Explicit Evidence Systems with Common Knowledge

Samuel Bucheli, Roman Kuznets*, and Thomas Studer

Institut für Informatik und angewandte Mathematik, Universität Bern
Bern, Switzerland
{bucheli, kuznets, tstuder}@iam.unibe.ch

Abstract. Justification logics are epistemic logics that explicitly include justifications for agents' knowledge. We present a multi-agent justification logic with evidence terms for individual agents as well as for common knowledge. A Kripke-style semantics similar to Fitting's semantics for the Logic of Proofs LP is defined and soundness and completeness with respect to it are shown along with the finite model property. Furthermore, we demonstrate that our logic is a conservative extension of Yavorskaya's minimal bimodal explicit evidence logic, a two-agent version of LP. We discuss the relationship to multi-agent modal logic S4 with common knowledge. Finally, as an illustration we give a brief analysis of the problem of coordinated attack in the new language.

1 Introduction

Justification logics [6] are epistemic logics that explicitly include justifications for agents' knowledge. The first such logic, the *Logic of Proofs* LP, was developed by Artemov [3, 4] to provide the modal logic S4 with provability semantics. The language of justification logics has also been used to provide a new approach to the logical omniscience problem [7] and to study self-referential proofs [15].

Instead of statements *A is known* denoted $\Box A$, justification logics reason about justifications directly by using the construct $[t]A$ to formalize statements *t is a justification for A*, where *evidence term t* can be viewed as an informal justification or a formal mathematical proof depending on the application. Evidence terms are built by means of operations that correspond to the axioms of S4 as can be seen from Fig. 1.

Artemov [4] showed that the Logic of Proofs LP is an *explicit counterpart* of the modal logic S4 in the following formal sense: each theorem of LP becomes a theorem of S4 if all terms are replaced with modality $\Box$ and, vice versa, each theorem of S4 can be turned into a theorem of LP by replacing occurrences of modality with suitable terms. The latter process is called *realization*, and the statement of correspondence is called a *realization theorem*. Although the operation $+$ introduced by the sum axiom in Fig. 1 does not have a modal

* The first and second authors are supported by Swiss National Science Foundation grant 200021–117699.

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)

Fig. 1. Axioms of S4 and LP

analog, it is an essential part of the proof of the realization theorem in [4]. Explicit counterparts for many normal modal logics between K and S5 have been developed, see a recent survey in [6] and a uniform proof of realization theorems for all single-agent justification logics forthcoming in [8].

The notion of *common knowledge* is essential in the area of multi-agent systems, where coordination among a set of agents is a central issue. The textbooks [11, 17] provide excellent introductions to epistemic logics in general and common knowledge in particular. 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

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

Artemov [5] created an explicit counterpart of McCarthy's *any fool knows* common knowledge modality [16], where common knowledge of A is defined as an arbitrary fixed point of (1). The relationship between the traditional common knowledge from [11, 17] and McCarthy's version is studied in [2].

In this paper, we present a multi-agent justification logic with evidence terms for individual agents as well as for common knowledge, with the intension to provide an explicit counterpart of the modal logic of traditional common knowledge $\mathsf{S4}_h^{\mathsf{C}}$.

To the best of our knowledge, the only existing multi-agent justification logics with evidence terms for each agent are due to Yavorskaya [20], where the two-agent case is considered. We will show that in the case of two agents our system is a conservative extension of Yavorskaya's minimal bimodal explicit evidence logic, which is an explicit counterpart of $\mathsf{S4}_2$.

An epistemic semantics for LP, *F-models*, was created by Fitting in [12] by augmenting Kripke models with an *evidence function* that specifies which formulae are evidenced by a term at a given world. It is easily extended to the whole family of single-agent justification logics (for details, see [6]). In [5] Artemov extends F-models to justification terms for McCarthy's common knowledge modality in the presence of several ordinary modalities, creating the most general type of epistemic models, sometimes called *AF-models*, where common evidence terms are given their own accessibility relation not directly dependent on the accessibility relations for individual modalities. Yavorskaya in [20] proves a stronger completeness theorem with respect to singleton F-models, independently introduced by Mkrtychev [18] and now known as *M-models*, where the role of the accessibility relation is completely taken over by the evidence function.

The paper is organized as follows. In Sect. 2, we introduce the language and give the 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 give a Fitting-style semantics similar to AF-models and prove soundness and completeness with respect to this semantics as well as with the 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 bimodal explicit evidence logic. In Sect. 6 we show how our logics are related to traditional modal logics of common knowledge and discuss the problem of realization. Finally, in Sect. 7 we provide an analysis of the problem of coordinated attack in our logic.

2 Syntax

To create an explicit counterpart of modal logic of common knowledge we use its axiomatization via the induction axiom from [17] rather than via the induction rule to facilitate the proof of 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. As motivated in [10], a proof of $\mathsf{C}A$ can be thought of as an infinite list of proofs of the conjuncts $\mathsf{E}^m A$ in 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 $\mathsf{ind}(\cdot,\cdot)$. On the other hand, to access 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 head and tail,

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

by means of two unary co-closure operations $\mathsf{ccl}_1(\cdot)$ and $\mathsf{ccl}_2(\cdot)$. Evidence terms for mutual knowledge are represented as tuples of individual agents' evidence terms with the standard operation of tupling and h unary projections. While only two of the three operations on LP terms are adopted for common knowledge evidence and none for mutual knowledge evidence, it will be shown in Sect. 3 that most remaining operations are definable with the notable exception of inspection for mutual knowledge.

We consider a system of h agents. Throughout the whole paper, 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 $\circledast$. The sets Tm_1, $\dots$, Tm_h, Tm_{E}, and Tm_{C} of *evidence terms for individual agents*, *mutual*, and *common knowledge* respectively are inductively defined as follows:

1. $\mathrm{Cons}_{\circledast} \subseteq \mathrm{Tm}_{\circledast}$;
2. $\mathrm{Var}_{\circledast} \subseteq \mathrm{Tm}_{\circledast}$;
3. $!_i t \in \mathrm{Tm}_i$ for any $t \in \mathrm{Tm}_i$;
4. $t +_* s \in \mathrm{Tm}_*$ and $t \cdot_* s \in \mathrm{Tm}_*$ for any $t, s \in \mathrm{Tm}_*$;
5. $\langle t_1, \dots, t_h \rangle \in \mathrm{Tm}_{\mathsf{E}}$ for any $t_1 \in \mathrm{Tm}_1$, $\dots$, $t_h \in \mathrm{Tm}_h$;
6. $\pi_i t \in \mathrm{Tm}_i$ for any $t \in \mathrm{Tm}_{\mathsf{E}}$;
7. $\mathsf{ccl}_1(t) \in \mathrm{Tm}_{\mathsf{E}}$ and $\mathsf{ccl}_2(t) \in \mathrm{Tm}_{\mathsf{E}}$ for any $t \in \mathrm{Tm}_{\mathsf{C}}$;
8. $\mathsf{ind}(t, s) \in \mathrm{Tm}_{\mathsf{C}}$ for any $t \in \mathrm{Tm}_{\mathsf{C}}$ and $s \in \mathrm{Tm}_{\mathsf{E}}$.

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

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

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

where $t \in \mathrm{Tm}_{\circledast}$. The set of all formulae is denoted by $\mathrm{Fm}_{\mathsf{LPC}_h}$. 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

1. all propositional tautologies
2. $[t]_*(A \to B) \to ([s]_* A \to [t \cdot s]_* B)$ (application)
3. $[t]_* A \to [t + s]_* A, \quad [s]_* A \to [t + s]_* A$ (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

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

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 *pure*, 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. The deductive system $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS})$ is the Hilbert systems given by the axioms of $\mathsf{LP}_h^{\mathsf{C}}$ above and rules modus ponens and axiom necessitation:

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

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

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

For an arbitrary $\mathcal{CS}$, we write $\Delta \vdash_{\mathcal{CS}} A$ to state that A is derivable from Δ in $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS})$ and omit the mention of $\mathcal{CS}$ when working with the constant specification from (2) by writing $\Delta \vdash A$. We use Δ, A to mean $\Delta \cup \{A\}$.

3 Basic Properties

In this section we show that our logics possess 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 exist, which follows from Lemma 26 in Sect. 6 and the fact that $\mathsf{E}A \to \mathsf{EE}A$ is not a valid modal formula.

We begin with the following easy observation.

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* $\vdash_{\mathcal{CS}} [t]_{\mathsf{E}} A \to [t +_{\mathsf{E}} s]_{\mathsf{E}} A$ *and* $\vdash_{\mathcal{CS}} [s]_{\mathsf{E}} A \to [t +_{\mathsf{E}} s]_{\mathsf{E}} A$. ($\mathsf{E}$-*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* $t \in \mathrm{Tm}_{\mathsf{C}}$. ($\mathsf{C}$-*reflexivity*)

Proof. 1. Immediate by projection and reflexivity axiom.
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 the statements from the previous lemma, the ones from the next lemma require the constant specification $\mathcal{CS}$ to 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* $\vdash_{\mathcal{CS}} [t]_{\mathsf{C}} A \to [!_{\mathsf{C}} t]_{\mathsf{C}} [t]_{\mathsf{C}} A$. ($\mathsf{C}$-*inspection*)
2. *For any* $t \in \mathrm{Tm}_{\mathsf{C}}$ *there is a term* $\Leftarrow t \in \mathrm{Tm}_{\mathsf{C}}$ *such that* $\vdash_{\mathcal{CS}} [t]_{\mathsf{C}} A \to [\Leftarrow t]_{\mathsf{C}} [\mathsf{ccl}_1(t)]_{\mathsf{E}} A$. ($\mathsf{C}$-*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 following two theorems are standard in justification logics. Their proofs can be taken almost word for word from [4] 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 $t \in \mathrm{Tm}_{\circledast}$,*

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

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, accordingly for $\Delta(x/t, P/B)$ and $\mathcal{CS}(x/t, P/B)$.

The following lemma states an important property, namely that our logic can internalize its own proofs. It can be shown by induction on the derivation of A.

Lemma 5 (C-lifting). *Let $\mathcal{CS}$ be a pure C-axiomatically appropriate constant specification. 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 $t_{\circledast}(x_1, \ldots, x_n, y_1, \ldots, y_m) \in \mathrm{Tm}_{\circledast}$ 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 $y_1, \ldots, y_m \in \mathrm{Tm}_{\circledast}$.

Corollary 6 (Constructive necessitation). *Let $\mathcal{CS}$ be a pure 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 $\vdash_{\mathcal{CS}} [t]_{\circledast} A$.*

Corollary 7 (Internalized induction rule). *Let $\mathcal{CS}$ be a pure C-axiomatically appropriate constant specification. For any formula A, if $\vdash_{\mathcal{CS}} A \to [s]_{\mathsf{E}} A$, there is a term $t \in \mathrm{Tm}_{\mathsf{C}}$ such that $\vdash_{\mathcal{CS}} A \to [\mathsf{ind}(t, s)]_{\mathsf{C}} A$.*

4 Soundness and Completeness

Definition 8. *An* AF-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 $W \neq \varnothing$, *a function $R : \{1, \ldots, h\} \to \mathcal{P}(W \times W)$ that assigns a reflexive and transitive* accessibility relation *on W to each agent $i \in \{1, \ldots, h\}$, and a* truth valuation $\nu : \mathrm{Prop} \to \mathcal{P}(W)$. *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 $\mathcal{E} \colon W \times \mathrm{Tm} \to \mathcal{P}\left(\mathrm{Fm}_{\mathsf{LP}^{\mathsf{c}}_h}\right)$ *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 9. *A ternary relation* $\mathcal{M}, w \Vdash A$ *for* formula A being satisfied at a world $w \in W$ in an AF-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. $\Vdash$ *respects 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* $\Vdash_{\mathcal{CS}} A$ *and say that the formula* A *is* valid with respect to $\mathcal{CS}$ *if* $\mathcal{M} \Vdash A$ *for all AF-models* $\mathcal{M}$ *meeting* $\mathcal{CS}$.

Lemma 10 (Soundness). *Provable formulae are valid:* $\vdash_{\mathcal{CS}} A$ *implies* $\Vdash_{\mathcal{CS}} A$.

Proof. Let $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ be an AF-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 modus ponens and axiom necessitation rules are the same as for the single agent case in [12] and are, therefore, omitted. Of the remaining four axioms we show two representative cases:

(co-closure) Assume $\mathcal{M}, w \Vdash [t]_\mathsf{C} A$. Then 1) $\mathcal{M}, v \Vdash A$ for all $v \in W$ with $(w,v) \in R_\mathsf{C}$ and 2) $A \in \mathcal{E}_\mathsf{C}(w,t)$. It follows from 1) that, for all $v' \in W$ with $(w,v') \in R_\mathsf{E}$, we have $\mathcal{M}, v' \Vdash A$ since $R_\mathsf{E} \subseteq R_\mathsf{C}$; also, due to the monotonicity closure condition, $\mathcal{M}, v' \Vdash [t]_\mathsf{C} A$ since $R_\mathsf{E} \circ R_\mathsf{C} \subseteq R_\mathsf{C}$. From 2), by the co-closure closure condition, $A \in \mathcal{E}_\mathsf{E}(w, \mathsf{ccl}_1(t))$ 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 $A \in \mathcal{E}_\mathsf{C}(w, \mathsf{ind}(t,s))$. It remains to prove $\mathcal{M}, v \Vdash A$ for all $v \in W$ with $(w,v) \in R_\mathsf{C}$. This can easily be done by showing $\mathcal{M}, v \Vdash A$ for all $v \in W$ with $(w,v) \in (R_\mathsf{E})^n$ by induction on $n \in \mathbb{N}$. Finally, we conclude that $\mathcal{M}, w \Vdash [\mathsf{ind}(t,s)]_\mathsf{C} A$. □

Definition 11. *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 but 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 12. *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 13. *Let $\mathcal{CS}$ be a constant specification. The* canonical AF-model $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ meeting $\mathcal{CS}$ *is defined as follows:*

1. $W := \{w \subseteq \mathrm{Fm}_{\mathsf{LP}_h^{\mathsf{C}}} \ : \ w \textit{ is a maximal } \mathcal{CS}\textit{-consistent set}\}$;
2. $R_i := \{(w, v) \in W \times W \ : \ w/i \subseteq v\}$;
3. $\mathcal{E}_{\circledast}(w, t) := \{A \in \mathrm{Fm}_{\mathsf{LP}_h^{\mathsf{C}}} \ : \ [t]_{\circledast} A \in w\}$;
4. $\nu(P_n) := \{w \in W \ : \ P_n \in w\}$.

Lemma 14. *Let $\mathcal{CS}$ be a constant specification. The canonical AF-model meeting $\mathcal{CS}$ is an AF-model meeting $\mathcal{CS}$.*

Proof. The proof of 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 [12]). Of the remaining five closure conditions we show two representative cases:

(induction) Assume $A \in \mathcal{E}_{\mathsf{E}}(w, s)$ and $(A \to [s]_{\mathsf{E}} A) \in \mathcal{E}_{\mathsf{C}}(w, t)$. Then 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 [12]. It is sufficient to prove by induction on $n \in \mathbb{N}$ that

$$\text{if } [t]_{\mathsf{C}} A \in w \text{ and } (w, v) \in (R_{\mathsf{E}})^n, \text{ then } [t]_C A \in v \ . \tag{3}$$

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 $A \in \mathcal{E}_{\mathsf{C}}(w, t)$, i.e. $[t]_{\mathsf{C}} A \in w$. As shown above, $[t]_{\mathsf{C}} A \in v$. Thus, $A \in \mathcal{E}_{\mathsf{C}}(v, t)$. □

Remark 15. Let R'_{C} denote the binary relation on W given 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, the converse does not hold for any pure C-axiomatically appropriate constant specification, which we show by adapting an example from [17]. Let

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

This set is consistent for any $P \in \mathrm{Prop}$ because it is easy to construct a model for each finite subset of Φ. Thus, there is a maximal consistent set $w \supseteq \Phi$. Set $\Psi := \{\neg P\} \cup (w/\mathsf{C})$. Ψ is also consistent, otherwise, by Corollary 6, there would have existed a term s such that $[s]_{\mathsf{C}} P \in w$, which would contradict the consistency of w. Let v be a maximal consistent set that contains Ψ, i.e. $v \supseteq \Psi$.

Clearly, $w/\mathsf{C} \subseteq v$, i.e. $(w, v) \in R'_{\mathsf{C}}$, but $(w, v) \notin R_{\mathsf{C}}$ because that would imply $P \in v$, which cannot happen. It follows that $R_{\mathsf{C}} \subsetneq R'_{\mathsf{C}}$.

Similarly, we could 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.

Lemma 16 (Truth Lemma). *Let $\mathcal{CS}$ be a constant specification and $\mathcal{M}$ be the canonical AF-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 the definition of $\Vdash$ and 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, $\mathcal{M}, v \Vdash B$. From this $\mathcal{M}, w \Vdash A$ immediately follows.

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 get $B \in \mathcal{E}_{\mathsf{E}}(w, t)$. By 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 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 $n \in \mathbb{N}$. As in the previous cases, $B \in \mathcal{E}_{\mathsf{C}}(w, t)$ by definition of $\mathcal{E}_{\mathsf{C}}$. By (3) we find $A \in v$ and thus, by C-reflexivity and maximal consistency, also $B \in v$. 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. [17]. The last case, in particular, usually requires more sophisticated methods that guarantee the finiteness of the model.

Theorem 17 (Completeness). $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS})$ *is sound and complete with respect to the class of AF-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 has already been shown in Lemma 10. For completeness let $\mathcal{M}$ be the canonical AF-model meeting $\mathcal{CS}$ and assume $\nvdash_{\mathcal{CS}} A$. Then $\{\neg A\}$ is $\mathcal{CS}$-consistent and hence contained in some maximal $\mathcal{CS}$-consistent set $w \in W$. So, by Lemma 16, $\mathcal{M}, w \Vdash \neg A$ and hence, by Lemma 14, $\nVdash_{\mathcal{CS}} A$. □

M-models were introduced as semantics for LP by Mkrtychev [18]. They form a subclass of F-models (see [12]).

Definition 18. *An* M-model *is a singleton AF-model.*

Theorem 19 (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 10. 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 AF-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. $W' := \{w_0\}$, $R'_{\circledast} := \{(w_0, w_0)\}$ for any $\circledast$, $\mathcal{E}' := \mathcal{E} \restriction (W' \times \mathrm{Tm})$, and $\nu'(P_n) := \nu(P_n) \cap W'$.

Since $\mathcal{M}'$ is clearly an M-model meeting $\mathcal{CS}$, it 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

$$\mathcal{M}, w_0 \Vdash [t]_{\circledast} C \text{ if and only if } C \in \mathcal{E}'_{\circledast}(w_0, t) \ . \tag{4}$$

Indeed, by Lemma 16, $\mathcal{M}, w_0 \Vdash [t]_{\circledast} C$ if and only if $[t]_{\circledast} C \in w_0$ which, by definition of the canonical AF-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 (4) 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 (4) we have $C \notin \mathcal{E}'_{\circledast}(w, t)$ and thus $\mathcal{M}', w_0 \nVdash [t]_{\circledast} C$. □

Corollary 20 (Finite model property). $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS})$ *enjoys the finite model property with respect to AF-models.*

5 Conservativity

Yavorskaya in [20] introduced a 2-agent version of LP, which we extend to an arbitrary h in the natural way:

Definition 21. *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 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 a technique from [13].

Definition 22. *The mapping* $\times : \mathrm{Fm}_{\mathsf{LP}_h^{\mathsf{C}}} \to \mathrm{Fm}_{\mathsf{LP}_h}$ *is defined as follows:*

1. $P^{\times} = P$ *for propositional variables* $P \in \mathrm{Prop}$*;*
2. $\times$ *commutes with propositional connectives;*
3. $([t]_{\circledast} A)^{\times} = \begin{cases} A^{\times} & \text{if } t \text{ contains a subterm } s \in \mathrm{Tm}_{\mathsf{E}} \cup \mathrm{Tm}_{\mathsf{C}}, \\ [t]_{\circledast} A^{\times} & \text{otherwise.} \end{cases}$

Theorem 23. *Let $\mathcal{CS}$ be a constant specification for $\mathsf{LP}_h^{\mathsf{C}}$. For an arbitrary formula $A \in \mathrm{Fm}_{\mathsf{LP}_h}$, if $\mathsf{LP}_h^{\mathsf{C}}(\mathcal{CS}) \vdash A$ then $\mathsf{LP}_h(\mathcal{CS}^\times) \vdash A$.*

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 an instance of the application axiom

$$[t]_*(B \to C) \to ([s]_* B \to [t \cdot s]_* C).$$

We distinguish the following possibilities:
1. Both t and s contain a subterm from $\mathrm{Tm}_{\mathsf{E}} \cup \mathrm{Tm}_{\mathsf{C}}$. Then $D^\times$ has the form $(B^\times \to C^\times) \to (B^\times \to C^\times)$, which is a 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 $D^\times$ is $(B^\times \to C^\times) \to ([s]_i B^\times \to C^\times)$, 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 case, $\times$ 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.

Cases of other axioms are similar and, thus, omitted here. The only other situation when an axiom is not mapped to an axiom is an instance of the tupling axiom $[t_1]_1 B \wedge \cdots \wedge [t_h]_h B \to [\langle t_1, \ldots, t_h \rangle]_{\mathsf{E}} B$ with none of t_i's containing any subterms from $\mathrm{Tm}_{\mathsf{E}} \cup \mathrm{Tm}_{\mathsf{C}}$.
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 $\mathsf{LP}_h^{\mathsf{C}}$ and hence $B^\times$ is derivable in $\mathsf{LP}_h(\mathcal{CS}^\times)$ as shown above; in the latter case, $[c]_i B^\times \in \mathcal{CS}^\times$. **Case** of modus ponens is trivial. □

Remark 24. 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 $(A \to B) \to ([s]_i A \to B)$ and $[t_1]_1 A \wedge \cdots \wedge [t_h]_h A \to A$ 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 [13] 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 terms are replaced with the corresponding modalities, which is one direction of the realization theorem. We also discuss approaches to the harder opposite direction.

We start with recalling the modal language of common knowledge. Modal formulae are defined by the following 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 $P_j \in \mathrm{Prop}$. The set of all modal formulae is denoted by $\mathrm{Fm}_{\mathsf{S4}^{\mathsf{C}}_h}$.

The Hilbert system $\mathsf{S4}^{\mathsf{C}}_h$ [17] is given by the modal axioms of $\mathsf{S4}$ for individual agents, necessitation rule for $\Box_1, \ldots, \Box_h$ and C, modus ponens and 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 25 (Forgetful projection). *The mapping* $\circ : \mathrm{Fm}_{\mathsf{LP}^{\mathsf{C}}_h} \to \mathrm{Fm}_{\mathsf{S4}^{\mathsf{C}}_h}$ *is defined as follows:*
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. $([t]_{\mathsf{E}} A)^\circ = \mathsf{E}A^\circ$*;*
5. $([t]_{\mathsf{C}} A)^\circ = \mathsf{C}A^\circ$.

Lemma 26. *Let* $\mathcal{CS}$ *be any 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 induction on the derivation of A. □

Definition 27 (Realization). *A realization is a mapping* $r : \mathrm{Fm}_{\mathsf{S4}^{\mathsf{C}}_h} \to \mathrm{Fm}_{\mathsf{LP}^{\mathsf{C}}_h}$ *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) by evidence terms of the corresponding type. The problem of realization for a given pure C-axiomatically appropriate constant specification $\mathcal{CS}$ can be stated 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 constitutes the harder direction of the realization theorem, which is often demonstrated using induction on a cut-free sequent proof of the modal formula.

Cut-free systems for $\mathsf{S4}^{\mathsf{C}}_h$ are presented in [1] and [9]. They 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, realization of such a rule meets with serious difficulties in reaching uniformity among the realizations of the approximants $\mathsf{E}^m A$.

A finitary cut-free system is obtained in [14] by finitizing this ω-rule via the finite model property. Unfortunately, the "somewhat unusual" structural

properties of the resulting system (see discussion in [14]) make it hard to use for realization.

The non-constructive, semantical realization method from [12] cannot be applied directly because of the non-standard behavior of the canonical model, see Remark 15.

Perhaps the infinitary system presented in [10], which is finitely branching but admits infinite branches, can help in proving the realization theorem for $\mathsf{LP}_h^{\mathsf{C}}$. For now this remains work in progress.

7 Coordinated attack

To illustrate our logic we will now analyze the problem of coordinated attack along the lines of [11], where also additional references can be found. Let us briefly recall this classical problem. Suppose two divisions of an army, located at distinct places, are about to attack an 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 an 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 was not delivered. But if general H sends an acknowledgment m_2^H, he in turn cannot be sure whether the acknowledgement reached G prompting yet another acknowledgement 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 among 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: $[t]_{\mathsf{C}}(att \to [s]_{\mathsf{E}}\, att)$ for some terms s and t. So by the induction axiom we get $att \to [\mathsf{ind}(t,s)]_{\mathsf{C}}\, att$. 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 $att \to [r \cdot \mathsf{ind}(t,s)]_{\mathsf{C}}\, del$.

We now show that common knowledge of *del* cannot be achieved and, therefore, no attack will take place, no matter how many messages and acknowledgements m_1^G, m_2^H, m_3^G, ... are sent by the generals even in the case all 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 [11], we provide a countermodel where m_2^H is the last message, it has been delivered, but H is unsure of that, i.e. $\left[m_1^G\right]_H del$, $\left[m_2^H\right]_G \left[m_1^G\right]_H del$, but $\neg\, [s]_H \left[m_2^H\right]_G \left[m_1^G\right]_H del$ for all terms s. Indeed, consider the model $\mathcal{M}$ with

$W := \{0,1,2,3\}$, $\nu(del) := \{0,1,2\}$, R_G being the reflexive closure of $\{(1,2)\}$, R_H being the reflexive closure of $\{(0,1),(2,3)\}$, and any evidence function $\mathcal{E}$ such that $del \in \mathcal{E}_H(0, m_1^G)$ and $\big[m_1^G\big]_H del \in \mathcal{E}_G(0, m_2^H)$. Then, whatever $\mathcal{E}_{\mathsf{C}}$ is, $\mathcal{M}, 0 \nVdash [s]_H \big[m_2^H\big]_G \big[m_1^G\big]_H del$ and $\mathcal{M}, 0 \nVdash [t]_{\mathsf{C}} del$ for any s and t because $\mathcal{M}, 3 \nVdash del$.

In our models with explicit evidence there is an alternative possibility for the lack of knowledge: the absence of an acceptable evidence. To give an example, G may receive the acknowledgment m_2^H but not consider it to be an admissible evidence for $\big[m_1^G\big]_H del$ 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 admissible common knowledge evidence. A corresponding M-model $\mathcal{M} = (W, R, \mathcal{E}, \nu)$ is obtained as follows: $W := \{w\}$, $R_i := \{(w,w)\}$, $\nu(del) := \{w\}$, and $\mathcal{E}$ is the minimal evidence functions such that $del \in \mathcal{E}_H(w, m_1^G)$ and $\big[m_1^G\big]_H del \in \mathcal{E}_G(w, m_2^H)$. In this model $M, w \nVdash [t]_{\mathsf{C}} del$ for any evidence term t because $del \notin \mathcal{E}_{\mathsf{C}}(w,t)$ for any t. To show the latter statement, note that for any term t, by Lemma 26,

$$\nvdash \big[m_1^G\big]_H del \wedge \big[m_2^H\big]_G \big[m_1^G\big]_H del \to [t]_{\mathsf{C}} del \tag{5}$$

because $\mathsf{S4}_h^{\mathsf{C}} \nvdash \Box_H del \wedge \Box_G \Box_H del \to \mathsf{C} del$, which is easy to show. Thus, the negation of the formula from (5) is satisfiable, and for each t there is a world w_t in the canonical AF-model with evidence function $\mathcal{E}^{\mathrm{can}}$ such that $del \in \mathcal{E}_H^{\mathrm{can}}(w_t, m_1^G)$ and $\big[m_1^G\big]_H del \in \mathcal{E}_G^{\mathrm{can}}(w_t, m_2^H)$, but by the Truth Lemma 16, $del \notin \mathcal{E}_{\mathsf{C}}^{\mathrm{can}}(w_t, t)$. Since $\mathcal{E}^{\mathrm{can}} \restriction (\{w_t\} \times \mathrm{Tm})$ satisfies all the closure conditions, 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.

8 Conclusions

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

Our analysis of the problem of coordinated attack in the language of $\mathsf{LP}_h^{\mathsf{C}}$ shows that the access to explicit 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 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 mixed constant specification, where only certain aspects of reasoning are common knowledge.

We established soundness and completeness with respect to AF-models and singleton M-models. Can other semantics for justification logics such as (arithmetical) provability semantics [3, 4] and game semantics [19] be adapted to $\mathsf{LP}_h^{\mathsf{C}}$?

There are further interesting questions: Is $\mathsf{LP}_h^{\mathsf{C}}$ decidable and, if yes, what is its complexity compared to that of $\mathsf{S4}_h^{\mathsf{C}}$? How robust is our treatment of common knowledge if the individual modalities are taken to be of type K, $\mathsf{K5}$, etc.?

References

[1] Luca Alberucci and Gerhard Jäger. About cut elimination for logics of common knowledge. *Annals of Pure and Applied Logic*, 133(1–3):73–99, 2005.

[2] Evangelia Antonakos. Justified and common knowledge: Limited conservativity. In S. N. Artemov and A. Nerode, editors, *LFCS 2007*, volume 4514 of *LNCS*, pages 1–11. Springer, 2007.

[3] Sergei N. Artemov. Operational modal logic. Technical Report MSI 95–29, Cornell University, December 1995.

[4] Sergei N. Artemov. Explicit provability and constructive semantics. *Bulletin of Symbolic Logic*, 7(1):1–36, 2001.

[5] Sergei [N.] Artemov. Justified common knowledge. *Theoretical Computer Science*, 357(1–3):4–22, 2006.

[6] Sergei [N.] Artemov. The logic of justification. *The Review of Symbolic Logic*, 1(4):477–513, 2008.

[7] Sergei [N.] Artemov and Roman Kuznets. Logical omniscience as a computational complexity problem. In A. Heifetz, editor, *TARK 2009*, pages 14–23. ACM, 2009.

[8] Kai Brünnler, Remo Goetschi, and Roman Kuznets. A Syntactic Realization Theorem for Justification Logics. Submitted, 2010.

[9] Kai Brünnler and Thomas Studer. Syntactic cut-elimination for common knowledge. *Annals of Pure and Applied Logic*, 160(1):82–95, 2009.

[10] Samuel Bucheli, Roman Kuznets, and Thomas Studer. Two ways to common knowledge. In T. Bolander and T. Braüner, editors, *M4M 2009*, number 128 in Comp. Sci. Research Reports, pages 73–87. Roskilde University, Denmark, 2009.

[11] Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning about Knowledge*. MIT Press, 1995.

[12] Melvin Fitting. The logic of proofs, semantically. *Annals of Pure and Applied Logic*, 132(1):1–25, 2005.

[13] Melvin Fitting. Justification logics, logics of knowledge, and conservativity. *Annals of Mathematics and Artificial Intelligence*, 53(1–4):153–167, 2008.

[14] Gerhard Jäger, Mathis Kretz, and Thomas Studer. Cut-free common knowledge. *Journal of Applied Logic*, 5(4):681–689, 2007.

[15] Roman Kuznets. Self-referential justifications in epistemic logic. *Theory of Computing Systems*, 46(4):636–661, 2010.

[16] John McCarthy, Masahiko Sato, Takeshi Hayashi, and Shigeru lgarashi. On the model theory of knowledge. Technical Report CS–TR–78–657, Stanford University Computer Science Department, April 1978.

[17] J.-J. Ch. Meyer and W. van der Hoek. *Epistemic Logic for AI and Computer Science*. Cambridge University Press, 1995.

[18] Alexey Mkrtychev. Models for the logic of proofs. In S. Adian and A. Nerode, editors, *LFCS 1997*, volume 1234 of *LNCS*, pages 266–275. Springer, 1997.

[19] Bryan Renne. Propositional games with explicit strategies. *Information and Computation*, 207(10):1015–1043, 2009.

[20] Tatiana Yavorskaya (Sidon). Interacting explicit evidence systems. *Theory of Computing Systems*, 43(2):272–293, 2008.