-

Justifications for Common Knowledge

Samuel Bucheli — Roman Kuznets — Thomas Studer

Institut für Informatik und angewandte Mathematik, Universität Bern
Neubrückstrasse 10, CH-3012 Bern (Switzerland)
{*bucheli, kuznets, tstuder*}*@iam.unibe.ch*

ABSTRACT. *Justification logics are epistemic logics that explicitly include justifications for the agents' knowledge. We develop a multi-agent justification logic with evidence terms for individual agents as well as for common knowledge. We define a Kripke-style semantics that is similar to Fitting's semantics for the Logic of Proofs* LP. *We show the soundness, completeness, and finite model property of our multi-agent justification logic with respect to this Kripke-style semantics. We demonstrate that our logic is a conservative extension of Yavorskaya's minimal bimodal explicit evidence logic, which is a two-agent version of* LP. *We discuss the relationship of our logic to the multi-agent modal logic* S4 *with common knowledge. Finally, we give a brief analysis of the coordinated attack problem in the newly developed language of our logic.*

KEYWORDS: *justification logic, epistemic modal logic, multi-agent systems, common knowledge.*

DOI:10.3166/JANCL.-.1–25

1. Introduction

Justification logics are epistemic logics that explicitly include justifications for the agents' knowledge (Artemov, 2008). The first logic of this kind, the *Logic of Proofs* LP, was developed by Artemov to provide the modal logic S4 with provability semantics (Artemov, 1995; Artemov, 2001). The language of justification logics has also been used to create a new approach to the logical omniscience problem (Artemov *et al.*, 2009) and to study self-referential proofs (Kuznets, 2010).

Instead of statements 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 S4, as is illustrated in Fig. 1.

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 1. *Axioms of* S4 *and* LP

Artemov has shown that the Logic of Proofs LP is an *explicit*[1] *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 (Artemov, 2001). The latter process is called *realization*, and the statement of correspondence is called a *realization theorem*. Note that the operation $+$ introduced by the sum axiom in Fig. 1 does not have a modal analog, but it is an essential part of the proof of the realization theorem in (Artemov, 2001). Explicit counterparts for many normal modal logics between K and S5 have been developed (see a recent survey in (Artemov, 2008) and a uniform proof of realization theorems for all single-agent justification logics in (Brünnler *et al.*, 2010)).

The notion of *common knowledge* is essential in the area of multi-agent systems, where coordination among agents is a central issue. For a thorough introduction to epistemic logics in general and to common knowledge in particular, one can refer to the standard textbooks (Fagin *et al.*, 1995; Meyer *et al.*, 1995). Informally, 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

$$\lambda X.(\text{everybody knows } A \text{ and everybody knows } X) \ . \tag{1}$$

An explicit counterpart of McCarthy's *any fool knows* common knowledge modality (McCarthy *et al.*, 1978), where common knowledge of A is defined as an arbitrary fixed point of (1), is presented in (Artemov, 2006). The relationship between the traditional common knowledge from (Fagin *et al.*, 1995; Meyer *et al.*, 1995) and McCarthy's version is studied in (Antonakos, 2007).

In this paper, 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 $\mathsf{S4}_h^{\mathsf{C}}$. For the sake of compactness and readability, we will not treat groups of agents.

1. For other meanings of "explicit" see Sect. 8.

Multi-agent justification logics with evidence terms for each agent are considered in (Yavorskaya (Sidon), 2008; Renne, 2009a; Artemov, 2010), but common knowledge is not present in any of them. Renne's system combines features of modal and dynamic epistemic logics (Renne, 2009a) 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 (Artemov, 2010). Yavorskaya studies various operations of evidence transfer between agents (Yavorskaya (Sidon), 2008). Yavorskaya's minimal[2] 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.

An epistemic semantics for LP, *F-models*, was created by Fitting by augmenting Kripke models with an *evidence function* that specifies which formulae are evidenced by a term at a given world (Fitting, 2005). Independently, Mkrtychev proved a stronger completeness result for LP with respect to singleton F-models (Mkrtychev, 1997), now known as *M-models*, where the role of the accessibility relation is completely taken over by the evidence function. The semantics of F-models has been adapted to the whole family of single-agent justification logics (for details, see (Artemov, 2008)). Artemov extends F-models to the language with both evidence terms for McCarthy's common knowledge modality and ordinary modalities for the individual agents (Artemov, 2006), 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 (Yavorskaya (Sidon), 2008).

The paper is organized as follows. In Sect. 2, we introduce a language and give an axiomatization of a family of multi-agent justification logics with common knowledge. In Sect. 3, we prove their basic properties including the internalization property, which is characteristic of all justification logics. In Sect. 4, 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 Sect. 5, we show that for the two-agent case, our logic is a conservative extension of Yavorskaya's minimal two-agent justification logic. In Sect. 6, we demonstrate how our logic is related to the modal logic of traditional common knowledge and discuss the problem of realization. In Sect. 7, we provide an analysis of the coordinated attack problem in our logic. Finally, in Sect. 8, we discuss how the newly introduced terms affect the agents, including their ability to communicate information in various communication modes.

2. Minimality here is understood in the sense of the minimal transfer of evidence.

2. Syntax

To create an explicit counterpart of the modal logic of common knowledge $\mathsf{S4}^{\mathsf{C}}_h$, we use its axiomatization *via* the induction axiom from (Meyer *et al.*, 1995) rather than *via* the induction rule 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. (Yavorskaya (Sidon), 2008)).

As motivated in (Bucheli *et al.*, 2010b), 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

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

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

$$[t]_{\mathsf{C}}A \to [\mathsf{ccl}_1(t)]_{\mathsf{E}}A \ , \quad [t]_{\mathsf{C}}A \to [\mathsf{ccl}_2(t)]_{\mathsf{E}}\,[t]_{\mathsf{C}}A \ ,$$

by means of two unary co-closure operations $\mathsf{ccl}_1(\cdot)$ and $\mathsf{ccl}_2(\cdot)$.

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 Fig. 1) are adopted for common knowledge evidence and none is adopted for mutual knowledge evidence, it will be shown in Sect. 3 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, 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 Sect. 6 for details).

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

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 $\circledast$. 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. $\mathsf{ccl}_1(t) \in \mathrm{Tm}_\mathsf{E}$ and $\mathsf{ccl}_2(t) \in \mathrm{Tm}_\mathsf{E}$ 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 $s \in \mathrm{Tm}_\mathsf{E}$.

$\mathrm{Tm} := \mathrm{Tm}_1 \cup \dots \cup \mathrm{Tm}_h \cup \mathrm{Tm}_\mathsf{E} \cup \mathrm{Tm}_\mathsf{C}$ 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 $\mathrm{Prop} := \{P_1, P_2, \dots\}$ be a countable set of *propositional variables*. *Formulae* are denoted by $A, B, C, \dots$ and are defined by the grammar

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

where $t \in \mathrm{Tm}_\circledast$ and $P_j \in \mathrm{Prop}$. The set of all formulae is denoted by $\mathrm{Fm}_{\mathsf{LP}_h^\mathsf{C}}$. 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.

Axioms of LP_h^C:

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. $[t_1]_1 A \wedge \dots \wedge [t_h]_h A \to [\langle t_1, \dots, t_h \rangle]_\mathsf{E} A$ (tupling)
7. $[t]_\mathsf{E} A \to [\pi_i t]_i A$ (projection)
8. $[t]_\mathsf{C} A \to [\mathsf{ccl}_1(t)]_\mathsf{E} A, \quad [t]_\mathsf{C} A \to [\mathsf{ccl}_2(t)]_\mathsf{E}\, [t]_\mathsf{C} A$ (co-closure)
9. $A \wedge [t]_\mathsf{C}(A \to [s]_\mathsf{E} A) \to [\mathsf{ind}(t, s)]_\mathsf{C} A$ (induction)

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

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

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 LP_h^C above and by the rules modus ponens and axiom necessitation:

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

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

$$\mathcal{CS} = \left\{ [c]_\mathsf{C} A \ : \ c \in \mathrm{Cons}_\mathsf{C} \text{ and } A \text{ is an axiom of } \mathsf{LP}_h^\mathsf{C} \right\} \ . \tag{2}$$

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 (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 LP_h^C respectively. We use Δ, A to mean $\Delta \cup \{A\}$.

3. 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 28 in Sect. 6 and the fact that $\mathsf{E}A \to \mathsf{EE}A$ is not a valid modal formula.

LEMMA 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* $t \in \mathrm{Tm}_\mathsf{E}$*;* ($\mathsf{E}$-*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* $\vdash_{\mathcal{CS}} [t]_\mathsf{E}(A \to B) \to ([s]_\mathsf{E} A \to [t \cdot_\mathsf{E} s]_\mathsf{E} B)$*;* ($\mathsf{E}$-*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* $\vdash_{\mathcal{CS}} [t]_\mathsf{E} A \vee [s]_\mathsf{E} A \to [t +_\mathsf{E} s]_\mathsf{E} A$*;* ($\mathsf{E}$-*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 $\downarrow_i t := \pi_i \mathsf{ccl}_1(t)$.
5. Immediate by 4. and the reflexivity axiom. ■

Unlike Lemma 1, Lemma 2 requires that a constant specification $\mathcal{CS}$ be C-axiomatically appropriate.

LEMMA 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* $\vdash_{\mathcal{CS}} [t]_{\mathsf{C}} A \to [\Leftarrow t]_{\mathsf{C}} [\mathsf{ccl}_1(t)]_{\mathsf{E}} A$. (C-*shift*)

PROOF. —

1. Set $!_{\mathsf{C}} t := \mathsf{ind}(c, \mathsf{ccl}_2(t))$, where $[c]_{\mathsf{C}}([t]_{\mathsf{C}} A \to [\mathsf{ccl}_2(t)]_{\mathsf{E}} [t]_{\mathsf{C}} A) \in \mathcal{CS}$.
2. Set $\Leftarrow t := c' \cdot_{\mathsf{C}} (!_{\mathsf{C}} t)$, where $[c']_{\mathsf{C}}([t]_{\mathsf{C}} A \to [\mathsf{ccl}_1(t)]_{\mathsf{E}} A) \in \mathcal{CS}$.

The existence of constants c and c' is guaranteed by the C-appropriateness of $\mathcal{CS}$. ■

The following two lemmas are standard in justification logics. Their proofs can be taken almost word for word from (Artemov, 2001) and are, therefore, omitted here.

LEMMA 3 (DEDUCTION THEOREM). — *Let* $\mathcal{CS}$ *be a constant specification and* $\Delta \cup \{A, B\} \subseteq \mathrm{Fm}_{\mathsf{LP}^{\mathsf{C}}_h}$*. Then* $\Delta, A \vdash_{\mathcal{CS}} B$ *if and only if* $\Delta \vdash_{\mathcal{CS}} A \to B$.

LEMMA 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* $t \in \mathrm{Tm}_{\circledast}$*,*

$$\text{if } \Delta \vdash_{\mathcal{CS}} A, \quad \text{then } \Delta(x/t, P/B) \vdash_{\mathcal{CS}(x/t, P/B)} A(x/t, P/B) \ ,$$

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 *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 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*

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

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

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

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

$$t_{\mathsf{C}} := c, \qquad t_i :=\downarrow_i c, \qquad t_{\mathsf{E}} := \mathsf{ccl}_1(c)$$

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

$$t_{\mathsf{C}} :=!_{\mathsf{C}} x_j, \qquad t_i :=\downarrow_i !_{\mathsf{C}} x_j, \qquad t_{\mathsf{E}} := \mathsf{ccl}_2(x_j)$$

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 1 for $\circledast = \mathsf{E}$.

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

$$t_{\mathsf{C}} :=!_{\mathsf{C}} c, \qquad t_i :=\downarrow_i !_{\mathsf{C}} c, \qquad t_{\mathsf{E}} := \mathsf{ccl}_2(c)$$

and 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 (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 used in various axiomatizations of $\mathsf{S4}_h^{\mathsf{C}}$ (see (Bucheli *et al.*, 2010b) for a discussion of several axiomatizations of this kind).

LEMMA 7 (INTERNALIZED INDUCTION RULE 1). — *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 $t \in \mathrm{Tm}_{\mathsf{C}}$. It remains to use the induction axiom and propositional reasoning. ■

LEMMA 8 (INTERNALIZED INDUCTION RULE 2). — *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

$$\vdash_{\mathcal{CS}} B \to [s]_{\mathsf{E}}(A \wedge B) \ . \tag{3}$$

From this we immediately get $\vdash_{\mathcal{CS}} A \wedge B \to [s]_{\mathsf{E}}(A \wedge B)$. Thus, by Lemma 7, there is a $t \in \mathrm{Tm}_{\mathsf{C}}$ with

$$\vdash_{\mathcal{CS}} A \wedge B \to [\mathsf{ind}(t,s)]_{\mathsf{C}}(A \wedge B) \ . \tag{4}$$

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

$$\vdash_{\mathcal{CS}} [c]_{\mathsf{C}}(A \wedge B \to A) \ . \tag{5}$$

Making use of C-application, we find by (4) and (5) that

$$\vdash_{\mathcal{CS}} A \wedge B \to [c \cdot \mathsf{ind}(t,s)]_{\mathsf{C}}A \ . \tag{6}$$

From (3) we get by E-reflexivity that $\vdash_{\mathcal{CS}} B \to A \wedge B$. This, together with (6), finally yields $\vdash_{\mathcal{CS}} B \to [c \cdot \mathsf{ind}(t,s)]_{\mathsf{C}}A$. ■

4. Soundness and completeness

DEFINITION 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 $R\colon \{1, \ldots, h\} \to \mathcal{P}(W \times W)$ that assigns a reflexive and transitive* accessibility relation *on W to each agent $i \in \{1, \ldots, h\}$, 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: $R_{\mathsf{E}} := R_1 \cup \cdots \cup R_h$ 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* $1 \leq i \leq h$, *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* $A \in \mathcal{E}_{\mathsf{E}}(w, \mathsf{ccl}_1(t))$ *and* $[t]_{\mathsf{C}} A \in \mathcal{E}_{\mathsf{E}}(w, \mathsf{ccl}_2(t))$; (*co-closure*)
9. *if* $A \in \mathcal{E}_{\mathsf{E}}(w,s)$ *and* $(A \to [s]_{\mathsf{E}} A) \in \mathcal{E}_{\mathsf{C}}(w,t)$, *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 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 11 (SOUNDNESS). — *All theorems are valid:* $\vdash_{\mathcal{CS}} A$ *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 (Fitting, 2005) 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 $A \in \mathcal{E}_{\mathsf{E}}(w, \mathsf{ccl}_1(t))$ and $[t]_{\mathsf{C}} A \in \mathcal{E}_{\mathsf{E}}(w, \mathsf{ccl}_2(t))$. Hence, $\mathcal{M}, w \Vdash [\mathsf{ccl}_1(t)]_{\mathsf{E}} A$ and $\mathcal{M}, w \Vdash [\mathsf{ccl}_2(t)]_{\mathsf{E}} [t]_{\mathsf{C}} A$.

(induction) Assume $\mathcal{M}, w \Vdash A$ and $\mathcal{M}, w \Vdash [t]_{\mathsf{C}}(A \to [s]_{\mathsf{E}} A)$. From the second assumption and the reflexivity of R_{C}, we get $\mathcal{M}, w \Vdash A \to [s]_{\mathsf{E}} A$; thus, $\mathcal{M}, w \Vdash [s]_{\mathsf{E}} A$ by the first assumption. So $A \in \mathcal{E}_{\mathsf{E}}(w, s)$ and, by the second assumption, $A \to [s]_{\mathsf{E}} A \in \mathcal{E}_{\mathsf{C}}(w, t)$. 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 $\mathcal{M}, w \Vdash [t]_{\mathsf{C}}(A \to [s]_{\mathsf{E}} A)$, we get $\mathcal{M}, v' \Vdash A \to [s]_{\mathsf{E}} A$. Thus, $\mathcal{M}, v' \Vdash [s]_{\mathsf{E}} A$, 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 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 13. — *For a set Φ of formulae, we define*

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

DEFINITION 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. $W := \{w \subseteq \mathrm{Fm}_{\mathsf{LP}^{\mathsf{C}}_h} \ : \ w$ *is a maximal* $\mathcal{CS}$*-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 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 (Fitting, 2005)). We show the remaining five closure conditions:

(tupling) Assume $A \in \mathcal{E}_i(w,t_i)$ for all $1 \leq i \leq h$. By definition of $\mathcal{E}_i$, we have $[t_i]_i A \in w$ for all $1 \leq i \leq h$. 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, $[\mathsf{ccl}_1(t)]_{\mathsf{E}}A \in w$ and $[\mathsf{ccl}_2(t)]_{\mathsf{E}}\,[t]_{\mathsf{C}}A \in w$. Thus, $A \in \mathcal{E}_{\mathsf{E}}(w,\mathsf{ccl}_1(t))$ and $[t]_{\mathsf{C}}A \in \mathcal{E}_{\mathsf{E}}(w,\mathsf{ccl}_2(t))$.

(induction) Assume $A \in \mathcal{E}_{\mathsf{E}}(w,s)$ and $(A \to [s]_{\mathsf{E}}A) \in \mathcal{E}_{\mathsf{C}}(w,t)$. By definition of $\mathcal{E}_{\mathsf{E}}$ and $\mathcal{E}_{\mathsf{C}}$, we have $[s]_{\mathsf{E}}A \in w$ and $[t]_{\mathsf{C}}(A \to [s]_{\mathsf{E}}A) \in w$. From $\vdash_{\mathcal{CS}} [s]_{\mathsf{E}}A \to A$ (Lemma 1.1) and the induction axiom, it follows by maximal consistency that $A \in w$ and $[\mathsf{ind}(t,s)]_{\mathsf{C}}A \in w$. Therefore, $A \in \mathcal{E}_{\mathsf{C}}(w,\mathsf{ind}(t,s))$.

(monotonicity) We show only the case of $* = \mathsf{C}$ since the other cases are the same as in (Fitting, 2005). It is sufficient to prove by induction on the positive integer n that

$$\text{if } [t]_{\mathsf{C}}A \in w \text{ and } (w,v) \in (R_{\mathsf{E}})^n, \text{ then } [t]_C A \in v \ . \tag{7}$$

Base case $n = 1$**.** Assume $(w,v) \in R_{\mathsf{E}}$: *i.e.*, $w/i \subseteq v$ for some i. As $[t]_{\mathsf{C}}A \in w$, $[\pi_i \mathsf{ccl}_2(t)]_i\,[t]_{\mathsf{C}}A \in w$ 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 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 (Meyer *et al.*, 1995). 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, 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 $[(\dots (s \cdot t_1) \cdots t_{n-1}) \cdot t_n]_{\mathsf{C}} P \in w$—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 $R_{\mathsf{C}} \subsetneq R'_{\mathsf{C}}$.

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 $A_1, \dots, A_h$ such that $[t_i]_i A_i \in w$ for some $t_i \in \mathrm{Tm}_i$, but $A_i \notin v$. Now let $[c_i]_{\mathsf{C}}(A_i \to A_1 \vee \dots \vee A_h) \in \mathcal{CS}$ for constants $c_1, \dots, c_h$. Then $[\downarrow_i c_i \cdot t_i]_i (A_1 \vee \dots \vee A_h) \in w$ for all $1 \leq i \leq h$, so $[\langle \downarrow_1 c_1 \cdot t_1, \dots, \downarrow_h c_h \cdot t_h \rangle]_{\mathsf{E}} (A_1 \vee \dots \vee A_h) \in w$. However, $A_i \notin v$ for any $1 \leq i \leq h$; therefore, by the maximal consistency of v, $A_1 \vee \dots \vee A_h \notin v$ either. Hence, $w/\mathsf{E} \not\subseteq v$, so $(w,v) \notin R'_{\mathsf{E}}$. □

LEMMA 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 $w \in W$,*

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

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 (7) in the proof of Lemma 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.* (Meyer *et al.*, 1995)). 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 18 (COMPLETENESS). — $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS})$ *is sound and complete with respect to the class of epistemic models meeting* $\mathcal{CS}$*:* i.e.*, for all formulae* $A \in \mathrm{Fm}_{\mathsf{LP}_h^{\mathsf{C}}}$,

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

PROOF. — Soundness was already shown in Lemma 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 17, $\mathcal{M}, w \Vdash \neg A$, and hence, by Lemma 15, $\nVdash_{\mathcal{CS}} A$. ■

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 (Mkrtychev, 1997). We will now adapt M-models to our logic and prove the finite model property for it.

DEFINITION 19. — *An* M-model *is a singleton epistemic model.*

THEOREM 20 (COMPLETENESS WITH RESPECT TO M-MODELS). — $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS})$ *is also sound and complete with respect to the class of M-models meeting* $\mathcal{CS}$.

PROOF. — Soundness follows immediately from Lemma 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, $\mathcal{E}' := \mathcal{E} \restriction (W' \times \mathrm{Tm})$, and $\nu'(P_n) := \nu(P_n) \cap W'$.

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

$$\mathcal{M}, w_0 \Vdash [t]_{\circledast} C \text{ if and only if } C \in \mathcal{E}'_{\circledast}(w_0, t) \ . \tag{8}$$

Indeed, by Lemma 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 (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 (8) we have $C \notin \mathcal{E}'_{\circledast}(w_0, t)$, so $\mathcal{M}', w_0 \nVdash [t]_{\circledast} C$. ■

COROLLARY 21 (FINITE MODEL PROPERTY). — $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS})$ *enjoys the finite model property with respect to epistemic models.*

REMARK 22. — Note 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

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

being consistent, as shown in Remark 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. □

5. Conservativity

We extend the two-agent version LP^2 of the Logic of Proofs (Yavorskaya (Sidon), 2008) to an arbitrary h in the natural way and rename it in accordance with our naming scheme:

DEFINITION 23. — *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 (Fitting, 2008), for which evidence terms are essential.

DEFINITION 24. — *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. $([t]_{\circledast} A)^{\times} := \begin{cases} A^{\times} & \text{if } t \text{ contains a subterm } s \in \mathrm{Tm}_{\mathsf{E}} \cup \mathrm{Tm}_{\mathsf{C}}, \\ [t]_{\circledast} A^{\times} & \text{otherwise.} \end{cases}$

THEOREM 25. — *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 $D = [t_1]_1 B \wedge \cdots \wedge [t_h]_h B \to [\langle t_1, \ldots, t_h \rangle]_{\mathsf{E}} B$ 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 $C_1 \wedge \cdots \wedge C_h \to B^{\times}$ 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 $[t_1]_1 B^{\times} \wedge \cdots \wedge [t_h]_h B^{\times} \to B^{\times}$, 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.*, $[t]_{\mathsf{C}} B \to [\mathsf{ccl}_1(t)]_{\mathsf{E}} B$ or $[t]_{\mathsf{C}} B \to [\mathsf{ccl}_2(t)]_{\mathsf{E}}\, [t]_{\mathsf{C}} B$. Then $D^{\times}$ is the propositional tautology $B^{\times} \to B^{\times}$.

Case when $D = B \wedge [t]_{\mathsf{C}}(B \to [s]_{\mathsf{E}} B) \to [\mathsf{ind}(t,s)]_{\mathsf{C}} B$ is an instance of the induction axiom. Then $D^{\times}$ is the propositional tautology $B^{\times} \wedge (B^{\times} \to B^{\times}) \to B^{\times}$.

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 26. — 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 al-

ways an axiom of LP_h. To ensure that $\mathcal{CS}^\times$ is a proper constant specification, all formulae of the forms

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

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 (Fitting, 2008) for details). □

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 Sect. 1. 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.

In the modal language of common knowledge, modal formulae are defined by the grammar

$$A ::= P_j \mid \neg A \mid (A \wedge A) \mid (A \vee A) \mid (A \to A) \mid \Box_i A \mid \mathsf{E}A \mid \mathsf{C}A\ ,$$

where $P_j \in$ Prop. The set of all modal formulae is denoted by $\mathrm{Fm}_{\mathsf{S4}_h^{\mathsf{C}}}$. The Hilbert system $\mathsf{S4}_h^{\mathsf{C}}$ (Meyer *et al.*, 1995) is given by the modal axioms of $\mathsf{S4}$ for individual agents, by the necessitation rule for $\Box_1, \ldots, \Box_h$, and C, by modus ponens, and by the axioms

$$\mathsf{C}(A \to B) \to (\mathsf{C}A \to \mathsf{C}B), \qquad \mathsf{C}A \to A, \qquad \mathsf{E}A \leftrightarrow \Box_1 A \wedge \cdots \wedge \Box_h A,$$
$$A \wedge \mathsf{C}(A \to \mathsf{E}A) \to \mathsf{C}A, \qquad \mathsf{C}A \to \mathsf{E}(A \wedge \mathsf{C}A).$$

DEFINITION 27 (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* $P_j \in$ 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 28. — *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. ■

DEFINITION 29 (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 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.

The cut-free systems for $\mathsf{S4}^{\mathsf{C}}_h$ presented in (Alberucci *et al.*, 2005) and (Brünnler *et al.*, 2009) are 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 a finitary cut-free system (Jäger *et al.*, 2007). Unfortunately, the "somewhat unusual" structural properties of the resulting system (see discussion in (Jäger *et al.*, 2007)) make it hard to use it for realization.

The non-constructive, semantic realization method from (Fitting, 2005) cannot be applied directly because of the non-standard behavior of the canonical model (see Remark 16).

Perhaps the infinitary system presented in (Bucheli *et al.*, 2010b), which is finitely branching but admits infinite branches, can help in proving the realization theorem for $\mathsf{LP}^{\mathsf{C}}_h$. For now this remains work in progress.

7. Coordinated attack

To illustrate our logic, we will now analyze the coordinated attack problem along the lines of (Fagin *et al.*, 1995), 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 attack at all, *i.e.*, that they attack only if both know that they attack: $[t]_{\mathsf{C}}(\mathit{att} \rightarrow [s]_{\mathsf{E}}\, \mathit{att})$ for some terms s and t. Thus, by the induction axiom, we get $\mathit{att} \rightarrow [\mathsf{ind}(t,s)]_{\mathsf{C}}\, \mathit{att}$. Another reasonable assumption is that it is common knowledge that neither general attacks unless the message with the time of attack has been delivered: $[r]_{\mathsf{C}}(\mathit{att} \rightarrow \mathit{del})$ for some term r. Using the application axiom, we obtain $\mathit{att} \rightarrow [r \cdot \mathsf{ind}(t,s)]_{\mathsf{C}}\, \mathit{del}$.

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 (Fagin *et al.*, 1995), we provide a countermodel where m_2^H is the last message, it has been delivered, but H is unsure of that: *i.e.*,

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

for all terms s. Consider any model $\mathcal{M}$ where $W := \{0,1,2,3\}$, $\nu(\mathit{del}) := \{0,1,2\}$, 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 $\mathit{del} \in \mathcal{E}_H\left(0, m_1^G\right)$ and $\left[m_1^G\right]_H \mathit{del} \in \mathcal{E}_G\left(0, m_2^H\right)$. Whatever $\mathcal{E}_{\mathsf{C}}$ is, we have $\mathcal{M}, 0 \nVdash [s]_H \left[m_2^H\right]_G \left[m_1^G\right]_H \mathit{del}$ and $\mathcal{M}, 0 \nVdash [t]_{\mathsf{C}}\, \mathit{del}$ for any s and t because $\mathcal{M}, 3 \nVdash \mathit{del}$.

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 $\left[m_1^G\right]_H \mathit{del}$ 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: $W := \{w\}$, $R_i := \{(w,w)\}$, $\nu(\mathit{del}) := \{w\}$, and $\mathcal{E}$ is the minimal evidence function such that $\mathit{del} \in \mathcal{E}_H\left(w, m_1^G\right)$

and $\left[m_1^G\right]_H \mathit{del} \in \mathcal{E}_G\left(w, m_2^H\right)$. In this model, $M, w \nVdash [t]_{\mathsf{C}} \mathit{del}$ for any evidence term t because $\mathit{del} \notin \mathcal{E}_{\mathsf{C}}(w,t)$ for any t. To prove the latter statement, it is sufficient to note that for any term t, by Lemma 28,

$$\nvdash \left[m_1^G\right]_H \mathit{del} \wedge \left[m_2^H\right]_G \left[m_1^G\right]_H \mathit{del} \rightarrow [t]_{\mathsf{C}} \mathit{del} \tag{9}$$

because

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

which is easy to demonstrate. Let $\mathcal{M}^{\text{can}}$ be the canonical epistemic model meeting the empty constant specification and $\mathcal{E}^{\text{can}}$ be its evidence function. Since the negation of the formula from (9) must be satisfiable, for each t there is a world w_t from $\mathcal{M}^{\text{can}}$ such that $\mathit{del} \in \mathcal{E}_H^{\text{can}}\left(w_t, m_1^G\right)$ and $\left[m_1^G\right]_H \mathit{del} \in \mathcal{E}_G^{\text{can}}\left(w_t, m_2^H\right)$, but by the Truth Lemma 17, $\mathit{del} \notin \mathcal{E}_{\mathsf{C}}^{\text{can}}(w_t, t)$. Since $\mathcal{E}^{\text{can}} \restriction (\{w_t\} \times \mathrm{Tm})$ satisfies all the closure conditions, the minimality of $\mathcal{E}$ implies that $\mathcal{E}_{\mathsf{C}}(w,s) \subseteq \mathcal{E}_{\mathsf{C}}^{\text{can}}(w_t, s)$ for any term s. In particular, $\mathit{del} \notin \mathcal{E}_{\mathsf{C}}(w,t)$ for any term t.

8. Discussion

In this paper, 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 (Artemov, 2008), 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[3] and is contrasted with *tacit knowledge*. As described in (Nonaka, 1991), "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

3. 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 (Fagin *et al.*, 1995).

received by them[4], 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 (Fagin *et al.*, 1995). Clearly, an introduction of evidence terms cannot and should not change this general phenomenon.

However, there exist 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. A prime example of such a mode is, of course, public announcements, 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 (Plaza, 2007), 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 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 (Blass *et al.*, 2011). 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 considered a situation where agents can verify each other's evidence: $[t]_i A \to \left[!_i^j t\right]_j [t]_i A$ for $i \neq j$ (Yavorskaya (Sidon), 2008). 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: for any term $t \in \mathrm{Tm}_i$, there is a term $s(x) \in \mathrm{Tm}_{\mathsf{C}}$ such that $\vdash [t]_i A \to [s(t)]_{\mathsf{C}} A$. However, the mode of communication necessary for the

4. 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).

$!_i^j$-operation to work must be reliable and immediate, which restricts the applicability of such a logic; for instance, it precludes an analysis of asynchronous systems. 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 (Artemov, 2008), 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 in this paper 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 (Fagin *et al.*, 1995).

9. Conclusions

We have presented a justification logic $\mathsf{LP}_h^{\mathsf{C}}$ with common knowledge, which is a conservative extension of the multi-agent justification logic LP_h. The major open problem at the moment remains proving the realization theorem, one direction of which we have demonstrated.

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 concentrated 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 remains whether other semantics for justification logics such as (arithmetical) provability semantics (Artemov, 1995; Artemov, 2001) and game semantics (Renne, 2009b) can be adapted to $\mathsf{LP}_h^{\mathsf{C}}$. Further avenues of research include but are not limited to the decidability of $\mathsf{LP}_h^{\mathsf{C}}$, the comparison of its complexity to that of $\mathsf{S4}_h^{\mathsf{C}}$, and the extension of our treatment of common knowledge to the logics with the individual modalities of type K, $\mathsf{K5}$, *etc.*

A long-term goal of our research is to find justification counterparts of dynamic epistemic logics with common knowledge. A step in this direction (although still without common knowledge) was made in (Bucheli *et al.*, 2010a) by proposing a justification counterpart to public announcement logic. 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.

Acknowledgements

Bucheli and Kuznets are supported by the Swiss National Science Foundation grant 200021–117699. The authors would like to thank the anonymous reviewers for their helpful comments. Many thanks to Galina Savukova for editing the paper.

10. References

Alberucci L., Jäger G., "About cut elimination for logics of common knowledge", *Annals of Pure and Applied Logic*, vol. 133, num. 1–3, pp. 73–99, May, 2005.

Antonakos E., "Justified and Common Knowledge: Limited Conservativity", *in* S. N. Artemov, A. Nerode (eds), *Logical Foundations of Computer Science, International Symposium, LFCS 2007, New York, NY, USA, June 4–7, 2007, Proceedings*, vol. 4514 of *Lecture Notes in Computer Science*, Springer, pp. 1–11, 2007.

Artemov S. N., Operational modal logic, Technical Report num. MSI 95–29, Cornell University, December, 1995.

Artemov S. N., "Explicit Provability and Constructive Semantics", *Bulletin of Symbolic Logic*, vol. 7, num. 1, pp. 1–36, March, 2001.

Artemov S. N., "Justified common knowledge", *Theoretical Computer Science*, vol. 357, num. 1–3, pp. 4–22, July, 2006.

Artemov S. N., "The Logic of Justification", *The Review of Symbolic Logic*, vol. 1, num. 4, pp. 477–513, December, 2008.

Artemov S. N., "Tracking Evidence", *in* A. Blass, N. Dershowitz, W. Reisig (eds), *Fields of Logic and Computation, Essays Dedicated to Yuri Gurevich on the Occasion of His 70th Birthday*, vol. 6300 of *Lecture Notes in Computer Science*, Springer, pp. 61–74, 2010.

Artemov S. N., Kuznets R., "Logical Omniscience as a Computational Complexity Problem", *in* A. Heifetz (ed.), *Theoretical Aspects of Rationality and Knowledge, Proceedings of the Twelfth Conference (TARK 2009)*, ACM, pp. 14–23, 2009.

Blass A., Gurevich Y., Moskal M., Neeman I., "Evidential Authorization", *in* S. Nanz (ed.), *The Future of Software Engineering*, Springer, pp. 73–99, 2011.

Brünnler K., Goetschi R., Kuznets R., "A Syntactic Realization Theorem for Justification Logics", *in* L. Beklemishev, V. Goranko, V. Shehtman (eds), *Advances in Modal Logic, Volume 8*, College Publications, pp. 39–58, 2010.

Brünnler K., Studer T., "Syntactic cut-elimination for common knowledge", *Annals of Pure and Applied Logic*, vol. 160, num. 1, pp. 82–95, July, 2009.

Bucheli S., Kuznets R., Renne B., Sack J., Studer T., "Justified Belief Change", *in* X. Arrazola, M. Ponte (eds), *LogKCA-10, Proceedings of the Second ILCLI International Workshop on Logic and Philosophy of Knowledge, Communication and Action*, University of the Basque Country Press, pp. 135–155, 2010a.

Bucheli S., Kuznets R., Studer T., "Two Ways to Common Knowledge", *in* T. Bolander, T. Braüner (eds), *Proceedings of the 6th Workshop on Methods for Modalities (M4M–6 2009), Copenhagen, Denmark, 12–14 November 2009*, num. 262 in *Electronic Notes in Theoretical Computer Science*, Elsevier, pp. 83–98, May, 2010b.

Fagin R., Halpern J. Y., Moses Y., Vardi M. Y., *Reasoning about Knowledge*, MIT Press, 1995.

Fitting M., "The logic of proofs, semantically", *Annals of Pure and Applied Logic*, vol. 132, num. 1, pp. 1–25, February, 2005.

Fitting M., "Justification logics, logics of knowledge, and conservativity", *Annals of Mathematics and Artificial Intelligence*, vol. 53, num. 1–4, pp. 153–167, August, 2008.

Jäger G., Kretz M., Studer T., "Cut-free common knowledge", *Journal of Applied Logic*, vol. 5, num. 4, pp. 681–689, December, 2007.

Kuznets R., "Self-Referential Justifications in Epistemic Logic", *Theory of Computing Systems*, vol. 46, num. 4, pp. 636–661, May, 2010.

McCarthy J., Sato M., Hayashi T., Igarashi S., On the model theory of knowledge, Technical Report num. CS–TR–78–657, Stanford University Computer Science Department, April, 1978.

Meyer J.-J. Ch., van der Hoek W., *Epistemic Logic for AI and Computer Science*, vol. 41 of *Cambridge Tracts in Theoretical Computer Science*, Cambridge University Press, 1995.

Mkrtychev A., "Models for the Logic of Proofs", *in* S. Adian, A. Nerode (eds), *Logical Foundations of Computer Science, 4th International Symposium, LFCS'97, Yaroslavl, Russia, July 6–12, 1997, Proceedings*, vol. 1234 of *Lecture Notes in Computer Science*, Springer, pp. 266–275, 1997.

Nonaka I., "The Knowledge-Creating Company", *Harvard Business Review*, November–December, 1991.

Plaza J., "Logics of public communications", *Synthese*, vol. 158, num. 2, pp. 165–179, September, 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.

Renne B., "Evidence Elimination in Multi-Agent Justification Logic", *in* A. Heifetz (ed.), *Theoretical Aspects of Rationality and Knowledge, Proceedings of the Twelfth Conference (TARK 2009)*, ACM, pp. 227–236, 2009a.

Renne B., "Propositional games with explicit strategies", *Information and Computation*, vol. 207, num. 10, pp. 1015–1043, October, 2009b.

Yavorskaya (Sidon) T., "Interacting Explicit Evidence Systems", *Theory of Computing Systems*, vol. 43, num. 2, pp. 272–293, August, 2008.