua Sack [Buc+10].

2. $t = \Uparrow s$. The operator $\Uparrow$ represents the inference Bob has to make from the regulations after the elite card is shown. This approach is taken by the logic $\mathsf{OPAL(K)}$, which we introduced in [BKS11b].

3. $t = \Uparrow_A s$. The inference process explicitly mentions both the regulations, s, and the demonstration of Ann's elite card, A. We do not consider this variant since it would make schematic reasoning impossible. Indeed (B.1) is an axiom scheme that does not depend on the announcement. Therefore, the operation that represents the update on the level of terms should not depend on the announcement either.

As already argued in [BKS11b], the simplicity of the first option, axiomatized by $\mathsf{JPAL(K)}$, may not always be sufficient. Imagine that Ann has been upgraded to business class (say, as a reward for postponing her original flight, which had been overbooked). So, according to the same regulations, she can check in with Bob based on her ticket alone without announcing her elite status, which in our notation is represented by $s : B$. But Ann may choose to announce her elite status anyways, or $[A]s : B$ in our notation. In $\mathsf{JPAL(K)}$, where $t = s$, after the elite status is announced, t encodes two different reasons for Bob to check Ann in: as a business-class customer and as an elite flyer. By contrast, in $\mathsf{OPAL(K)}$, these two reasons are represented by two different terms, s and $\Uparrow s$, of which the latter depends on Ann's elite status while the former is due to the ticket alone. And Bob would want to distinguish between the two reasons because of the difference in baggage allowances: an elite frequent flyer is often allowed to check more luggage for free than an owner of a business-class ticket who has been upgraded from economy.

In addition, in this and similar cases, the approach of $\mathsf{JPAL(K)}$ implies that the meaning of the regulations changes after public announcements: if Ann has an economy ticket, the regulations do not allow her a business-counter check-in until she shows her elite card, and then they do. This is a little counterintuitive since the regulations are a legal document whose meaning should not be changed by each public announcement. The use of reason $\Uparrow s$ enables us to separate the permanent status of the regulations from their momentary applications influenced by public announcements.

Let us now look at the other direction of (B.1)—from right to left—and see how the first two options manifest themselves there. The implication states that an agent who will believe B after an announcement of A must believe that, if A is true and announced, B holds after the announcement. For instance, if Charlie, while standing in a long line at the economy check-in counter, sees Ann showing her elite card and being served by Bob at the business counter, $[A]\Box B$, then Charlie has empirical evidence e that Ann is served at the business counter, $[A]e : B$. It would be natural for Charlie to believe that having an elite status and showing it gets one to the business counter, $\Box(A \to [A]B)$. But it seems even clearer in this case that Charlie's empirical observation e cannot explain the causality of the implication $A \to [A]B$. If before Ann showed up, Charlie had read the sign that invited elite members to the business counter, then Charlie's memory of this sign, refreshed by Ann's actions, could serve as such an explanation. Thus, instead of using e, as in JPAL(K), in this example it also seems better to use $\Downarrow e$, where $\Downarrow$ is yet another new operation of our logic OPAL(K).

Besides the explicit justifications for dynamic epistemic logic in the already mentioned [Buc+10; BKS11b; BKS12b], there is Renne's earlier research on introducing new evidence [Ren08] and eliminating unreliable evidence [Ren] in the framework of justification logic. He also presents expressivity results for certain justification logics with announcements [Ren11]. However, the modal counterparts of Renne's systems do not correspond to any traditional public announcement logic whereas both JPAL(K) and OPAL(K) are intended as justification logics with public announcement operators whose belief dynamics closely corresponds to the modal belief dynamics of Gerbrandy–Groeneveld's modal public announcement logic PAL(K) [GG97].

In the next section, we recall the axiomatization and basic properties of PAL(K). In particular, we present the reduction of a PAL(K) formula A to a provably equivalent formula $\mathsf{red}(A)$ that does not contain public announcement operators. This reduction facilitates a simple completeness proof for PAL(K) by reducing it to completeness of the basic modal logic K.

As mentioned before, JPAL(K) and OPAL(K) (both with additional positive introspection axioms) were introduced in [Buc+10] and [BKS11b], respectively, where we also established soundness and completeness for these two logics. We give the definitions of OPAL(K) and JPAL(K) and their semantics in Sections B.2 and B.3. Soundness and completeness for OPAL(K) and JPAL(K) is proved in Section B.4. Since the replacement property does not hold in justification logics, we cannot establish completeness of either logic by reducing it to completeness of the basic justification logic J. Instead we perform a canonic model construction for each of the two logics. In Section B.5, we a partial realization theorem in the following way. First, we reduce the PAL(K) formula A to a provably equivalent formula $\mathsf{red}(A)$ that has no announcement operators, i.e., $\mathsf{red}(A)$ is a traditional modal logic formula. Then we use realization for modal logic (without public announcements) to obtain a justification logic formula $r(\mathsf{red}(A))$ that realizes $\mathsf{red}(A)$. Finally, we 'invert' the reduction from A to $\mathsf{red}(A)$ on the justification logic side to obtain a formula $r(A)$ that realizes A, see Figure B.1 on page 174.

We call this approach *realization by reduction*, the closest analog of this method can be found in [Fit11], where S5 is realized by reducing it to K45. However, there the reversal of the reduction is trivial, while in our setting it requires an involved extension of Fitting's replacement theorem [Fit09]. First, we need replacement also for formulas with public announcements and, second, we need replacement also in negative positions (the original proof in [Fit09] only deals with replacement in positive positions). While we only show this extended replacement theorem for JPAL(K), there seems to be little or no extra work required to prove the same extended replacement theorem for OPAL(K). The problem lies in the application of this replacement theorem to reverse the modal reduction on the justification side for OPAL(K). The exact nature of the problem is too technical to be explained in the introduction and is pointed out in the proof of Theorem B.44, Footnote 3. We only mention here that the problem concerns reversing in OPAL(K) the modal update reduction in a negative position. Thus, we obtain realization only for JPAL(K). It is open whether a similar result can be shown for OPAL(K).

B.1. Modal Public Announcement Logic

In this section, we recall some of the basic definitions and facts concerning the Gerbrandy–Groeneveld modal logic of public introspective announcements [Ger99; GG97; DHK07], i.e, public announcements that need not be truthful but are trusted by all the agents.

Definition B.1 (PAL(K) Language)**.** We fix a countable set Prop of *atomic propositions.* The *language of* PAL(K) consists of the *formulas* $A \in \mathsf{Fml}_{\Box,[\cdot]}$ formed by the grammar

$$A ::= p \mid \neg A \mid (A \to A) \mid \Box A \mid [A]A \ ,$$

where $p \in \mathsf{Prop}$. The language $\mathsf{Fml}_{\Box}$ of modal formulas without announcements is obtained from the same grammar without the $[A]A$ constructor.

The Gerbrandy–Groeneveld theory PAL(K) of Public Announcement Logic uses the language $\mathsf{Fml}_{\Box,[\cdot]}$ to reason about belief change and public announcements.

Definition B.2 ($\mathsf{PAL(K)}$ Deductive System)**.** The *axioms of* $\mathsf{PAL(K)}$ consist of all $\mathsf{Fml}_{\Box,[\cdot]}$-instances of the following schemes:

1. Axiom schemes for the modal logic K
2. $[A]p \leftrightarrow p$ (independence)
3. $[A](B \to C) \leftrightarrow ([A]B \to [A]C)$ (normality)
4. $[A]\neg B \leftrightarrow \neg[A]B$ (functionality)
5. $[A]\Box B \leftrightarrow \Box(A \to [A]B)$ (update)
6. $[A][B]C \leftrightarrow [A \wedge [A]B]C$ (iteration)

The *deductive system* $\mathsf{PAL(K)}$ is a Hilbert system that consists of the above axioms of $\mathsf{PAL(K)}$ and the following rules of *modus ponens* (MP) and *necessitation* (Nec):

$$\frac{A \quad A \to B}{B}\,(\mathrm{MP}) \ , \qquad \frac{A}{\Box A}\,(\mathsf{Nec}) \ .$$

We write $\mathsf{PAL(K)} \vdash A$ to state that $A \in \mathsf{Fml}_{\Box,[\cdot]}$ is a theorem of $\mathsf{PAL(K)}$.

Lemma B.3 (Admissible Announcement Necessitation, [DHK07])**.** *Announcement necessitation is admissible in* $\mathsf{PAL(K)}$*: that is, for all formulas* $A, B \in \mathsf{Fml}_{\Box,[\cdot]}$*, we have*

$$\mathsf{PAL(K)} \vdash A \ \textit{implies}\ \mathsf{PAL(K)} \vdash [B]A \ .$$

$\mathsf{PAL(K)}$, like many traditional modal public announcement logics, features the so-called *reduction property*: $\mathsf{Fml}_{\Box,[\cdot]}$-formulas can be reduced to provably equivalent $\mathsf{Fml}_{\Box}$-formulas [Ger99; GG97; DHK07]. That means one can express what the situation is after an announcement by saying what the situation was before the announcement. The following lemma formally describes this reduction procedure (for a proof, see, for instance, [DHK07]). The proof method was first introduced by Plaza in [Pla07b].

Definition B.4 (Announcement Redexes and their Reducts)**.** The following are five pairs of redexes and their reducts:

Redex	Its reduct
$[A]p$	p
$[A]\neg B$	$\neg[A]B$
$[A](B \to C)$	$[A]B \to [A]C$
$[A]\Box B$	$\Box(A \to [A]B)$
$[A][B]C$	$[A \wedge [A]B]C$

(B.3)

Definition B.5 (Reduction)**.** The one-step reduction function

$$\mathsf{red}_1 : \mathsf{Fml}_{\Box,[\cdot]} \to \mathsf{Fml}_{\Box,[\cdot]}$$

is defined as follows:

- If no subformula of $A \in \mathsf{Fml}_{\Box,[\cdot]}$ is a redex, then $\mathsf{red}_1(A) := A$.
- Otherwise, let R be the outermost leftmost subformula occurrence of A that is a redex, i.e.,
  1) if R is a proper subformula occurrence of R', which is a subformula occurrence of A, R' is not a redex;
  2) if R is a subformula occurrence of C and $B \to C$ is a subformula occurrence of A, no redex can occur in B.

  In this case, $\mathsf{red}_1(A)$ is defined as the result of replacing the formula occurrence R in A with its reduct.

Note that R is outside of announcements in A. Indeed, if R occured in B with $[B]C$ being a subformula occurrence of A, then $[B]C$ would itself be a redex, which is prohibited by item 1) above.

In order to show that the reduction function has the intended behavior, we need the following notion.

Definition B.6 (Rank)**.** The *rank* $\mathsf{rk}^{\Box}(A)$ of a formula $A \in \mathsf{Fml}_{\Box,[\cdot]}$ is defined as follows:

1. $\mathsf{rk}^{\Box}(p) := 1$ for each $p \in \mathsf{Prop}$
2. $\mathsf{rk}^{\Box}(\neg A) := \mathsf{rk}^{\Box}(A) + 1$

3. $\mathsf{rk}^\Box(A \to B) := \max(\mathsf{rk}^\Box(A), \mathsf{rk}^\Box(B)) + 1$

4. $\mathsf{rk}^\Box(\Box A) := \mathsf{rk}^\Box(A) + 1$

5. $\mathsf{rk}^\Box([A]B) := (2 + \mathsf{rk}^\Box(A)) \cdot \mathsf{rk}^\Box(B)$

The following properties can be verified by easy and straightforward calculations.

Lemma B.7 (Reductions Reduce Rank)**.** *For all formulas $A, B, C \in \mathsf{Fml}_{\Box,[\cdot]}$ we have the following:*

1. $\mathsf{rk}^\Box(A) > \mathsf{rk}^\Box(B)$ *if B is a proper subformula of A*

2. $\mathsf{rk}^\Box([A]\neg B) > \mathsf{rk}^\Box(\neg[A]B)$

3. $\mathsf{rk}^\Box([A](B \to C)) > \mathsf{rk}^\Box([A]B \to [A]C)$

4. $\mathsf{rk}^\Box([A]\Box B) > \mathsf{rk}^\Box(\Box(A \to [A]B))$

5. $\mathsf{rk}^\Box([A][B]C) > \mathsf{rk}^\Box([A \wedge [A]B]C)$

Using this formula rank it is easy to show, that for any formula $A \in \mathsf{Fml}_{\Box,[\cdot]}$, there exists $N > 0$ such that $\mathsf{red}_1^{N+1}(A) = \mathsf{red}_1^N(A)$, which must then be an $\mathsf{Fml}_\Box$-formula. In this case, $\mathsf{red}_1^n(A) = \mathsf{red}_1^N(A)$ for any $n \geq N$ and we define $\mathsf{red}(A) := \mathsf{red}_1^N(A)$.

Lemma B.8 (Provable Equivalence of Reductions)**.** *For all formulae $A \in \mathsf{Fml}_{\Box,[\cdot]}$, we have $\mathsf{PAL(K)} \vdash A \leftrightarrow \mathsf{red}_1(A)$, and, consequently,*

$$\mathsf{PAL(K)} \vdash A \leftrightarrow \mathsf{red}(A) \quad .$$

Remark B.9. The above lemma facilitates a completeness proof for PAL(K) by reducing it to completeness of K. The completeness is proved with respect to the class of all Kripke models. To evaluate validity of formulas with announcements, the standard Kripke semantics is extended with a model update operation for introspective announcements. For the sake of brevity we refer to [DHK07, Section 4.9] for a full completeness proof and give only a sketch. Suppose that $A \in \mathsf{Fml}_{\Box,[\cdot]}$ is valid. Then $\mathsf{red}(A)$ is also valid by Lemma B.8 and by soundness of PAL(K), which is easy to show directly. Since $\mathsf{red}(A)$ is a formula of $\mathsf{Fml}_\Box$, completeness of K yields $\mathsf{K} \vdash \mathsf{red}(A)$ and, hence, $\mathsf{PAL(K)} \vdash \mathsf{red}(A)$ because $\mathsf{PAL(K)}$ extends K. Applying Lemma B.8 again, we conclude that $\mathsf{PAL(K)} \vdash A$. As a corollary of the soundness of $\mathsf{PAL(K)}$, since the semantics for $\mathsf{PAL(K)}$ extends the standard Kripke semantics, $\mathsf{PAL(K)}$ is a conservative extension of K.

B.2. Justification Public Announcement Logic

Our language extends the language typically used in justification logic by adding public announcement formulas $[A]B$ and two unary operations on terms, $\Uparrow$ and $\Downarrow$, to express the update dynamics of evidence. However, of the two logics $\mathsf{JPAL(K)}$ and $\mathsf{OPAL(K)}$ introduced in this section, only $\mathsf{OPAL(K)}$ uses the $\Uparrow$ and $\Downarrow$ operations to record the event of an announcement, whereas $\mathsf{JPAL(K)}$ does not explicitly record such an event by an operation on terms.

Definition B.10 (Language)**.** In addition to the set of propositions Prop, we fix countable sets Cons of *constants* and Vars of *variables*. Our language consists of the *terms* given by the grammar

$$t ::= x \mid c \mid (t \cdot t) \mid (t + t) \mid \Uparrow t \mid \Downarrow t \ ,$$

where $x \in \mathsf{Vars}$ and $c \in \mathsf{Cons}$ and the *formulae* formed by the grammar

$$A ::= p \mid \neg A \mid (A \to A) \mid t : A \mid [A]A \ ,$$

where $p \in \mathsf{Prop}$. We denote the *set of terms* by Tm and the *set of formulae* by Fml_J.

A term is a *ground term* if it does not contain variables. The language introduced in [Buc+10; BKS11b] for justification logics with public announcements includes additionally an operation ! on terms that is used for positive introspection. Since the logics $\mathsf{OPAL(K)}$ and $\mathsf{JPAL(K)}$, to be introduced below, do not have an introspection axiom, we can dispense with the ! operation.

Remark B.11. To state axioms of our systems $\mathsf{JPAL(K)}$ and $\mathsf{OPAL(K)}$, we use arbitrary finite sequences of announcements, which is not done in modal public announcement logics. This use of sequences may seem puzzling, especially given that the iteration axiom of $\mathsf{PAL(K)}$, which normally allows to replace any such finite sequence with a single announcement, is transferred to JPAL(K) and OPAL(K) as is. But recall that the replacement property does not hold for justification logics, as already mentioned earlier. Replacing single announcements with sequences of announcements in axioms is the minimally invasive solution we have found to ensure the admissibility of the announcement necessitation rule, which is clearly valid semantically. For instance, if a theorem B is obtained by modus ponens from $C \to B$ and C, it follows by announcement necessitation that $[A_1] \dots [A_n]B$ should be derivable. For $n = 1$, the normality axiom of PAL(K) takes care of the transition from $[A_1](C \to B)$ and $[A_1]C$ to $[A_1]B$. In order to make such a transition possible for an arbitrary $n > 0$ we generalize the normality axiom to allow an arbitrary finite sequence of announcements $[A_1] \dots [A_n]$.

σ and τ (with and without subscripts) will denote finite sequences of formulas. ε denotes the empty sequence. Given such a sequence $\sigma = (A_1, \dots, A_n)$ and a formula B, the formula $[\sigma]B$ is defined as follows:

$$[\sigma]B := [A_1] \dots [A_n]B \text{ if } n > 0 \qquad \text{and} \qquad [\varepsilon]B := B.$$

Further, we define

$$\sigma, B \quad := \quad (A_1, \dots, A_n, B) \quad \text{and} \quad B, \sigma := (B, A_1, \dots, A_n) \ .$$

For a sequence $\tau = (C_1, \dots, C_m)$, we define

$$\tau, \sigma \quad := \quad (C_1, \dots, C_m, A_1, \dots, A_n) \ .$$

We will also need the length $|\sigma|$ of a sequence σ, which is given by $|\varepsilon| := 0$ and $|(A_1, \dots, A_n)| := n$.

Definition B.12 (OPAL(K))**.** The *axioms of* OPAL(K) consist of all $\mathsf{Fml_J}$-instances of the following schemes:

1. $[\sigma]A$, where A is a classical propositional tautology
2. $[\sigma](t : (A \to B) \to (s : A \to t \cdot s : B))$ (application)
3. $[\sigma](t : A \to t + s : A)$, $\quad [\sigma](s : A \to t + s : A)$ (sum)

4. $[\sigma]p \leftrightarrow p$ (independence)
5. $[\sigma](B \to C) \leftrightarrow ([\sigma]B \to [\sigma]C)$ (normality)
6. $[\sigma]\neg B \leftrightarrow \neg[\sigma]B$ (functionality)
7. $[\sigma]t:(A \to [A]B) \to [\sigma][A]\Uparrow t:B$ (update $\Uparrow$)
8. $[\sigma][A]t:B \to [\sigma]\Downarrow t:(A \to [A]B)$ (update $\Downarrow$)
9. $[\sigma][A][B]C \leftrightarrow [\sigma][A \wedge [A]B]C$ (iteration)

The *deductive system* $\mathsf{OPAL(K)}$ is a Hilbert system that consists of the above axioms of $\mathsf{OPAL(K)}$ and the following rules of *modus ponens* (MP) and *axiom necessitation* (AN):

$$\frac{A \quad A \to B}{B}\ (\text{MP})\ ,$$

$$\frac{c_1, \ldots, c_n \in \mathsf{Cons} \quad C \text{ is an } \mathsf{OPAL(K)}\text{-axiom}}{[\sigma_1]c_1 : \cdots : [\sigma_n]c_n : C}\ (\text{AN})\ ,$$

where σ_i's are (possibly empty) finite sequences of formulas.

We sometimes use the same names for both axioms of $\mathsf{OPAL(K)}$ and axioms of $\mathsf{PAL(K)}$ because it will always be clear from the context which of the two is meant.

Besides $\mathsf{OPAL(K)}$, we also consider the deductive system $\mathsf{JPAL(K)}$, which does not assign any particular meaning to the two term operations $\Uparrow$ and $\Downarrow$.

Definition B.13 ($\mathsf{JPAL(K)}$)**.** The *axioms of* $\mathsf{JPAL(K)}$ are the axioms of $\mathsf{OPAL(K)}$ where the two update axiom schemes are replaced by the single scheme

$$[\sigma]t:(A \to [A]B) \leftrightarrow [\sigma][A]t:B\ . \qquad \text{(update)}$$

The *deductive system* $\mathsf{JPAL(K)}$ is a Hilbert system that consists of the axioms of $\mathsf{JPAL(K)}$ and the rules (MP) and (AN), where the formula C in (AN) now stands for an axiom of $\mathsf{JPAL(K)}$.

We will use $\mathsf{OPAL(K)} \vdash A$ and $\mathsf{JPAL(K)} \vdash A$ to express that A is derivable in $\mathsf{OPAL(K)}$ and $\mathsf{JPAL(K)}$, respectively. If the deductive system does not matter, for instance when A is derivable in both of them, then we use $\vdash A$.

The following example gives some intuition as to how the deductive systems work and what their differences are.

Example B.14. For any $p \in \mathsf{Prop}$ and any $c_1, c_2 \in \mathsf{Cons}$, we have

1. $\mathsf{OPAL(K)} \vdash [p] \Uparrow(c_1 \cdot c_2) : p$ and
2. $\mathsf{JPAL(K)} \vdash [p](c_1 \cdot c_2) : p$.

Proof. We use PR to denote the use of propositional reasoning. By AN for the tautology $([p]p \leftrightarrow p) \to (p \to [p]p)$ we have

$$\vdash c_1 : \big(([p]p \leftrightarrow p) \to (p \to [p]p)\big) \ . \tag{B.4}$$

By AN for the independence axiom $[p]p \leftrightarrow p$ we have

$$\vdash c_2 : ([p]p \leftrightarrow p) \ . \tag{B.5}$$

From (B.4) and (B.5) we obtain by the application axiom and PR

$$\vdash (c_1 \cdot c_2) : (p \to [p]p) \ . \tag{B.6}$$

In $\mathsf{OPAL(K)}$ we get from (B.6) by the update axiom $\Uparrow$ and PR

$$\mathsf{OPAL(K)} \vdash [p] \Uparrow(c_1 \cdot c_2) : p \ .$$

In $\mathsf{JPAL(K)}$ we get from (B.6) by the update axiom and PR

$$\mathsf{JPAL(K)} \vdash [p](c_1 \cdot c_2) : p \ . \qquad \square$$

We see that, independent of the truth value of an atomic proposition p, after p is announced, there is a reason to believe p. In $\mathsf{JPAL(K)}$ this reason is given by the term $c_1 \cdot c_2$. In $\mathsf{OPAL(K)}$ the term is $\Uparrow(c_1 \cdot c_2)$. In the latter case, the presence of $\Uparrow$ in the evidence term clearly signifies that this evidence for p is contingent on a prior public announcement. However, the exact content of such a public announcement, p in our case, is not recorded in the term. This design decision enables us to avoid the overcomplexification of the language and is similar to the introspection operation in the traditional justification logics: $!t$ is evidence for $t : A$ whenever t is evidence for A; however, the formula A is not recorded in the term $!t$.

Remark B.15. The announcement-free fragment of $\mathsf{OPAL(K)}$ and $\mathsf{JPAL(K)}$ (that is the first three axiom schemes with $\sigma = \varepsilon$, rule MP, and rule AN, restricted to $c_1 : \cdots : c_n : C$) is the justification logic J (see 5).

The internalization property is standard for justification logics, it can be proved by an easy induction on the length of derivation that it holds for OPAL(K) and JPAL(K).

Lemma B.16 (Internalization)**.** *If* $C_1, \dots, C_n \vdash A$*, then there is a term* $t(y_1, \dots, y_n)$ *for fresh variables* $y_1, \dots, y_n$ *such that*

$$y_1 : C_1, \dots, y_n : C_n \vdash t(y_1, \dots, y_n) : A \ .$$

Corollary B.17 (Constructive Necessitation)**.** *For any formula* A*, if* $\vdash A$*, then there is a ground term* t *such that* $\vdash t : A$*.*

B.3. Semantics

We adapt the Kripke-style semantics for Justification Logic due to Fitting [Fit05].

Definition B.18 (Frame)**.** A *frame* is a pair (W, R) that consists of a set $W \neq \emptyset$ of *(possible) worlds* and of an *accessibility relation* $R \subseteq W \times W$.

Definition B.19 (Evidence Function)**.** A function $\mathcal{E} \colon W \times \mathrm{Tm} \to \mathcal{P}(\mathsf{Fml_J})$ is called *evidence function* if it satisfies the following closure conditions:

1. *Axioms:* if $c : A$ is derivable by the AN-rule, then $A \in \mathcal{E}(w, c)$ for any $w \in W$.

2. *Application:* if $(A \to B) \in \mathcal{E}(w, t)$ and $A \in \mathcal{E}(w, s)$, then $B \in \mathcal{E}(w, t \cdot s)$.

3. *Sum:* $\mathcal{E}(w, s) \cup \mathcal{E}(w, t) \subseteq \mathcal{E}(w, s + t)$ for any $s, t \in \mathrm{Tm}$ and any $w \in W$.

In a model of OPAL(K) or JPAL(K), there is an evidence function $\mathcal{E}^\sigma$ for each finite sequence σ of formulas. The idea is that the evidence function $\mathcal{E}^\sigma$ models the "evidential situation" that arises after the formulas in σ have been publicly announced.

Definition B.20 (Model)**.** A *model* is a structure $\mathcal{M} = (W, R, \mathcal{E}, \nu)$, where (W, R) is a frame, $\nu : \mathsf{Prop} \to \mathcal{P}(W)$ is a *valuation*, and function $\mathcal{E}$

maps finite sequences σ of formulas to evidence functions $\mathcal{E}^\sigma$. An $\mathsf{OPAL(K)}$ *model* satisfies the following three conditions:

$$A \to [A]B \in \mathcal{E}^\sigma(w,t) \text{ implies } B \in \mathcal{E}^{\sigma,A}(w,\Uparrow t) \ , \tag{B.7}$$

$$B \in \mathcal{E}^{\sigma,A}(w,t) \text{ implies } A \to [A]B \in \mathcal{E}^\sigma(w,\Downarrow t) \ , \tag{B.8}$$

$$\mathcal{E}^{\sigma,A,B}(w,t) = \mathcal{E}^{\sigma,A\wedge[A]B}(w,t) \ . \tag{B.9}$$

A $\mathsf{JPAL(K)}$ *model* satisfies (B.9) and, instead of (B.7) and (B.8), the condition

$$A \to [A]B \in \mathcal{E}^\sigma(w,t) \text{ if and only if } B \in \mathcal{E}^{\sigma,A}(w,t) \ . \tag{B.10}$$

Conditions (B.7), (B.8), (B.9), and (B.10) correspond to the update axiom $\Uparrow$, the update axiom $\Downarrow$, the iteration axiom, and the update axiom of $\mathsf{JPAL(K)}$ respectively.

Remark B.21. Our notion of model is non-empty. The following example used in [BKS11b] and [Buc+10] gives a simple sample model for $\mathsf{OPAL(K)}$ and $\mathsf{JPAL(K)}$, respectively.

Define the structure $\mathcal{M} = (W,R,\mathcal{E},\nu)$ as follows: $W := \{w\}$, $R := \emptyset$, $\mathcal{E}^\sigma(w,t) := \mathsf{Fml_J}$ for all σ and all $t \in \mathrm{Tm}$, $\nu(p) := \{w\}$ for all $p \in \mathsf{Prop}$.

Definition B.22 (Truth in a Model)**.** A ternary relation $\mathcal{M}, w \Vdash A$ for *formula A being satisfied at a world $w \in W$ in a model $\mathcal{M} = (W,R,\mathcal{E},\nu)$* is defined by induction on the structure of A:

- $\mathcal{M}, w \Vdash p$ if and only if $w \in \nu(p)$.
- $\Vdash$ behaves classically with respect to Boolean connectives.
- $\mathcal{M}, w \Vdash t\!:\!A$ if and only if
  1) $A \in \mathcal{E}^\varepsilon(w,t)$,
  2) and $\mathcal{M}, v \Vdash A$ for all $v \in W$ with $R(w,v)$.
- $\mathcal{M}, w \Vdash [A]B$ if and only if $\mathcal{M}_A, w \Vdash B$, where the model $\mathcal{M}_A = (W_A, R_A, \mathcal{E}_A, \nu_A)$ is defined as follows:

$$\begin{aligned} W_A &:= W \ ; \\ R_A &:= \{(s,t) \mid R(s,t) \text{ and } \mathcal{M}, t \Vdash A\} \ ; \\ (\mathcal{E}_A)^\sigma &:= \mathcal{E}^{A,\sigma} \ ; \\ \nu_A &:= \nu \ . \end{aligned}$$

> Note that if $\mathcal{M}$ is an $\mathsf{OPAL(K)}$ model, then $\mathcal{M}_A$ satisfies conditions (B.7)–(B.9) from Def. B.20 and hence also is an $\mathsf{OPAL(K)}$ model. Similarly, if $\mathcal{M}$ is a $\mathsf{JPAL(K)}$ model, then $\mathcal{M}_A$ satisfies conditions (B.9) and (B.10) and hence also is a $\mathsf{JPAL(K)}$ model.

We write $\mathcal{M} \Vdash A$ to mean that $\mathcal{M}, w \Vdash A$ for all $w \in W$. We say that formula A is $\mathsf{OPAL(K)}$ *valid*, written $\mathsf{OPAL(K)} \Vdash A$, to mean that $\mathcal{M} \Vdash A$ for all $\mathsf{OPAL(K)}$ models $\mathcal{M}$. Formula A is $\mathsf{JPAL(K)}$ *valid*, written $\mathsf{JPAL(K)} \Vdash A$, if $\mathcal{M} \Vdash A$ for all $\mathsf{JPAL(K)}$ models $\mathcal{M}$.

For a sequence $\tau = (A_1, \dots, A_n)$ of formulas we denote by $\mathcal{M}_\tau = (W_\tau, R_\tau, \mathcal{E}_\tau, \nu_\tau)$ the model $(\cdots((\mathcal{M}_{A_1})_{A_2})\cdots)_{A_n}$. Note that $(\mathcal{E}_\tau)^\sigma = \mathcal{E}^{\tau,\sigma}$; in particular, $(\mathcal{E}_\tau)^\varepsilon = \mathcal{E}^\tau$.

To illustrate how the semantics works, we prove a semantic version of the result from Example B.14.

Example B.23. For any $p \in \mathsf{Prop}$ and any $c_1, c_2 \in \mathsf{Cons}$, we have

1. $\mathsf{OPAL(K)} \Vdash [p] \Uparrow(c_1 \cdot c_2) : p$ and
2. $\mathsf{JPAL(K)} \Vdash [p](c_1 \cdot c_2) : p$.

Proof. Let $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ be an arbitrary model and let $w \in W$. By Def. B.19.1, we have $([p]p \leftrightarrow p) \to (p \to [p]p) \in \mathcal{E}^\varepsilon(w, c_1)$ and $([p]p \leftrightarrow p) \in \mathcal{E}^\varepsilon(w, c_2)$. Thus, $(p \to [p]p) \in \mathcal{E}^\varepsilon(w, c_1 \cdot c_2)$ by Def. B.19.2.

Assume now that $\mathcal{M}$ is an $\mathsf{OPAL(K)}$ model. Then, by condition (B.7) from Def. B.20, we have $p \in \mathcal{E}^p(w, \Uparrow(c_1 \cdot c_2))$. Since $R_p(w, v)$ implies $\mathcal{M}, v \Vdash p$, i.e., $v \in \nu(p) = \nu_p(p)$, we have $\mathcal{M}_p, w \Vdash \Uparrow(c_1 \cdot c_2) : p$ by Def. B.22 and, hence, $\mathcal{M}, w \Vdash [p] \Uparrow(c_1 \cdot c_2) : p$.

Assume that $\mathcal{M}$ is a $\mathsf{JPAL(K)}$ model. Then, by condition (B.10) from Def. B.20, we have $p \in \mathcal{E}^p(w, c_1 \cdot c_2)$. As above we then find $\mathcal{M}_p, w \Vdash (c_1 \cdot c_2) : p$ and $\mathcal{M}, w \Vdash [p](c_1 \cdot c_2) : p$. ☐

B.4. Soundness and Completeness

The soundness proof is as usual by induction on the length of the derivation of A

Lemma B.24 (Soundness)**.** *For all formulas* $A \in \mathsf{Fml_J}$*, we have*

1. $\mathsf{OPAL(K)} \vdash A$ *implies* A *is* $\mathsf{OPAL(K)}$ *valid,*

2. $\mathsf{JPAL(K)} \vdash A$ *implies* A *is* $\mathsf{JPAL(K)}$ *valid.*

The traditional modal logic reduction approach (see Remark B.9) to establishing completeness is not possible in the presence of justifications since the replacement property does not hold in Justification Logic (see [Fit09, Section 6] for a detailed discussion of the replacement property in Justification Logic). That means, in particular, that $\vdash A \leftrightarrow B$ does not imply $\vdash t : A \leftrightarrow t : B$, which would be an essential step in the proof of a justification-analog of Lemma B.8. Thus, it is not possible to transfer the completeness of J (see [Fit05; Pac05]) to $\mathsf{OPAL(K)}$ or $\mathsf{JPAL(K)}$. We will, instead, provide a canonical model construction to prove the completeness of $\mathsf{OPAL(K)}$ and $\mathsf{JPAL(K)}$. In the following we let S stand for either $\mathsf{OPAL(K)}$ or $\mathsf{JPAL(K)}$.

Definition B.25 (Maximal S-Consistent Sets)**.** A set Φ of $\mathsf{Fml_J}$-formulas is called S-*consistent* if there is a formula that cannot be derived from Φ in S. A set Φ is called *maximal* S-*consistent* if it is consistent but has no consistent proper extensions.

It can be easily shown that maximal S-consistent sets contain all axioms of S and are closed under modus ponens and axiom necessitation.

Definition B.26 (Canonical S Model)**.** We define the *canonical* S *model* $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ as follows:

1. $W := \{w \subseteq \mathsf{Fml_J} \mid w \text{ is a maximal } \mathsf{S}\text{-consistent set}\}$,

2. $R(w, v)$ if and only if for all finite sequences σ and all $t \in \mathrm{Tm}$, we have $[\sigma]t : A \in w$ implies $[\sigma]A \in v$,

3. $\mathcal{E}^\sigma(w, t) := \{A \in \mathsf{Fml_J} \mid [\sigma]t : A \in w\}$,

4. $\nu(p) := \{w \in W \mid p \in w\}$.

To establish completeness, we need to know that the canonical model is a model.

Lemma B.27 (Correctness of the Canonical Model)**.**

1. *The canonical* $\mathsf{OPAL(K)}$ *model is an* $\mathsf{OPAL(K)}$ *model.*

2. *The canonical* $\mathsf{JPAL(K)}$ *model is a* $\mathsf{JPAL(K)}$ *model.*

Remark B.28. The canonical $\mathsf{OPAL(K)}$ model and the canonical $\mathsf{JPAL(K)}$ model are both degenerate, i.e., the canonical model consists of isolated irreflexive worlds.

Proof. $\bot \to F$ is an axiom for any F. In particular, $\vdash \bot \to [\bot]F$ is an axiom for any F. By (AN), for any F, we have $\vdash c\!:\!(\bot \to [\bot]F)$, where c is a constant. By the update axiom and the update axiom $\Uparrow$ respectively, for any formula F, we have $\vdash [\bot]s\!:\!F$ for some ground term s. Hence, for any F, the formula $[\bot]s\!:\!F$ is contained in each maximal consistent set, that is in each world of the canonical model. Let w be such a world and assume towards a contradiction that v is accessible from w. We then have by Definition B.26 that $[\bot]F \in v$ for any F. In particular, both $[\bot]F \in v$ and $[\bot]\neg F \in v$. By the functionality axiom, the latter implies $\neg[\bot]F \in v$, which contradicts $[\bot]F \in v$ since v is consistent. Thus, there cannot be a world v that is accessible from w. □

We can adapt the notion and properties of a rank from Definition B.6 and Lemma B.7, respectively, to formulae in the language of justification public announcement logic:

Definition B.29 (Rank)**.** The *rank* $\mathsf{rk}(A)$ of a formula A is defined as follows:

1. $\mathsf{rk}(p) := 1$ for each $p \in \mathsf{Prop}$
2. $\mathsf{rk}(\neg A) := \mathsf{rk}(A) + 1$
3. $\mathsf{rk}(A \to B) := \max(\mathsf{rk}(A), \mathsf{rk}(B)) + 1$
4. $\mathsf{rk}(t\!:\!A) := \mathsf{rk}(A) + 1$
5. $\mathsf{rk}([A]B) := (2 + \mathsf{rk}(A)) \cdot \mathsf{rk}(B)$

Lemma B.30 (Reductions Reduce Rank)**.** *For all formulas A, B, C and all terms s, t, we have the following:*

1. $\mathsf{rk}(A) > \mathsf{rk}(B)$ *if B is a proper subformula of A*
2. $\mathsf{rk}([A]\neg B) > \mathsf{rk}(\neg[A]B)$

3. $\mathsf{rk}([A](B \to C)) > \mathsf{rk}([A]B \to [A]C)$

4. $\mathsf{rk}([A]s:B) > \mathsf{rk}(t:(A \to [A]B))$

5. $\mathsf{rk}([A][B]C) > \mathsf{rk}([A \wedge [A]B]C)$

The following proof is done by induction on $\mathsf{rk}(D)$ and a case distinction on the structure of D.

Lemma B.31 (Truth Lemma)**.** *Let $\mathcal{M}$ be the canonical $\mathsf{OPAL(K)}$ model or the canonical $\mathsf{JPAL(K)}$ model. For all formulas D and all worlds w in $\mathcal{M}$, we have $D \in w$ if and only if $\mathcal{M}, w \Vdash D$.*

As usual, the Truth Lemma implies completeness, which, as a corollary, yields announcement necessitation.

Theorem B.32 (Completeness)**.** *For all formulas $A \in \mathsf{Fml_J}$, we have*

1. $\mathsf{OPAL(K)} \vdash A$ *if and only if A is* $\mathsf{OPAL(K)}$ *valid,*

2. $\mathsf{JPAL(K)} \vdash A$ *if and only if A is* $\mathsf{JPAL(K)}$ *valid.*

Proof. Soundness was already shown in Lemma B.24. For completeness of $\mathsf{OPAL(K)}$, consider the canonical $\mathsf{OPAL(K)}$ model $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ and assume that $\mathsf{OPAL(K)} \nvdash A$. Then $\{\neg A\}$ is consistent and, hence, contained in some maximal consistent set $w \in W$. By Lemma B.31, it follows that $\mathcal{M}, w \Vdash \neg A$ and, hence, that $\mathcal{M}, w \nVdash A$. Since $\mathcal{M}$ is an $\mathsf{OPAL(K)}$ model (Lemma B.27), we have shown that $\mathsf{OPAL(K)} \nvdash A$ implies $\mathsf{OPAL(K)} \nVdash A$. Completeness of $\mathsf{OPAL(K)}$ follows by contraposition. Completeness of $\mathsf{JPAL(K)}$ is established similarly. ☐

Corollary B.33 (Announcement Necessitation)**.** *Announcement necessitation is admissible: that is, for all formulas $A, B \in \mathsf{Fml_J}$, we have*

1. $\mathsf{OPAL(K)} \vdash A$ *implies* $\mathsf{OPAL(K)} \vdash [B]A$,

2. $\mathsf{JPAL(K)} \vdash A$ *implies* $\mathsf{JPAL(K)} \vdash [B]A$.

Proof. Assume $\mathsf{OPAL(K)} \vdash A$. By soundness, $\mathsf{OPAL(K)} \Vdash A$. Therefore, $\mathcal{M} \Vdash A$ for all $\mathsf{OPAL(K)}$ models $\mathcal{M}$. In particular, $\mathcal{M}_B, w \Vdash A$ for all $\mathsf{OPAL(K)}$ models of the form $\mathcal{M}_B$ and worlds w in them. Thus, we obtain $\mathcal{M}, w \Vdash [B]A$ for all $\mathcal{M}, w$. By completeness, we conclude $\mathsf{OPAL(K)} \vdash [B]A$. The case for $\mathsf{JPAL(K)}$ is shown similarly. ☐

B.5. Forgetful Projection and Realization

This section deals with the relationship between PAL(K) and dynamic justification logics.

Definition B.34 (Forgetful Projection)**.** The mapping $\circ : \mathsf{Fml_J} \to \mathsf{Fml}_{\Box,[\cdot]}$ is defined as follows:

- $p^\circ := p$ for all $p \in \mathsf{Prop}$,
- $\circ$ commutes with connectives $\neg$ and $\to$,
- $(t : A)^\circ := \Box A^\circ$,
- $([A]B)^\circ := [A^\circ]B^\circ$.

For a sequence $\sigma = (A_1, \ldots, A_n)$ of $\mathsf{Fml_J}$-formulas, we define σ° to be the sequence $({A_1}^\circ, \ldots, {A_n}^\circ)$ of $\mathsf{Fml}_{\Box,[\cdot]}$-formulas. In particular, $\varepsilon^\circ := \varepsilon$.

By induction on the derivation of A in JPAL(K), respectively in OPAL(K), we can show that for any theorem of either OPAL(K) or JPAL(K), its forgetful projection is a theorem of PAL(K). Note that the fact that for each sequence σ of $\mathsf{Fml_J}$-formulas, there exists a formula $U_\sigma \in \mathsf{Fml}_{\Box,[\cdot]}$ such that $\mathsf{PAL(K)} \vdash [\sigma^\circ]D \leftrightarrow [U_\sigma]D$ for any formula $D \in \mathsf{Fml}_{\Box,[\cdot]}$ simplifies the proof.

Theorem B.35 (Forgetful Projection of JPAL(K) and OPAL(K))**.** *For all formulas $A \in \mathsf{Fml_J}$,*

$$\mathsf{JPAL(K)} \vdash A \quad\Longrightarrow\quad \mathsf{PAL(K)} \vdash A^\circ \ ,$$
$$\mathsf{OPAL(K)} \vdash A \quad\Longrightarrow\quad \mathsf{PAL(K)} \vdash A^\circ \ .$$

A much more difficult question is whether a dynamic justification logic, such as JPAL(K) or OPAL(K), can realize PAL(K): that is, whether for any theorem A of PAL(K), it is possible to replace each $\Box$ in A with some term such that the resulting formula is a dynamic justification validity.

In the remainder of this section we present the first realization technique for dynamic justification logics and establish a partial realization result for JPAL(K): it can realize formulas A that do not contain $\Box$ operators within announcements. Our main idea is to reduce realization of PAL(K) to realization of K. In our proof, we rely on notions and techniques introduced by Fitting [Fit09].

Definition B.36 (Substitution)**.** A *substitution* is a mapping from variables to terms. If A is a formula and σ is a substitution, we write $A\sigma$ to denote the result of simultaneously replacing each variable x in A with the term $x\sigma$.

The following lemma is standard in justification logics and can be proved by a simple induction on the derivation of A.

Lemma B.37 (Substitution Lemma)**.** *For every formula A of* $\mathsf{Fml_J}$ *and every substitution σ,*

$$\mathsf{JPAL(K)} \vdash A \textit{ implies } \mathsf{JPAL(K)} \vdash A\sigma,$$
$$\mathsf{OPAL(K)} \vdash A \textit{ implies } \mathsf{OPAL(K)} \vdash A\sigma.$$

In most justification logics, in addition to this substitution of proof terms for proof variables, the substitution of formulas for propositions is also possible (see [Art08]). However, the latter type of substitution typically fails in logics with public announcements, as it does in both $\mathsf{JPAL(K)}$ and $\mathsf{OPAL(K)}$.

Definition B.38 (Annotations)**.** An *annotated formula* is an formula $A \in \mathsf{Fml}_{\Box,[\cdot]}$ in which each modal operator is annotated by a natural number. An annotated formula is *properly annotated* if modalities in negative positions are annotated with even numbers, modalities in positive positions are annotated with odd numbers, and no index i annotates two modality occurrences. Positions within an announcement $[A]$ are considered neither positive nor negative: i.e., the parity of indices within announcements in properly annotated formulas is not regulated. If A' is the result of replacing all indexed modal operators $\Box_i$ with $\Box$ in a (properly) annotated formula A, then A is called a *(properly) annotated version of A'*.

Definition B.39 (Realization Function)**.** A *realization function* r is a mapping from natural numbers to terms such that $r(2i) = x_i$, where $x_1, x_2, \dots$ is a fixed enumeration of all variables. For a realization function r and an annotated formula A, $r(A)$ denotes the result of replacing each indexed modal operator $\Box_i$ in A with the term $r(i)$. For instance, $r(\Box_i B) = r(i) : r(B)$. A realization function r is called *non-self-referential on variables over A* if, for each subformula $\Box_{2i} B$ of A, the variable $x_i = r(2i)$ does not occur in $r(B)$.

The following realization result for the logic K is due to Brezhnev [Bre00]; the additional result about non-self-referentiality on variables follows from the stronger statement that K can be realized without any self-referential cycles of arbitrary terms, proved in [Kuz10].

Theorem B.40 (Realization for K)**.** *If A' is a theorem of K, then for any properly annotated version A of A', there is a realization function r that is non-self-referential on variables over A and such that $r(A)$ is provable in J. Clearly, $(r(A))^\circ = A'$.*

In order to formulate the replacement theorem for $\mathsf{JPAL}(\mathsf{K})$, a technical result necessary for demonstrating the partial realization theorem for $\mathsf{JPAL}(\mathsf{K})$, we use the following standard convention: whenever $D(q)$ and A are formulas in the same language, $D(A)$ is the result of replacing all occurrences of the proposition q in $D(q)$ with A. In most cases, q has only one occurrence in $D(q)$ that is not within announcements.

For the rest of this section, we consider only formulas $A \in \mathsf{Fml}_{\Box,[\cdot]}$ and their annotated versions that do not contain modal operators within announcements: i.e., if $[B]C$ is a subformula of A, then B does not contain modal operators.

We will use a theorem that was first proved by Fitting [Fit09] for replacement in positive positions in LP. We use his method in a richer language and for a different logic and also use replacement in both positive and negative positions. See [BKS12b] for a full proof of the following theorem:

Theorem B.41 (Restricted Realization Modification for $\mathsf{JPAL}(\mathsf{K})$)**.** *Assume the following:*

H-1. *A proposition p has exactly[1] one occurrence in a properly annotated formula $X(p)$ that is outside of announcements and $X(A)$ and $X(B)$ are properly annotated formulas with no modalities within announcements.*

H-2. *r_1 is a realization function, non-self-referential on variables over $X(A)$.*

[1] While the proof of this theorem does not depend on whether p actually occurs in $X(p)$, the formulation of H-1 is simpler when it does.

H-3. *If p occurs positively in $X(p)$,* $\mathsf{JPAL(K)} \vdash r_1(A) \to r_1(B)$.
If p occurs negatively in $X(p)$, $\mathsf{JPAL(K)} \vdash r_1(B) \to r_1(A)$.

Then for each subformula $\phi(p)$ of $X(p)$ that occurs outside of announcements, there exists some realization/substitution pair $\langle r_\phi, \sigma_\phi \rangle$ such that:

C-1. $\mathsf{JPAL(K)} \vdash r_1(\phi(A))\sigma_\phi \to r_\phi(\phi(B))$ *if $\phi(p)$ occurs positively in $X(p)$.*
$\mathsf{JPAL(K)} \vdash r_\phi(\phi(B)) \to r_1(\phi(A))\sigma_\phi$ *if $\phi(p)$ occurs negatively in $X(p)$.*

C-2. *σ_ϕ lives on input positions in $\phi(p)$, i.e., $x_i\sigma_\phi = x_i$ if $\Box_{2i}$ does not occur in $\phi(p)$;*

C-3. *σ_ϕ meets the no new variable condition, i.e., the only variable that may occur in $x\sigma_\phi$ is x;*

C-4. *If r_1 is non-self-referential on variables over $X(B)$, then r_ϕ is also non-self-referential on variables over $X(B)$.*

The generality of the theorem above is needed to carry through the induction. However, to prove realization, we will only need the following weaker version:

Corollary B.42 (Replacement for $\mathsf{JPAL(K)}$)**.** *Assume the following:*

1. *A proposition p has exactly one occurrence in a properly annotated formula $X(p)$ that is outside of announcements. $X(A)$ and $X(B)$ are properly annotated formulas with no modalities within announcements.*

2. *r_1 is a realization function that is non-self-referential on variables over $X(A)$ and over $X(B)$.*

3. *If p occurs positively in $X(p)$,* $\mathsf{JPAL(K)} \vdash r_1(A) \to r_1(B)$.
If p occurs negatively in $X(p)$, $\mathsf{JPAL(K)} \vdash r_1(B) \to r_1(A)$.

Then there exists a realization function r and a substitution σ such that

$$\mathsf{JPAL(K)} \vdash r_1(X(A))\sigma \to r(X(B))$$

and r is non-self-referential on variables over $X(B)$.

It remains to extend the notions of one-step reduction and of reduction to annotated modal formulas (with announcements). To achieve this it is sufficient to replace the 4th row in the table in (B.3) by

Redex	Its reduct
$[A]\Box_i B$	$\Box_i(A \to [A]B)$

(B.11)

The functions red_1 and red for annotated formulas are defined the same way as in Definition B.5 but based on the new set of reductions. Accordingly, $\mathsf{red}_1(A)$ is an annotated formula and $\mathsf{red}(A)$ is an annotated formula without announcements whenever A is an annotated formula.

Since the only difference in how red_1 works on $\mathsf{Fml}_{\Box,[\cdot]}$-formulas and on annotated formulas is such that erasing annotations in a pair of redex/reduct of the annotated red_1 yields a pair of redex/reduct of the unannotated red_1, the following lemma is not very surprising:

Lemma B.43. *Let D be a properly annotated variant of $D' \in \mathsf{Fml}_{\Box,[\cdot]}$ and let neither one contain modalities within announcements.*[2] *Then* $\mathsf{red}_1(D)$ *and* $\mathsf{red}(D)$ *are properly annotated variants of* $\mathsf{red}_1(D')$ *and* $\mathsf{red}(D')$ *respectively, neither* $\mathsf{red}_1(D)$ *nor* $\mathsf{red}_1(D')$ *contains modalities within announcements, and neither* $\mathsf{red}(D)$ *nor* $\mathsf{red}(D')$ *contains announcements.*

Proof. That $\mathsf{red}_1(D)$ is an annotated version of $\mathsf{red}_1(D')$ and, hence, $\mathsf{red}(D)$ is an annotated version of $\mathsf{red}(D')$ is clear from the definition. Thus, it remains to prove that red_1 preserves the properness of annotations and the property of not having modalities within announcements. We will only show this for the pair of redex/reduct from (B.11). The other four cases are even simpler. Let $D = X([A]\Box_i B)$ for some $X(p)$ and red_1 maps it to $X(\Box_i(A \to [A]B))$. By assumption, A contains no modalities and p is outside of announcements in $X(p)$. Hence, after the replacement $\Box_i$ and the modalities in B remain outside of announcements. Further, all annotations in $\Box_i(A \to [A]B)$, i.e., i and all annotations in B, do not occur in $X(p)$ because D is properly annotated. The duplication of A does not violate the properness of annotation because A is modality-free. ∎

We now have all the ingredients sufficient to establish our realization theorem. The following diagram shows how we obtain it. We start with

[2] If one does not contain them, then the other does not either.

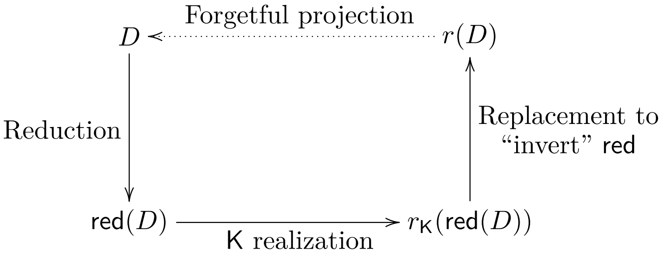

Figure B.1.: Realizing $\mathsf{PAL}(\mathsf{K})$

a formula $D' \in \mathsf{Fml}_{\Box,[\cdot]}$. Taking its arbitrary properly annotated version D, we use annotated reduction from Lemma B.43, K realization from Theorem B.40, and replacement from Corollary B.42 to construct a formula $r(D) \in \mathsf{Fml}_{\mathsf{J}}$ that realizes D. It is easy to see that $(r(D))^\circ = D'$.

Theorem B.44 (Realization for $\mathsf{PAL}(\mathsf{K})$)**.** *Let D' be a theorem of $\mathsf{PAL}(\mathsf{K})$ that does not contain modalities within announcements. Then for any properly annotated version D of D', there is a realization function r such that $r(D)$ is provable in $\mathsf{JPAL}(\mathsf{K})$ and $(r(D))^\circ = D'$.*

Proof. Let D be a properly annotated version of D'. Clearly, D contains no modalities within announcements. By Lemma B.43, there exists a sequence of properly annotated formulas $D_0, D_1, \dots, D_N$ that do not contain modalities within announcements and a sequence of $\mathsf{Fml}_{\Box,[\cdot]}$-formulas $D'_0, D'_1, \dots, D'_N$ that do not contain modalities within announcements such that $D_0 = D, D'_0 = D'$, $D_{i+1} = \mathsf{red}_1(D_i)$ and $D'_{i+1} = \mathsf{red}_1(D'_i)$ for each $i = 0, \dots, N-1$, $D_N = \mathsf{red}(D)$, $D'_N = \mathsf{red}(D')$, neither D_N nor D'_N contains announcements, and D_i is a properly annotated version of D'_i for each $i = 0, \dots, N$. By Lemma B.8, $\mathsf{PAL}(\mathsf{K}) \vdash D' \leftrightarrow D'_N$. Hence, D'_N is also a theorem of $\mathsf{PAL}(\mathsf{K})$. Given that D'_N contains no announcements, it is also a theorem of K due to the conservativity of $\mathsf{PAL}(\mathsf{K})$ over K. By Theorem B.40, there exists a realization function r_N that is non-self-referential on variables over the properly annotated version D_N of D'_N such that $r_N(D_N)$ is provable in J and, hence, in its extension $\mathsf{JPAL}(\mathsf{K})$. We now construct realization functions r_{N-i}, non-self-referential on variables over D_{N-i}, such that

$r_{N-i}(D_{N-i})$ is a theorem of $\mathsf{JPAL}(\mathsf{K})$ for $i = 0, \dots, N$ by induction on i. The base case is already established. Let $r_{N-(i-1)} = r_{N-i+1}$ be already constructed. Since $D_{N-i+1} = \mathsf{red}_1(D_{N-i})$, it follows that $D_{N-i} = X(\mathit{Redex})$ for some $X(p)$ with exactly one occurrence of a fresh proposition p whereas $D_{N-i+1} = X(\mathit{Reduct})$, where *Redex* is the outermost leftmost redex in D_{N-i} and *Reduct* is the reduct of *Redex*. We want to apply Corollary B.42 to $X(p)$, *Reduct*, *Redex*, and r_{N-i+1}. Assumption 1 is satisfied because D_{N-i} and D_{N-i+1} are properly annotated and by definition of red_1 (recall also that an outermost redex never occurs within announcements). That r_{N-i+1} is non-self-referential on variables over $X(\mathit{Reduct}) = D_{N-i+1}$ follows from the induction hypothesis. Looking at the five types of redexes and their reducts, it is easy to check that r_{N-i+1} is also non-self-referential on variables over $X(\mathit{Redex})$ because the substitution of *Redex* for *Reduct* at p in $X(p)$ never introduces new modalities and never moves modalities into the scope of other modalities (recall that announcements contain no modalities). Hence, assumption 2 is satisfied. It remains to note that $r(\mathit{Redex}) \leftrightarrow r(\mathit{Reduct})$ is one of the axioms of $\mathsf{JPAL}(\mathsf{K})$[3] for any realization function r, including r_{N-i+1}, so that assumption 3 is also satisfied, independent of the polarity of p in $X(p)$. By Corollary B.42, there exists a realization function r_{N-i}, non-self-referential on variables over $X(\mathit{Redex}) = D_{N-i}$, and a substitution σ_i such that $\mathsf{JPAL}(\mathsf{K}) \vdash r_{N-i+1}(X(\mathit{Reduct}))\sigma_i \to r_{N-i}(X(\mathit{Redex}))$. In other words, $\mathsf{JPAL}(\mathsf{K}) \vdash r_{N-i+1}(D_{N-i+1})\sigma_i \to r_{N-i}(D_{N-i})$. It remains to use the induction hypothesis, the Substitution Property, and MP to see that $r_{N-i}(D_{N-i})$ is a theorem of $\mathsf{JPAL}(\mathsf{K})$. In particular, $r_{N-N}(D_{N-N}) = r_0(D)$ is provable in $\mathsf{JPAL}(\mathsf{K})$. Set $r := r_0$. Clearly, $(r(D))^\circ = D'$. $\square$

Remark B.45. It is not clear how to generalize our proof to theorems of $\mathsf{PAL}(\mathsf{K})$ with modalities allowed within announcements. The problem is that a reduct $[\Box A]C \to [\Box A]D$ has two copies of $\Box A$, which need to be combined into only one copy in the corresponding redex $[\Box A](C \to D)$. In general, the outer $\Box$ in $\Box A$'s in the reduct will be realized by different terms, and we currently lack methods of merging terms within announcements.

[3]This is not the case in $\mathsf{OPAL}(\mathsf{K})$ for redex/reduct pairs from (B.11).

Remark B.46. Adapting this proof to OPAL(K) presents certain challenges. The problem is that in order to 'invert' the reduction from PAL(K) to K, we need to apply replacement also in negative positions. This is only possible because in the update axiom (update) of JPAL(K), we have the same evidence term on both sides of the equivalence. If, like in OPAL(K), we work with update operations $\Uparrow$ and $\Downarrow$ on terms, then we end up with different terms in the update axioms, which prevents the use of Fitting's replacement at negative positions.

B.6. Notes

We have presented JPAL(K) and OPAL(K), two alternative justification logic counterparts of Gerbrandy–Groeneveld's modal public announcement logic PAL(K). One of PAL(K)'s update principles is

$$\Box(A \to [A]B) \to [A]\Box B \ ,$$

which we render in JPAL(K) as

$$s : (A \to [A]B) \to [A]s : B$$

and in OPAL(K) as

$$s : (A \to [A]B) \to [A] \Uparrow s : B \ .$$

For the semantics, we have used a combination of the traditional semantics for public announcement logic (where an agent rejects as impossible the worlds that are inconsistent with the announcement made) and evidence functions from epistemic models for justification logic (that specify for each world which formulas an evidence term can justify). We then have shown soundness and completeness (by a canonical model construction) for JPAL(K) and OPAL(K).

The main result is a realization theorem stating that JPAL(K) realizes all the theorems of PAL(K) that do not contain modalities within announcements. To obtain this result we have to extend Fitting's replacement theorem such that, first, it works in the context of public announcements and, second, it allows replacement also in negative positions.

Finally, it should be noted that our novel realization method does not rely on a cut-free deductive system for $\mathsf{PAL}(\mathsf{K})$. Its constructiveness, however, depends on constructive realization for the modal logic K, to which we reduce $\mathsf{PAL}(\mathsf{K})$.

C. The Road to Realization?

In Chapter 6, Section 6.6, we have mentioned that the non-standard behavior of the canonical model prevents an application of Fitting's semantical realization method. In this chapter, we outline how such a proof would work and indicate what part causes problems. We also discuss possible solutions.

In order to facilitate notations in the following part, let us use the following convention: by $\circledast A$ we mean the (modal) formula

- $\Box_i A$, if $\circledast = i$,
- $\mathsf{E}A$, if $\circledast = \mathsf{E}$,
- $\mathsf{C}A$, if $\circledast = \mathsf{C}$.

C.1. Realizations and the Alternative Canonical Model

Definition C.1. For a given constant specification $\mathcal{CS}$, the *alternative canonical model* $\mathcal{M}' = (W, R_1, \dots, R_h, R'_{\mathsf{E}}, R'_{\mathsf{C}}, \mathcal{E}, \nu)$ meeting $\mathcal{CS}$ is defined as the canonical model meeting $\mathcal{CS}$ from Definition 6.14 but with R_{E} and R_{C} replaced by R'_{E} and R'_{C} as defined in Remark 6.16, i.e.,

$$(w, v) \in R'_{\mathsf{E}} \quad \text{if and only if} \quad w/\mathsf{E} \subseteq v$$

and

$$(w, v) \in R'_{\mathsf{C}} \quad \text{if and only if} \quad w/\mathsf{C} \subseteq v \ .$$

Remark C.2. While by Remark 6.16, the alternative canonical model is not an actual model, the Truth Lemma 6.17 nevertheless holds, i.e., for all formulae A and all worlds $w \in W$,

$$A \in w \text{ if and only if } \mathcal{M}', w \Vdash A \ .$$

The proof for this fact is even simpler for $\mathcal{M}'$, as the cases for mutual and common knowledge now can be treated in the same way as individual knowledge.

We first need the following simple auxiliary lemma.

Lemma C.3. *Let $\mathcal{CS}$ be a homogeneous, C-axiomatically appropriate constant specification, w a maximally $\mathcal{CS}$-consistent set and $B \in \mathrm{Fm}_{\mathsf{LP}^{\mathsf{C}}_h}$. If $w/\circledast \cup \{\neg B\}$ is not consistent, then there is a term $t \in \mathrm{Tm}_\circledast$ such that $[t]_\circledast B \in w$.*

Proof. We prove the contrapositive. So assume (towards a contradiction), that $[t]_\circledast B \notin w$ for every term $t \in \mathrm{Tm}_\circledast$, but $w/\circledast \cup \{\neg B\}$ is not consistent. Then there are formulae $A_1, \ldots, A_n \in w/\circledast$ such that $\{A_1, \ldots, A_n, \neg B\}$ is not consistent, i.e. $\mathsf{LP}^{\mathsf{C}}_h(\mathcal{CS}) \vdash A_1 \to (A_2 \to (\ldots \to (A_n \to B)))$. By constructive necessitation (Corollary 6.6), there is a ground term $t \in \mathrm{Tm}_\circledast$ such that $\mathsf{LP}^{\mathsf{C}}_h(\mathcal{CS}) \vdash [t]_\circledast(A_1 \to (A_2 \to (\ldots \to (A_n \to B))))$. As $A_1, \ldots, A_n \in w/\circledast$, there are terms $s_1, \ldots, s_n \in \mathrm{Tm}_\circledast$ such that $[s_j]_\circledast A_j \in w$ for $j = 1 \ldots n$. Using the application axiom and maximal consistency of w we then get $[(t \cdot s_1) \cdot s_2) \cdot \ldots) \cdot s_n)]_\circledast \in w$, contradicting our assumption that no such formula is in w. □

For the rest of this chapter, let us fix a modal formula $A \in \mathrm{Fm}_{\mathsf{S4}^{\mathsf{C}}_h}$.

Definition C.4. We assign polarities to subformulae of A as follows in the usual way.

- A itself is a positive subformula of A.
- If $\circledast B$ is a positive (negative) subformula of A, then B is also a positive (negative) subformula of A.
- If $B \to C$ is a positive (negative) subformula of A, then B is a negative (positive) subformula of A and C is a positive (negative) subformula of A.
- If $\neg B$ is a positive (negative) subformula of A, then B is a negative (positive) subformula of A.

Definition C.5. An annotation for A is a mapping $\mathcal{A}$ assigning proof variables to negatively occuring subformulae of the form $\circledast B$.

An annotation is called proper, if

1. different occurences are assigned different proof variables, and
2. types of modalities and proof variables match, e.g. a C-modality is assigned a proof variable from $\mathrm{Var}_{\mathsf{C}}$. Formally $\mathcal{A}(\circledast B) \in \mathrm{Var}_{\circledast}$ for a negative subformula occurrence $\circledast B$ of A.

Let us also fix a proper annotation $\mathcal{A}$ for A for the rest of this chapter. We will now define a mapping that assigns a set of "potential pre-realizations" to subformula occurrences of A with respect to the annotation $\mathcal{A}$.

Definition C.6. The mapping

$$|.| : \{B \text{ is a subformula occurrence of } A\} \to \mathcal{P}(\mathrm{Fm}_{\mathsf{LP}^{\mathsf{C}}_h})$$

from subformulae occurrences of A to sets of formulae is defined inductively as follows

- $|P| := \{P\}$,
- $|\neg B| := \{\neg B' \mid B' \in |B|\}$,
- $|B \to C| := \{B' \to C' \mid B' \in |B| \text{ and } C' \in |C|\}$,
- if $\circledast B$ is a negative subformula occurrence, then

$$| \circledast B| := \{[x]_{\circledast} B' \mid \mathcal{A}(\circledast B) = x \in \mathrm{Var}_{\circledast} \text{ and } B' \in |B|\},$$

- if $\circledast B$ is a positive subformula occurrence, then

$$| \circledast B| := \{[t]_{\circledast}(B'_1 \vee \ldots \vee B'_n) \mid t \in \mathrm{Tm}_{\circledast} \text{ and } B'_1, \ldots, B'_n \in |B|\}.$$

We use $\neg|B|$ to denote the set $\{\neg B' \mid B' \in |B|\}$.

Remark C.7. We can consider $\mathcal{M}'$ as a Kripke model by simply ignoring the evidence function. This model also suffers the defect that $R_{\mathsf{C}} \subsetneq R'_{\mathsf{C}}$ as stated in Remark 6.16. Even though it is always clear from the context whether $\Vdash$ is used with respect to Fitting semantics or Kripke semantics (i.e. with respect to formulae of justification logic or modal formulae), we will sometimes write $\Vdash_{\mathsf{LP}^{\mathsf{C}}_h}$ and $\Vdash_{\mathsf{S4}^{\mathsf{C}}_h}$, respectively, in order to emphasize this point.

Lemma C.8. *Let $\mathcal{CS}$ be a homogeneous, C-axiomatically appropriate constant specification and $\mathcal{M}'$ the alternative canoncial model meeting $\mathcal{CS}$.*

1. *If B is a positive subformula occurrence of A and $\mathcal{M}', w \Vdash_{\mathsf{LP}^{\mathsf{C}}_h} \neg|B|$, then $\mathcal{M}', w \nVdash_{\mathsf{S4}^{\mathsf{C}}_h} B$.*

2. *If B is a negative subformula occurrence of A and $\mathcal{M}', w \Vdash_{\mathsf{LP}^{\mathsf{C}}_h} |B|$, then $\mathcal{M}', w \Vdash_{\mathsf{S4}^{\mathsf{C}}_h} B$.*

Proof. We proceed by induction on the complexity of the subformula B of A.

- The case for atomic subformulae is trivial as we have $|P| = \{P\}$.

- Suppose $\neg B$ is a subformula occurring positively and $\mathcal{M}', w \Vdash \neg|\neg B|$. This means, for each $B' \in |B|$ we have $\mathcal{M}', w \Vdash \neg(\neg B')$. As B is a negatively occuring subformula, we get by induction hypothesis $\mathcal{M}', w \Vdash B$ and thus $\mathcal{M}' \nVdash \neg B$.

- Suppose $\neg B$ is a negative subformula occurrence and $\mathcal{M}', w \Vdash |\neg B|$. So, we have $\mathcal{M}' \Vdash \neg B'$ for all $B' \in |B|$. As B is a positive subformula occurrence, we get $\mathcal{M}' \Vdash \neg|B|$ and so, by induction hypothesis $\mathcal{M}', w \nVdash B$, i.e. $\mathcal{M}', w \Vdash \neg B$.

- Suppose $B \to C$ is a positively occuring subformula and $\mathcal{M}', w \Vdash \neg|B \to C|$. This means, we have $\mathcal{M}', w \Vdash \neg(B' \to C')$ for any $B' \in |B|$ and $C' \in |C|$. So, we have $\mathcal{M}', w \Vdash B'$ for any $B' \in |B|$ and $\mathcal{M}', w \Vdash \neg C'$ for any $C' \in |C|$, i.e. $\mathcal{M}', w \Vdash |B|$ and $\mathcal{M}', w \Vdash \neg|C|$. As B occurs negatively and C occurs positively, by the induction hypothesis we get $\mathcal{M}', w \Vdash B$ and $\mathcal{M}', w \nVdash C$ and thus $\mathcal{M}', w \nVdash B \to C$.

- Suppose $B \to C$ is a negatively occuring subformula and $\mathcal{M}', w \Vdash |B \to C|$. We will distinguish two cases. If $\mathcal{M}', w \Vdash \neg|B|$, then, as B occurs positively, we get $\mathcal{M}', w \nVdash B$ by the induction hypothesis and thus also $\mathcal{M}', w \Vdash B \to C$ and we are done. Otherwise, there is a $B' \in |B|$ such that $\mathcal{M}', w \Vdash B'$ and, as we have $\mathcal{M}', w \Vdash B' \to C'$ for any $C' \in |C|$ by assumption, we get $\mathcal{M}', w \Vdash C'$ for any $C' \in |C|$, i.e. $\mathcal{M}', w \Vdash |C|$. As C occurs negatively, we then get $\mathcal{M}', w \Vdash C$ by induction hypothesis and thus also $\mathcal{M}', w \Vdash B \to C$.

- Suppose $\circledast B$ occurs positively and $\mathcal{M}', w \Vdash \neg|\circledast B|$. We are done, if we can show

$$w/\circledast \cup \neg|B| \text{ is consistent.} \tag{C.1}$$

  For, then there must be a $v \in W$ such that $w/\circledast \cup \neg|B| \subseteq v$, as v is a maximal consistent set. By definition of $R_\circledast$, this means $R_\circledast(w, v)$ and, as in particular $\neg|B| \subseteq v$, we get $\mathcal{M}', v \Vdash \neg|B|$ by the Truth Lemma for the alternative canoncial model C.2. As B also occurs positively, by induction hypothesis we get $\mathcal{M}', v \nVdash B$.

  Let us now show the open claim (C.1). Assume (towards a contradiction), that $w/\circledast \cup \neg|B|$ is not consistent. Then, by Lemma C.3 there is a term $t \in \mathrm{Tm}_\circledast$ and $B_1, \ldots, B_n \in |B|$ such that $[t]_\circledast(B_1 \vee \ldots \vee B_n) \in w$ and so, by the Truth Lemma C.2 we have $\mathcal{M}', w \Vdash [t]_\circledast(B_1 \vee \ldots \vee B_n)$, contradicting the assumption $\mathcal{M}', w \Vdash \neg|\circledast B|$.

- Supppose $\circledast B$ occurs negatively and $\mathcal{M}', w \Vdash |\circledast B|$. Let $B' \in |B|$ and $\mathcal{A}(\circledast B) = x \in \mathrm{Var}_\circledast$. We have $\mathcal{M}', w \Vdash [x]_\circledast B'$ and so $\mathcal{M}', v \Vdash B'$ for any $v \in W$ with $R_\circledast(w, v)$. As B' was chosen arbitrarily, we get $\mathcal{M}', v \Vdash |B|$ and by induction hypothesis we obtain $\mathcal{M}', v \Vdash B$ for all $v \in W$ with $R_\circledast(w, v)$. Finally, we obtain $\mathcal{M}', w \Vdash \circledast B$. □

C.2. Where the Problem is

Following the usual semantical realization method, we would now use the previous lemma to prove a statement of the form

> Let $\mathcal{CS}$ be a homogeneous, C-axiomatically appropriate constant specification. If $\mathsf{S4}^{\mathsf{C}}_h \vdash A$, then there are $A_1, \ldots, A_n \in |A|$ such that $\mathsf{LP}^{\mathsf{C}}_h(\mathcal{CS}) \vdash A_1 \vee \ldots \vee A_n$. (∗)

This is usually done by proving the contrapositive: Assume for any $A_1, \ldots, A_n \in |A|$, we have

$$\mathsf{LP}^{\mathsf{C}}_h(\mathcal{CS}) \nvdash A_1 \vee \ldots \vee A_n.$$

Then $\neg|A|$ is consistent and so there is a world $w \in W$ in $\mathcal{M}'$ such that $\neg|A| \subseteq w$. By the Truth Lemma C.2 we get

$$\mathcal{M}', w \Vdash_{\mathsf{LP}^{\mathsf{C}}_h(\mathcal{CS})} \neg|A|.$$

As A is a positive subformula of itself, we can use Lemma C.8 to conclude

$$\mathcal{M}', w \nVdash_{\mathsf{S4}^{\mathsf{C}}_h} A.$$

However, this is where the problems start. As noted in Remark C.7, $\mathcal{M}'$ also is not in the right class of Kripke models, as we have $R_{\mathsf{C}} \subsetneq R'_{\mathsf{C}}$ and thus we can not use soundness to conclude

$$\mathsf{S4}^{\mathsf{C}}_h \nvdash A.$$

The usual approach in modal logic would be to use filtrations to obtain a model of the right form. However, this fails as we have maximal $\mathsf{LP}^{\mathsf{C}}_h$-consistent sets and not maximal $\mathsf{S4}^{\mathsf{C}}_h$-consistent sets and there is no obvious relationship between such maximal consistent sets. In particular, it is not clear whether the modal inducation axiom holds for the alternative canonical model. Another approach would be to use filtrations for justifications logics as introduced in Chapter A. For now, this remains work in progress and will need more model theoretic insights. For instance, filtrations will have to be set up in a way that preserves validity for certain forgetful projections. Furthermore, the lack of characteristic formulae (see Chapter A, Section A.2) prevents a simple adapation of the usual filtration techniques for models with non-standard behaviour (see [MH95; HKJ00]).

C.3. How the Proof Would Continue

Remarkably, the rest of the proof would work without any further problems. Let us outline the path. The next step is to look at the set of "potential realizations" and it remains to be shown, that the previously defined "potential pre-realizations" can be turned into such actual "potential realizations".

Definition C.9. The mapping

$$\|.\| : \{B \text{ is a subformula occurrence of } A\} \to \mathcal{P}(\mathrm{Fm}_{\mathsf{LP}^{\mathsf{C}}_h})$$

is defined as $|.|$ with $\|.\|$ in place of $|.|$ everywhere, except for the case of $\circledast B$ being a positive subformula of A, where we set

$$\| \circledast B\| := \{[t]_{\circledast} B' \mid B' \in \|B\| \text{ and } t \in \mathrm{Tm}_{\circledast}\}.$$

Definition C.10. 1. A substitution is a mapping $\sigma : \mathrm{Var} \to \mathrm{Tm}$ such that $\sigma(x) \in \mathrm{Tm}_{\circledast}$ whenever $x \in \mathrm{Var}_{\circledast}$. Substitutions can be extended inductively to terms and formulae in the obvious way. For a formula $C \in \mathrm{Fm}_{\mathsf{LP}^{\mathsf{C}}_h}$ we denote by $C\sigma$ the formula obtained by simultaneously replacing all occurrences of x by $\sigma(x)$ in C. For a set of formulae $\Phi \subseteq \mathrm{Fm}_{\mathsf{LP}^{\mathsf{C}}_h}$, we set $\Phi\sigma := \{A\sigma \mid A \in \Phi\}$.

2. The domain of a substitution σ is defined as

$$\mathrm{dom}(\sigma) := \{x \mid \sigma(x) \neq x\}.$$

3. A substitution σ with finite domain $\mathrm{dom}(\sigma) = \{x_1, \ldots, x_n\}$ will also be denoted by $\sigma = \{x_1/t_1, \ldots, x_n/t_n\}$ where $\sigma(x_j) = t_j$.

4. For two substitutions σ_1 and σ_2 with $\mathrm{dom}(\sigma_1) \cap \mathrm{dom}(\sigma_2) = \emptyset$ we define $\sigma_1 \cup \sigma_2$ by

$$(\sigma_1 \cup \sigma_2)(x) := \begin{cases} \sigma_1(x) & \text{if } x \in \mathrm{dom}(\sigma_1), \\ \sigma_2(x) & \text{if } x \in \mathrm{dom}(\sigma_2), \\ x & \text{otherwise.} \end{cases}$$

5. A substitution σ is said to live on a formula $C \in \mathrm{Fm}_{\mathsf{LP}^{\mathsf{C}}_h}$, if

$$\mathrm{dom}(\sigma) \subseteq \{x \mid x \text{ is a proof variable} \\ \text{occurring in negative position in } C\}.$$

6. A substitution σ meets the *no new variable* condition, if $\sigma(x)$ contains no variables except for x.

Remark C.11. A substitution σ is said to live away from a formula C, if all variables in its domain do not occur in negative position in C. If we have two substitutions σ_1 and σ_2 such that σ_1 lives on some formula C and σ_2 lives away from the same formula C and both meet the no new variable condition, we have $\sigma_1\sigma_2 = \sigma_2\sigma_1 = \sigma_1 \cup \sigma_2$. However, if σ lives away from C, this does not mean that $C\sigma = C$, as variables in the domain of σ might occur in positive position in C. This fact is important in the cases for implications and in the case for negatively occurring modalities in the following proof.

Lemma C.12. *Let $\mathcal{CS}$ be a homogeneous, C-axiomatically appropriate, schematic constant specification. For every subformula B of A and for all $B_1, \ldots, B_n \in |B|$ there is a formula $B' \in \|B\|$ and a substitution σ living on B' and meeting the no new variable condition such that*

1. *If B is a positive subformula occurrence of A, then*

$$\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash (B_1 \vee \ldots \vee B_n)\sigma \to B'.$$

2. *If B is a negative subformula occurrence of A, then*

$$\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash B' \to (B_1 \wedge \ldots \wedge B_n)\sigma.$$

Proof. We will proceed by induction on the complexity of the subformula B.

- If B is the atomic proposition P, then the statement is trivial, as $|B| = \|B\| = \{P\}$ and thus we can use the empty substitution (i.e. the identity function).

- Suppose $\neg B$ is a positively occurring subformula of A and let $\neg B_1, \ldots, \neg B_n \in |\neg B|$. Then B occurs negatively, $B_j \in |B|$ for $j = 1 \ldots n$, and by induction hypothesis there is $B' \in \|B\|$ and a substitution σ with the necessary properties such that $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash B' \to (B_1 \wedge \ldots \wedge B_n)\sigma$. Using the contrapositive, we get $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash (\neg B_1 \vee \ldots \vee \neg B_n)\sigma \to \neg B'$. As $\neg B' \in \|\neg B\|$ and σ satisfies the necessary properties by induction hypothesis, we are done.

- The case for $\neg B$ being a negatively occurring subformula is symmetric to the previous case and therefore omitted.

- Suppose $B \to C$ is a positive subformula occurrence and let $B_1 \to C_1, \ldots, B_n \to C_n \in |B \to C|$. As B occurs negatively, C

occurs positively, $B_j \in |B|$, and $C_j \in |C|$ for $j = 1 \ldots n$, there are $B' \in \|B\|$ and σ_B as well as $C' \in \|C\|$ and σ_C such that $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash B' \to (B_1 \wedge \ldots \wedge B_n)\sigma_B$ and $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash (C_1 \vee \ldots \vee C_n)\sigma_C \to C'$. As $\mathcal{CS}$ is schematic, we then also have $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash (B' \to (B_1 \wedge \ldots \wedge B_n)\sigma_B)\sigma_C$ and $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash ((C_1 \vee \ldots \vee C_n)\sigma_C \to C')\sigma_B$. As σ_B lives on B' and σ_C lives on C', both are different subformulae occurrences of A and $\mathcal{A}$ is a proper annotation, we have $\mathrm{dom}(\sigma_B) \cap \mathrm{dom}(\sigma_C) = \emptyset$ and we can thus define $\sigma := \sigma_B \cup \sigma_C$. We now can write the previous as $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash B'\sigma_C \to (B_1 \wedge \ldots \wedge B_n)\sigma$ and $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash (C_1 \vee \ldots \vee C_n)\sigma \to C'\sigma_B$. Using propositional reasoning, we obtain $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash ((B_1 \to C_1) \vee \ldots \vee (B_n \to C_n))\,\sigma \to (B'\sigma_C \to C'\sigma_B)$ and we are done as it is very easy to see that σ and $B'\sigma_C \to C'\sigma_B$ satisfy the necessary conditions.

- The case for $B \to C$ being a negative subformula occurrence is again symmetric to the previous case.

- Suppose $\circledast B$ is a positively occurring subformula and let

$$[t_1]_{\circledast} B_1, \ldots, [t_n]_{\circledast} B_n \in |\circledast B|.$$

By induction hypothesis there is a $B' \in \|B\|$ and a substitution σ such that $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash (B_1 \vee \ldots \vee B_n)\sigma \to B'$. Note that we can apply the induction hypothesis as each B_j is a disjunction of elements of $|B|$, hence in particular $B_1 \vee \ldots \vee B_n$ is a disjunction of elements of $|B|$. We immediately obtain $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash B_j\sigma \to B'$ for each $j = 1 \ldots n$. By constructive necessitation (Corollary 6.6) there is a ground proof term $u_j \in \mathrm{Tm}_{\circledast}$ such that $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash [u_j]_{\circledast}(B_j\sigma \to B')$. Using the application axiom (Lemma 6.1, respectively, for the case where $\circledast = \mathsf{E}$) we obtain $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash ([t_j]_{\circledast} B_j)\sigma \to [u_j \cdot t_j\sigma]_{\circledast} B'$. Now we can set $s := u_1 \cdot t_1\sigma + \ldots + u_n \cdot t_n\sigma$ and using the sum axiom (again Lemma 6.1, respectively), we get $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash ([t_j]_{\circledast} B_j)\sigma \to [s]_{\circledast} B'$ and so finally also $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash (([t_1]_{\circledast} B_1)\sigma \vee \ldots \vee ([t_n]_{\circledast} B_n)\sigma) \to [s]_{\circledast} B'$. Obviously we have $[s]_{\circledast} B' \in \|\circledast B\|$ and σ satisfies the necessary conditions by induction hypothesis.

- Suppose $\circledast B$ is a negatively occuring subformula and let

$$[x]_{\circledast} B_1, \ldots, [x]_{\circledast} B_n \in |\circledast B|$$

where $\mathcal{A}(\circledast B) = x \in \mathrm{Var}_{\circledast}$. By induction hypothesis there is a $B' \in \|B\|$ and a substitution σ such that $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash B' \to (B_1 \wedge \ldots \wedge B_n)\sigma$. As σ lives on B', meets the no new variable condition, B is a subformula of A and $\mathcal{A}$ is a proper annotation of A, we have $x \notin \mathrm{dom}(\sigma)$. Furthermore, for each $j = 1 \ldots n$ we have $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash B' \to B_j\sigma$ and so, by constructive necessitation (Corollary 6.6), there is a ground proof term $t_j \in \mathrm{Tm}_{\circledast}$ such that $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash [t_j]_{\circledast}(B' \to B_j\sigma)$. Now set $s := t_1 + \ldots + t_n$, then by the sum axiom (Lemma 6.1, respectively), we have $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash [s]_{\circledast}(B' \to B_j\sigma)$. Define $\sigma_0 := \{x/(s \cdot x)\}$ and by the schematicness of $\mathcal{CS}$ we get $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash ([s]_{\circledast}(B' \to B_j\sigma))\sigma_0$. As s is ground, we immediately have $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash [s]_{\circledast}(B'\sigma_0 \to B_j\sigma\sigma_0)$. By the application axiom (Lemma 6.1, respectively) also $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash [x]_{\circledast}(B'\sigma_0) \to [s \cdot x]_{\circledast}(B_j\sigma\sigma_0)$. Set $\sigma' := \sigma \cup \sigma_0$ (remember $x \notin \mathrm{dom}(\sigma)$). Now we can restate our previous result as $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash [x]_{\circledast}(B'\sigma_0) \to ([x]_{\circledast}B_j)\sigma'$. So we also have $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash [x]_{\circledast}(B'\sigma_0) \to ([x]_{\circledast}B_1 \wedge \ldots \wedge [x]_{\circledast}B_n)\sigma'$. It is easy to see that σ' lives on $[x]_{\circledast}(B'\sigma_0)$ and meets the no new variable condition. □

To conclude, let $\mathcal{CS}$ be a homogeneous, C-axiomatically appropriate, schematic constant specification. Our aim is to show

> If $\mathsf{S4}_h^{\mathsf{C}} \vdash A$, then there is a $A' \in \|A\|$ and a substitution σ such that $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash B'$.

This would be done in the following manner: if $\mathsf{S4}_h^{\mathsf{C}} \vdash A$ and if $(*)$ holds, there are $B_1, \ldots, B_n \in |B|$ such that

$$\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash B_1 \vee \ldots \vee B_n.$$

By Lemma C.12, there is a $B' \in \|B\|$ and a substitution σ, such that

$$\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash (B_1 \vee \ldots \vee B_n)\sigma \to B'.$$

Furthermore, by Lemma 6.4, and as $\mathcal{CS}\sigma \subseteq \mathcal{CS}$ by the schematicness of $\mathcal{CS}$ we also have

$$\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash (B_1 \vee \ldots \vee B_n)\sigma.$$

So, we finally obtain

$$\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash B'.$$

Note that this method can be easily adapted to also cover the logics introduced in Chapter 7, as it does not assume any particular aspects of the logics except constructive necessitation and the truth lemma with respect to the alternative canonical model.

Bibliography

[AF11] Sergei [N.] Artemov and Melvin Fitting. "Justification Logic". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2011. 2011. URL: `http://plato.stanford.edu/archives/fall2011/entries/logic-justification/` (cit. on pp. 51, 62).

[AGW07] Pietro Abate, Rajeev Goré, and Florian Widmann. "Cut-free single-pass tableaux for the logic of common knowledge". In: *In Workshop on Agents and Deduction at TABLEAUX*. 2007 (cit. on p. 20).

[AJ05] Luca Alberucci and Gerhard Jäger. "About cut elimination for logics of common knowledge". In: *Annals of Pure and Applied Logic* 133.1–3 (May 2005), pp. 73–99. DOI: `10.1016/j.apal.2004.10.004` (cit. on pp. 17, 18, 31, 32, 86).

[AK09] Sergei [N.] Artemov and Roman Kuznets. "Logical Omniscience as a Computational Complexity Problem". In: *Theoretical Aspects of Rationality and Knowledge, Proceedings of the Twelfth Conference (TARK 2009)*. Ed. by Aviad Heifetz. Stanford University, California: ACM, July 2009, pp. 14–23. DOI: `10.1145/1562814.1562821` (cit. on p. 54).

[AN05] Sergei [N.] Artemov and Elena Nogina. "Introducing Justification into Epistemic Logic". In: *Journal of Logic and Computation* 15.6 (Dec. 2005), pp. 1059–1073. DOI: `10.1093/logcom/exi053` (cit. on p. 62).

[Ant07] Evangelia Antonakos. "Justified and Common Knowledge: Limited Conservativity". In: *Logical Foundations of Computer Science, International Symposium, LFCS 2007, New York, NY, USA, June 4–7, 2007, Proceedings*. Ed. by Sergei N. Artemov and Anil Nerode. Vol. 4514. Lecture Notes in

Computer Science. Springer, 2007, pp. 1–11. DOI: 10.1007/978-3-540-72734-7_1 (cit. on p. 63).

[Art01] Sergei N. Artemov. "Explicit Provability and Constructive Semantics". In: *Bulletin of Symbolic Logic* 7.1 (Mar. 2001), pp. 1–36 (cit. on pp. 51, 53, 71, 104, 128).

[Art06] Sergei [N.] Artemov. "Justified common knowledge". In: *Theoretical Computer Science* 357.1–3 (July 2006), pp. 4–22. DOI: 10.1016/j.tcs.2006.03.009 (cit. on pp. 53, 62–64).

[Art08] Sergei [N.] Artemov. "The Logic of Justification". In: *The Review of Symbolic Logic* 1.4 (Dec. 2008), pp. 477–513. DOI: 10.1017/S1755020308090060 (cit. on pp. 53, 55, 60, 90, 94, 170).

[Art10] Sergei [N.] Artemov. "Tracking Evidence". In: *Fields of Logic and Computation, Essays Dedicated to Yuri Gurevich on the Occasion of His 70th Birthday*. Ed. by Andreas Blass, Nachum Dershowitz, and Wolfgang Reisig. Vol. 6300. Lecture Notes in Computer Science. Springer, 2010, pp. 61–74. DOI: 10.1007/978-3-642-15025-8_3 (cit. on pp. 53, 63, 64).

[Art95] Sergei N. Artemov. *Operational modal logic*. Tech. rep. MSI 95–29. Cornell University, Dec. 1995 (cit. on pp. 53, 128).

[Bal10] Roberta Ballarin. "Modern Origins of Modal Logic". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2010. 2010 (cit. on pp. 3, 4).

[Bar88] Jon Barwise. "Three views of common knowledge". In: *Proceedings of the 2nd conference on Theoretical aspects of reasoning about knowledge*. TARK '88. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1988, pp. 365–379. ISBN: 0-934613-66-9. URL: http://dl.acm.org/citation.cfm?id=1029718.1029753 (cit. on p. 11).

[BBW07] Patrick Blackburn, J. F. A. K. van Benthem, and Frank Wolter, eds. *Handbook of Modal Logic*. Vol. 3. Studies in logic and practical reasoning. Elsevier, 2007. ISBN: 978-0-444-51690-9 (cit. on p. 3).

[Bet53] Evert Willem Beth. "On Padoa's method in the theory of definitions". In: *Indagationes mathematicae.* A 56 (1953), pp. 330–339 (cit. on p. 21).

[BGK10] Kai Brünnler, Remo Goetschi, and Roman Kuznets. "A Syntactic Realization Theorem for Justification Logics". In: *Advances in Modal Logic, Volume 8*. Ed. by Lev [D.] Beklemishev, Valentin Goranko, and Valentin Shehtman. College Publications, 2010, pp. 39–58 (cit. on pp. 53, 57).

[Bie11] Ambrose Bierce. *The Collected Works of Ambrose Bierce.* Vol. 7. Neale Publishing Company, 1911 (cit. on pp. 63, 97).

[BKS10a] Samuel Bucheli, Roman Kuznets, and Thomas Studer. *Explicit Evidence Systems with Common Knowledge.* E-print 1005.0484. Later version published as [BKS11a]. arXiv.org, May 2010. eprint: `1005.0484` (cit. on p. 65).

[BKS10b] Samuel Bucheli, Roman Kuznets, and Thomas Studer. "Two Ways to Common Knowledge". In: *Proceedings of the 6th Workshop on Methods for Modalities (M4M–6 2009), Copenhagen, Denmark, 12–14 November 2009*. Ed. by Thomas Bolander and Torben Braüner. Electronic Notes in Theoretical Computer Science 262. Elsevier, May 2010, pp. 83–98. DOI: `10.1016/j.entcs.2010.04.007` (cit. on pp. 11, 32, 87).

[BKS11a] Samuel Bucheli, Roman Kuznets, and Thomas Studer. "Justifications for common knowledge". In: *Journal of Applied Non-Classical Logics* 21.1 (Feb. 2011), pp. 35–60. DOI: `10.3166/JANCL.21.35-60` (cit. on pp. 65, 193).

[BKS11b] Samuel Bucheli, Roman Kuznets, and Thomas Studer. "Partial Realization in Dynamic Justification Logic". In: *Logic, Language, Information and Computation, 18th International Workshop, WoLLIC 2011, Philadelphia, PA, USA, May 18–20, 2011, Proceedings*. Ed. by Lev D. Beklemishev and Ruy de Queiroz. Vol. 6642. Lecture Notes in Artificial Intelligence. Springer, 2011, pp. 35–51. DOI: `10.1007/978-3-642-20920-8_9` (cit. on pp. 151, 152, 154, 159, 164).

[BKS12a] Samuel Bucheli, Roman Kuznets, and Thomas Studer. "Decidability for Justification Logics Revisited". Submitted. 2012 (cit. on p. 131).

[BKS12b] Samuel Bucheli, Roman Kuznets, and Thomas Studer. "Realizing Public Announcements by Justifications". Submitted. 2012 (cit. on pp. 151, 154, 171).

[BL08] Kai Brünnler and Martin Lange. "Cut-free sequent systems for temporal logic". In: *Journal of Logic and Algebraic Programming* 76.2 (2008), pp. 216 –225. ISSN: 1567-8326. DOI: `DOI:10.1016/j.jlap.2008.02.004` (cit. on p. 20).

[Bla+11] Andreas Blass et al. "Evidential Authorization". In: *The Future of Software Engineering*. Ed. by Sebastian Nanz. Springer, 2011, pp. 73–99. DOI: `10.1007/978-3-642-15187-3_5` (cit. on p. 92).

[Bre00] Vladimir N. Brezhnev. *On explicit counterparts of modal logics*. Tech. rep. CFIS 2000–05. Cornell University, 2000 (cit. on p. 171).

[BRV02] P. Blackburn, M. Rijke, and Y. Venema. *Modal logic*. Cambridge tracts in theoretical computer science. Cambridge University Press, 2002. ISBN: 9780521527149 (cit. on pp. 3, 129, 131).

[Brü10] Kai Brünnler. *Nested Sequents*. E-print 1004.1845. arXiv, 2010. URL: `http://arxiv.org/abs/1004.1845` (cit. on p. 20).

[BS07] Julian Bradfield and Colin Stirling. "Modal Mu-Calculi". In: *Handbook of Modal Logic*. Ed. by Patrick Blackburn, Johan van Benthem, and Frank Wolter. Vol. 3. Studies in Logic and Practical Reasoning. Elsevier, 2007, pp. 721–756. DOI: `10.1016/S1570-2464(07)80015-2` (cit. on p. 31).

[BS09] Kai Brünnler and Thomas Studer. "Syntactic cut-elimination for common knowledge". In: *Annals of Pure and Applied Logic* 160.1 (July 2009), pp. 82–95. DOI: `10.1016/j.apal.2009.01.014` (cit. on pp. 20, 86).

[Buc+10] Samuel Bucheli et al. "Justified Belief Change". In: *LogKCA-10, Proceedings of the Second ILCLI International Workshop on Logic and Philosophy of Knowledge, Communication and Action*. Ed. by Xabier Arrazola and María Ponte. University of the Basque Country Press, 2010, pp. 135–155 (cit. on pp. 151, 152, 154, 159, 164).

[Car98] Lewis Carroll. *Alice's adventures in Wonderland ; and, Through the looking-glass, and what Alice found there*. London New York: Penguin Books, 1998. ISBN: 978-0-141-43976-1 (cit. on p. 1).

[Cra57] William Craig. "Three Uses of the Herbrand-Gentzen Theorem in Relating Model Theory and Proof Theory". In: *The Journal of Symbolic Logic* 22.3 (1957), pp. 269–285. ISSN: 00224812. URL: `http://www.jstor.org/stable/2963594` (cit. on p. 21).

[DHK07] Hans van Ditmarsch, Wiebe van der Hoek, and Barteld Kooi. *Dynamic Epistemic Logic*. Vol. 337. Synthese Library. Springer, 2007. DOI: `10.1007/978-1-4020-5839-4` (cit. on pp. 28, 30, 151, 155, 156, 158).

[DHL06] Christian Dax, Martin Hofmann, and Martin Lange. "A Proof System for the Linear Time μ-Calculus". In: *FSTTCS 2006: Foundations of Software Technology and Theoretical Computer Science, 26th International Conference, Kolkata, India, December 13-15, 2006, Proceedings*. Ed. by S. Arun-Kumar and Naveen Garg. Vol. 4337. Lecture Notes in Computer Science. Springer, 2006, pp. 273–284. DOI: `10.1007/11944836_26` (cit. on pp. 31, 46, 127).

[Fag+95] Ronald Fagin et al. *Reasoning about Knowledge*. MIT Press, 1995 (cit. on pp. 11, 13, 14, 18, 22, 23, 28, 63, 87, 89, 91, 95).

[Fef+86] S. Feferman et al., eds. *Kurt Gödel Collected Works*. Vol. I. Oxford University Press, 1986, pp. 301–303 (cit. on p. 197).

[Fit05] Melvin Fitting. "The logic of proofs, semantically". In: *Annals of Pure and Applied Logic* 132.1 (Feb. 2005), pp. 1–25. DOI: `10.1016/j.apal.2004.04.009` (cit. on pp. 53, 57, 60, 64, 75, 77, 87, 112, 115, 163, 166).

[Fit08a] Melvin Fitting. "A quantified logic of evidence". In: *Annals of Pure and Applied Logic* 152.1–3 (Mar. 2008), pp. 67–83. DOI: 10.1016/j.apal.2007.11.003 (cit. on p. 62).

[Fit08b] Melvin Fitting. "Justification logics, logics of knowledge, and conservativity". In: *Annals of Mathematics and Artificial Intelligence* 53.1–4 (Aug. 2008), pp. 153–167. DOI: 10.1007/s10472-009-9112-2 (cit. on pp. 82, 85, 121).

[Fit09] Melvin Fitting. "Realizations and LP". In: *Annals of Pure and Applied Logic* 161.3 (Dec. 2009), pp. 368–387. DOI: 10.1016/j.apal.2009.07.010 (cit. on pp. 129, 155, 166, 169, 171).

[Fit10] Melvin Fitting. "Justification Logics and Hybrid Logics". In: *Journal of Applied Logic* 8.4 (Dec. 2010). Published online August 2010, pp. 356–370. DOI: 10.1016/j.jal.2010.08.007 (cit. on pp. 57, 87).

[Fit11] Melvin Fitting. "The Realization Theorem for S5: A Simple, Constructive Proof". In: *Games, Norms and Reasons: Logic at the Crossroads*. Ed. by Johan van Benthem, Amitabha Gupta, and Eric Pacuit. Vol. 353. Synthese Library. Springer, 2011. Chap. 4, pp. 61–76. DOI: 10.1007/978-94-007-0714-6_4 (cit. on pp. 57, 155).

[Fit12] Melvin Fitting. "Prefixed tableaus and nested sequents". In: *Annals of Pure and Applied Logic* 163.3 (2012), pp. 291 –313. ISSN: 0168-0072. DOI: 10.1016/j.apal.2011.09.004. URL: http://www.sciencedirect.com/science/article/pii/S0168007211001266 (cit. on p. 20).

[FM98] Melvin Fitting and Richard L. Mendelsohn. *First-Order Modal Logic*. Vol. 277. Synthese Library. Kluwer Academic Publishers, 1998 (cit. on p. 3).

[Gar09] James Garson. "Modal Logic". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2009. 2009 (cit. on p. 3).

[Ger99] Jelle Gerbrandy. "Bisimulations on Planet Kripke". PhD thesis. Institute for Logic, Language, and Computation, University of Amsterdam, 1999. URL: `http://www.illc.uva.nl/Publications/Dissertations/DS-1999-01.text.pdf` (cit. on pp. 155, 156).

[Get63] Edmund Gettier. "Is Justified True Belief Knowledge?" In: *Analysis* 23.6 (1963), pp. 121–123 (cit. on p. 54).

[GG97] Jelle Gerbrandy and Willem Groeneveld. "Reasoning about Information Change". In: *Journal of Logic, Language and Information* 6.2 (Apr. 1997), pp. 147–169. DOI: `10.1023/A:1008222603071` (cit. on pp. 151, 154–156).

[GK12] Remo Goetschi and Roman Kuznets. "Realization for Justification Logics via Nested Sequents: Modularity through Embedding". In: *Annals of Pure and Applied Logic* (2012). ISSN: 0168-0072. DOI: `10.1016/j.apal.2012.02.002`. URL: `http://www.sciencedirect.com/science/article/pii/S0168007212000565` (cit. on p. 57).

[Gol03] Robert Goldblatt. "Mathematical modal logic: A view of its evolution". In: *Journal of Applied Logic* 1.5–6 (2003), pp. 309–392. ISSN: 1570-8683. DOI: `10.1016/S1570-8683(03)00008-9`. URL: `http://www.sciencedirect.com/science/article/pii/S1570868303000089` (cit. on p. 3).

[Göd33] Kurt Gödel. "Eine Interpretation des intuitionistischen Aussagenkalküls". In: *Ergebnisse Math. Kolloq.* 4 (1933), pp. 39–40. English translation in S. Feferman et al., eds. *Kurt Gödel Collected Works*. Vol. I. Oxford University Press, 1986, pp. 301–303 (cit. on p. 53).

[Hin62] Jaakko Hintikka. *Knowledge and Belief*. Ithaca, N.Y., Cornell University Press, 1962 (cit. on p. 3).

[HKJ00] David Harel, Dexter Kozen, and Tiuryn Jerzy. *Dynamic Logic*. Foundations of Computing. MIT Press, 2000. ISBN: 978-0-262-08289-1 (cit. on pp. 79, 184).

[HS09] Vincent Hendricks and John Symons. "Epistemic Logic". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Spring 2009. 2009 (cit. on p. 3).

[Hum39] David Hume. *A Treatise of Human Nature*. 1739 (cit. on p. 11).

[JKS07] Gerhard Jäger, Mathis Kretz, and Thomas Studer. "Cut-free common knowledge". In: *Journal of Applied Logic* 5.4 (Dec. 2007), pp. 681–689. DOI: 10.1016/j.jal.2006.02.003 (cit. on pp. 18, 19, 86, 87).

[JKS08] Gerhard Jäger, Mathis Kretz, and Thomas Studer. "Canonical completeness of infinitary μ". In: *Journal of Logic and Algebraic Programming* 76.2 (2008), pp. 270 –292. ISSN: 1567-8326. DOI: 10.1016/j.jlap.2008.02.005. URL: http://www.sciencedirect.com/science/article/pii/S1567832608000209 (cit. on p. 31).

[Joh10] Samuel Johnson. *The Collected Works of Samuel Johnson, LL.D. A New Edition in Twelve Volumes. With an Essay on his Life and Genius, by Arthur Murphy, Esq.* Vol. 4. London: Luke Hansard and Sons, 1810 (cit. on p. 127).

[Kri59] Saul A. Kripke. "A Completeness Theorem in Modal Logic". In: *The Journal of Symbolic Logic* 24.1 (1959), pp. 1–14. ISSN: 00224812. URL: http://www.jstor.org/stable/2964568 (cit. on p. 5).

[Kuz00] Roman Kuznets. "On the Complexity of Explicit Modal Logics". In: *Computer Science Logic, 14th International Workshop, CSL 2000, Annual Conference of the EACSL, Fischbachau, Germany, August 21–26, 2000, Proceedings.* Ed. by Peter G. Clote and Helmut Schwichtenberg. Vol. 1862. Lecture Notes in Computer Science. Springer, 2000, pp. 371–383. DOI: 10.1007/3-540-44622-2_25 (cit. on p. 131).

[Kuz08] Roman Kuznets. "Complexity Issues in Justification Logic". PhD thesis. CUNY Graduate Center, May 2008 (cit. on pp. 131, 139–141).

[Kuz10] Roman Kuznets. "Self-Referential Justifications in Epistemic Logic". In: *Theory of Computing Systems* 46.4 (May 2010), pp. 636–661. DOI: 10.1007/s00224-009-9209-3 (cit. on pp. 54, 171).

[Kuz12] Roman Kuznets. *Justification Logic Bibliography*. 2012. URL: http://sites.google.com/site/kuznets/justificationlogicbibliography (cit. on p. 62).

[Lew69] David Lewis. *Convention: A Philosophical Study*. Harvard University Press, 1969 (cit. on p. 11).

[Lit86] John Littlewood. *Littlewood's Miscellany*. Ed. by Béla Bollobás. Cambridge University Press, 1986. ISBN: 978-0-521-33702-1. DOI: 10.2277/052133702X (cit. on p. 3).

[LS77] Edward John Lemmon and Dana Stewart Scott. *An Introduction to Modal Logic*. Oxford: B. Blackwell, 1977 (cit. on p. 131).

[Mar75] Donald A. Martin. "Borel Determinacy". In: *The Annals of Mathematics*. Second Series 102.2 (1975), pp. 363–371. ISSN: 0003486X. URL: http://www.jstor.org/stable/1971035 (cit. on p. 46).

[McC+78] John McCarthy et al. *On the model theory of knowledge*. Tech. rep. CS–TR–78–657. Stanford University Computer Science Department, Apr. 1978. URL: ftp://reports.stanford.edu/pub/cstr/reports/cs/tr/78/657/CS-TR-78-657.pdf (cit. on p. 63).

[MH95] J.-J. Ch. Meyer and W. van der Hoek. *Epistemic Logic for AI and Computer Science*. Vol. 41. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 1995 (cit. on pp. 11, 14, 22, 23, 28, 63, 65, 78, 80, 117, 184).

[Mil09] Robert S. Milnikel. "Conservativity for Logics of Justified Belief". In: *Logical Foundations of Computer Science, International Symposium, LFCS 2009, Deerfield Beach, FL, USA, January 3–6, 2009, Proceedings*. Ed. by Sergei [N.] Artemov and Anil Nerode. Vol. 5407. Lecture Notes in Computer Science. Later version published as [Mil12]. Springer, 2009, pp. 354–364. DOI: 10.1007/978-3-540-92687-0_24 (cit. on p. 124).

[Mil12] Robert S. Milnikel. "Conservativity for logics of justified belief: Two approaches". In: *Annals of Pure and Applied Logic* 163.7 (July 2012). Published online October 2011. DOI: 10.1016/j.apal.2011.09.015 (cit. on p. 199).

[Mkr97] Alexey Mkrtychev. "Models for the Logic of Proofs". In: *Logical Foundations of Computer Science, 4th International Symposium, LFCS'97, Yaroslavl, Russia, July 6–12, 1997, Proceedings*. Ed. by Sergei Adian and Anil Nerode. Vol. 1234. Lecture Notes in Computer Science. Springer, 1997, pp. 266–275. DOI: 10.1007/3-540-63045-7_27 (cit. on pp. 53, 64, 80, 131).

[Non91] Ikujiro Nonaka. "The Knowledge-Creating Company". In: *Harvard Business Review* (1991) (cit. on p. 91).

[NW96] Damian Niwiński and Igor Walukiewicz. "Games for the μ-Calculus". In: *Theoretical Computer Science* 163.1–2 (1996), pp. 99–116. DOI: 10.1016/0304-3975(95)00136-0 (cit. on pp. 31, 41, 46, 127).

[Pac05] Eric Pacuit. "A Note on Some Explicit Modal Logics". In: *Proceedings of the 5th Panhellenic Logic Symposium*. Athens, Greece: University of Athens, July 2005, pp. 117–125 (cit. on pp. 60, 62, 166).

[Pla03] Plato. *Platonis Opera*. Ed. by John Burnet. Oxford University Press, 1903 (cit. on pp. 31, 51).

[Pla07a] Jan Plaza. "Logics of public communications". In: *Synthese* 158.2 (Sept. 2007). Reprinted from M. L. Emrich et al., editors, *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems (ISMIS '89)*, pages 201–216. Oak Ridge National Laboratory, ORNL/DSRD-24, 1989., pp. 165–179. DOI: 10.1007/s11229-007-9168-7 (cit. on p. 91).

[Pla07b] Jan Plaza. "Logics of public communications". In: *Synthese* 158.2 (Sept. 2007). Reprinted from M. L. Emrich et al., editors, *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems (ISMIS '89)*, pages 201–216. Oak Ridge National Laboratory, ORNL/DSRD-24, 1989., pp. 165–179. DOI: 10.1007/s11229-007-9168-7 (cit. on pp. 151, 156).

[Pla21] Plato. *Plato in Twelve Volumes*. Trans. by Harold N. Fowlder. Vol. 12. Translated by Harold N. Fowler. Harvard University Press, 1921 (cit. on pp. 31, 51).

[Pos44] Emil L. Post. "Recursively enumerable sets of positive integers and their decision problems." English. In: *Bull. Am. Math. Soc.* 50 (1944), pp. 284–316. DOI: 10.1090/S0002-9904-1944-08111-1 (cit. on p. 140).

[Ren] Bryan Renne. "Multi-agent Justification Logic: communication and evidence elimination". In: *Synthese* (), pp. 1–40. ISSN: 0039-7857. DOI: 10.1007/s11229-011-9968-7. URL: http://dx.doi.org/10.1007/s11229-011-9968-7 (cit. on p. 154).

[Ren08] Bryan Renne. "Dynamic Epistemic Logic with Justification". PhD thesis. CUNY Graduate Center, May 2008 (cit. on p. 154).

[Ren09a] Bryan Renne. "Evidence Elimination in Multi-Agent Justification Logic". In: *Theoretical Aspects of Rationality and Knowledge, Proceedings of the Twelfth Conference (TARK 2009)*. Ed. by Aviad Heifetz. Stanford University, California: ACM, July 2009, pp. 227–236. DOI: 10.1145/1562814.1562845 (cit. on pp. 63, 64).

[Ren09b] Bryan Renne. "Propositional games with explicit strategies". In: *Information and Computation* 207.10 (Oct. 2009), pp. 1015–1043. DOI: 10.1016/j.ic.2008.11.005 (cit. on p. 128).

[Ren11] Bryan Renne. "Public communication in justification logic". In: *Journal of Logic and Computation* 21.6 (Dec. 2011). Published online July 2010, pp. 1005–1034. DOI: 10.1093/logcom/exq026 (cit. on p. 154).

[Rub06] N[atalia] M. Rubtsova. "Evidence-based knowledge for S5". In: *2005 Summer Meeting of the Association for Symbolic Logic, Logic Colloquium '05, Athens, Greece, July 28–August 3, 2005*. Vol. 12(2). Bulletin of Symbolic Logic. Abstract. Association for Symbolic Logic, June 2006, pp. 344–345. DOI: 10.2178/bsl/1146620064 (cit. on p. 62).

[SE89] Robert S. Streett and E. Allen Emerson. "An Automata Theoretic Decision Procedure for the Propositional Mu-Calculus". In: *Information and Computation* 81.3 (1989), pp. 249–264. DOI: `10.1016/0890-5401(89)90031-X` (cit. on p. 31).

[Seg71] Krister Segerberg. *An Essay in Classical Modal Logic*. Uppsala,Filosofiska Föreningen Och Filosofiska Institutionen Vid Uppsala Universitet, 1971 (cit. on p. 131).

[SLB09] Yoav Shoham and Kevin Leyton-Brown. *Multiagent Systems. Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge: Cambridge University Press, 2009. ISBN: 9780521899437 (cit. on p. 11).

[ST96] H. Schwichtenberg and A.S. Troelstra. *Basic Proof Theory*. Vol. 43. Cambrdige Tracts in Theoretical Computer Science. 1996 (cit. on p. 21).

[Stu08] Thomas Studer. "On the Proof Theory of the Modal mu-Calculus". In: *Studia Logica* 89.3 (2008), pp. 343–363. DOI: `10.1007/s11225-008-9133-6` (cit. on pp. 31, 32, 127).

[Stu09] Thomas Studer. "Common Knowledge does not have the Beth Property". In: *Information Processing Letters* 109 (2009), 611–614 (cit. on p. 21).

[Stu11] Thomas Studer. "Decidability for some justification logics with negative introspection". Preprint. 2011. URL: `http://www.iam.unibe.ch/~tstuder/papers/J5_Decidability.pdf` (cit. on p. 149).

[SW91] Colin Stirling and David Walker. "Local model checking in the modal mu-calculus". In: *Theoretical Computer Science* 89.1 (1991), pp. 161–177. DOI: `10.1016/0304-3975(90)90110-4` (cit. on p. 31).

[VS09] Peter Vanderschraaf and Giacomo Sillari. "Common Knowledge". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Spring 2009. 2009 (cit. on p. 11).

[Weh10] Ricardo Wehbe. "Annotated Systems for Common Knowledge". phd. Institut für Informatik und angewandte Mathematik, 2010. URL: `http://www.iam.unibe.ch/tilpub/2010/weh10.pdf` (cit. on pp. 20, 127).

[Yav01] Rostislav E. Yavorsky. “Provability logics with quantifiers on proofs”. In: *Annals of Pure and Applied Logic* 113.1–3 (Dec. 2001), pp. 373–387. DOI: `10.1016/S0168-0072(01)00067-7` (cit. on p. 62).

[Yav08] Tatiana Yavorskaya (Sidon). “Interacting Explicit Evidence Systems”. In: *Theory of Computing Systems* 43.2 (Aug. 2008), pp. 272–293. DOI: `10.1007/s00224-007-9057-y` (cit. on pp. 62–65, 82, 92, 94).

List of Notations

Index

Erklärung

gemäss Art. 28 Abs. 1 RSL 05

Name/Vorname:	Bucheli Samuel
Matrikelnummer:	02-114-460
Studiengang:	Informatik, Dissertation
Titel der Arbeit:	Justification Logics with Common Knowledge
Leiter der Arbeit:	Prof. Dr. G. Jäger

Ich erkläre hiermit, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls der Senat gemäss Artikel 36 Absatz 1 Buchstabe o des Gesetzes vom 5. September 1996 über die Universität zum Entzug des auf Grund dieser Arbeit verliehenen Titels berechtigt ist.

Bern, den 1. Mai 2012	....
Ort/Datum	Unterschrift

Lebenslauf

von Samuel Bucheli

1982 Geboren am 2. September in Bern

1989–1995 Primarschule Schüpfheim

1995–1999 Kantonsschule Schüpfheim

1999–2002 Kantonsschule Beromünster
Abschluss mit Matura Typus C

2002–2008 Diplomstudium in Mathematik mit Nebenfächern Informatik und Physik an der Universität Bern

Titel der Diplomarbeit:
Explicit Mathematics with Positive Existential Stratified Comprehension, Join and Uniform Monotone Inductive Definitions
Leiter der Diplomarbeit: Prof. Dr. Thomas Strahm

2007–2009 Studium an der Pädagogischen Hochschule Bern
Abschluss mit Lehrdiplom für Maturitätsschulen in den Fächern Mathematik und Informatik

2008–2012 PhD-Studium in Informatik bei Prof. Dr. Gerhard Jäger in der Logic and Theory Group (ehemals Forschungsgruppe für Theoretische Informatik und Logik) am Institut für Informatik und angewandte Mathematik der Universität Bern