Evidence-Based Common Knowledge

(Technical Report TR-2004018 CUNY Ph.D. Program in Computer Science, revised version)

Sergei Artemov

Graduate Center CUNY
Ph.D. Program in Computer Science
365 Fifth Avenue, New York, NY 10016
SArtemov@gc.cuny.edu

April 4, 2005

Abstract

In this paper we introduce a new type of knowledge operator, called *evidence-based knowledge*, intended to capture the constructive core of common knowledge. An evidence-based knowledge system is obtained by augmenting a multi-agent logic of knowledge with a system of evidence assertions $t{:}\varphi$ ("t is an evidence for φ") based on the following plausible assumptions: 1) each axiom has evidence; 3) evidence is checkable; 3) any evidence implies individual knowledge for each agent. Normally, the following monotonicity property is also assumed: 4) any piece of evidence is compatible with any other evidence. We show that the evidence-based knowledge operator is a stronger version of the common knowledge operator. Evidence-based knowledge is free of logical omniscience, model-independent, and has a natural motivation. Furthermore, evidence-based knowledge can be presented by normal multi-modal logics, which are in the scope of well-developed machinery applicable to modal logic: epistemic models, normalized proofs, automated proof search, etc.

1 Introduction

Common knowledge is a fundamental feature of multi-agent systems of knowledge which was first discussed in [Lewis, 1969] and then studied in [Aumann, 1976; McCarthy *et al.*, 1979; Lehman, 1984; Halpern and Moses, 1990]. [Fagin *et al.*, 1995] provides an excellent introduction to logics of knowledge in general and to common knowledge phenomena in particular. Let $K_1, K_2, \ldots, K_n$ stand for knowledge operators in an n-agent logic of knowledge and

$$E\varphi \; = \; K_1\varphi \wedge K_2\varphi \wedge \ldots \wedge K_n\varphi.$$

Then the common knowledge operator C corresponding to $K_1, K_2, \ldots, K_n$ is informally defined as an infinite conjunction

$$C\varphi \; \Leftrightarrow \; \varphi \wedge E\varphi \wedge E^2\varphi \wedge \ldots \wedge E^n\varphi \ldots .$$

In a Kripke-style model for $K_1, K_2, \ldots, K_n$ the common knowledge operator is formally defined as the modality of reachability along paths that use accessibility edges corresponding to any of $K_1, K_2, \ldots, K_n$.

The traditional way to capture common knowledge deductively is to use the Fixed-Point Axiom

$$C\varphi \ \leftrightarrow \ E(\varphi \wedge C\varphi)$$

along with the Induction Rule

$$\frac{\varphi \rightarrow E(\psi \wedge \varphi)}{\varphi \rightarrow C\psi} \quad ,$$

capturing the greatest solution of the corresponding fixed-point equation (cf. [Fagin *et al.*, 1995]). This kind of deductive system does not behave well proof-theoretically. In particular, there is no conventional cut-elimination in the common-knowledge systems ([Alberucci and Jaeger, 2005]). This practically rules out automated proof search and severely limits the usage of formal methods in analyzing knowledge. Semi-formal model theoretical methods in this area have their own problems, both foundational and practical. For example, paradigmatic solutions of well-known puzzles like Muddy Children, Wise Men, Unfaithful Wives, etc. (cf. [Fagin *et al.*, 1995]), use a very strong, not formalized assumption that the agents possess a common knowledge of the same Kripke-style frame of possible situations.

As presented above, common knowledge captures the most liberal version of knowledge operator satisfying the Fixed Point Axiom, without imposing any conditions on the way this knowledge is attained. As a result, there might be nonconstructive versions of the common knowledge appearing by chance or for some unknown reasons or without any particular reasons at all. Here is an informal example[1] intended to show the difference between common knowledge and knowledge based on evidence.

> Bob, a graduate student, fails a standardized, multiple-choice qualifier in his area of research. His advisor, with whom Bob has worked extensively in this particular area, as well as his other professors–all of whom have awarded him top grades–know that it's not possible for Bob to have failed this exam. They do not even consider such a possibility. In fact, Bob hasn't failed the exam, rendering the faculty's belief as common knowledge. It is soon determined that a crease in Bob's test form was responsible for reversing his results; once this has been taken into account and the correct results made available, the faculty acquires evidence-based knowledge that Bob has indeed passed the qualifier.

We take this example to suggest that there is a certain need for evidence-based knowledge systems, in particular, for analyzing social situations.

In this paper we introduce a family of new knowledge operators representing so-called *evidence-based knowledge* (*EBK*). An (*EBK*)-system is obtained by augmenting a multi-agent logic of knowledge with a system of evidence assertions $t{:}\varphi$ ("t is an evidence for φ") based on the following plausible assumptions:

- all axioms have evidence;
- evidence is undeniable and implies individual knowledge of any agent.

[1]This version of the example was suggested by Karen Kletter.

• evidence is checkable;

• evidence is monotone, i.e., any new piece of evidence does not spoil a given one.

An important feature of *EBK*-systems is their graceful handling of the logical omniscience problem: an agent cannot claim to have evidence-based knowledge without having actually built a supporting evidence term.

In addition, we introduce a forgetful version of the evidence-based knowledge operator $J\varphi$ ("there is a justification for φ") obtained by collapsing all evidence terms into one modality J:

$$t{:}\varphi \mapsto J\varphi.$$

The forgetful *EBK*-systems are normal modal logics with standard Kripke-style semantics. Moreover, for any valid fact of forgetful *EBK*-systems, one could recover its constructive meaning by realizing all forgetful modalities $J\varphi$ by appropriate evidence terms $t{:}\varphi$.

Here is a brief comparison of the forgetful *EBK*-operator $J\varphi$ with the common knowledge operator $C\varphi$.

Informally,

$$J\varphi \;\Rightarrow \varphi \wedge E\varphi \wedge E^2\varphi \wedge \ldots \wedge E^n\varphi \ldots,$$

but the converse "$\Leftarrow$" does not necessarily hold. Such a $J\varphi$ is not necessarily unique, which means that we have a variety of evidence-based knowledge operators.

In the epistemic Kripke-style semantics, $J\varphi$ corresponds to any accessibility relation which contains (but does not necessarily coincides with) reachability.

Forgetful evidence-based logics postulate J as a normal (usually S4-like) modality, and contain enough machinery to prove the Fixed-Point Axiom for J:

$$J\varphi \;\leftrightarrow\; E(\varphi \wedge J\varphi)$$

The Induction Rule is not valid for J. This means that, unlike the common knowledge, the evidence-based knowledge is not committed to capturing the greatest solution of the corresponding fixed-point equation, but rather represents its generic solution.

In many traditional problems where the common knowledge has been used, the evidence-based knowledge operator is also applicable. Moreover, *EBK*-systems have certain advantages.

1. Forgetful *EBK*-systems are easier to justify since the question of whether a given real system has an evidence-based knowledge can be reduced to checking a manageable set of conditions. An axiomatic approach to common knowledge in the form of *EBK*-systems is model-independent and seems to avoid foundational loopholes of the standard model-theoretical reasoning about common knowledge.

2. Forgetful *EBK*-logics are simpler than the traditional common knowledge systems. Evidence-based knowledge is in the scope of well-developed methods in modal logic, both semantical and proof theoretic. These include cut-elimination theorems that yield the possibility of automated proof search and verification, which have been ruled out in the traditional common knowledge systems because of their prohibitive proof theoretical complexity.

3. Common-knowledge operator is a derivative of the agent knowledge operators and carries the features of the latter. The evidence-based knowledge component can be chosen independently of knowledge systems of individual agents, which provides an additional degree of flexibility.

2 The content of this paper

Arguably, the first paper raising the issue of epistemic logic with justification was [van Benthem, 1991]. Such systems, along with the usual knowledge modalities, were supposed to contain evidence assertions of the format "t is evidence for φ," denoted as $t{:}\varphi$. There are a variety of formal systems for describing evidence which could serve as a formal base for the "evidence component" here. The first system of explicit terms capturing a modal logic, $\mathsf{S4}$, was found in [Artemov, 1995; Artemov, 2001] and known as the logic of proofs LP. A similar system corresponding to $\mathsf{S5}$ was introduced in [Artemov *et al.*, 1999]. Finally, [Brezhnev, 2000; Brezhnev, 2001] describes systems of terms corresponding to K, $\mathsf{K4}$, T, D, $\mathsf{D4}$. These systems of explicit terms share several important features. Among these are the ability to internalize their own proofs as schematized by the Internalization Principle:

$$\textit{if} \vdash \varphi, \textit{ then} \vdash p{:}\varphi \textit{ for some proof term } p,$$

and the validity of Realization Theorems which assert that one can retrieve explicit evidence terms from the proof of any theorem provable in the underlying modal logic. As a result, the forgetful projection of the logic of explicit terms is exactly the counterpart modal logic, e.g., $\mathsf{S4}$ is the forgetful projection of LP. There are also other systems of explicit presentation of knowledge by evidence terms ("+"-free fragment of LP, [Artemov, 2001], functional logic of proofs [Krupski, 2002; Krupski, 2005], etc.), where compatibility of evidence is not required.

Along with the usual choices of K, T, $\mathsf{K4}$, $\mathsf{S4}$, and $\mathsf{S5}$ for base logics of knowledge of individual agents, this shows that the number of possible *EBK*-systems is rather high. We consider three representative cases, all using the logic of proofs LP as their evidence component: $\mathsf{T}_n\mathsf{LP}$, $\mathsf{S4}_n\mathsf{LP}$, and $\mathsf{S5}_n\mathsf{LP}$. In all these systems, the evidence logic is LP (which corresponds to $\mathsf{S4}$), whereas the base knowledge logics could be weaker (T in $\mathsf{T}_n\mathsf{LP}$), equal to ($\mathsf{S4}$ in $\mathsf{S4}_n\mathsf{LP}$), or stronger than ($\mathsf{S5}$ in $\mathsf{S5}_n\mathsf{LP}$) the evidence logic. All these *EBK*-systems are supplied with epistemic semantics capturing the notion of admissible evidence.

We also consider "forgetful counterparts" of the above *EBK*-systems: T_n^J, $\mathsf{S4}_n^J$, and $\mathsf{S5}_n^J$ obtained from $\mathsf{T}_n\mathsf{LP}$, $\mathsf{S4}_n\mathsf{LP}$, and $\mathsf{S5}_n\mathsf{LP}$ respectively, by collapsing

$$t{:}\varphi \mapsto J\varphi.$$

The intended epistemic semantics of $J\varphi$ is "there is a justification for φ." The forgetful *EBK*-systems T_n^J, $\mathsf{S4}_n^J$, and $\mathsf{S5}_n^J$ are normal modal logics with standard Kripke-style semantics. We show that T_n^J, $\mathsf{S4}_n^J$, and $\mathsf{S5}_n^J$ enjoy an important Realization Property: given a formula φ in the forgetful language derivable in T_n^J, $\mathsf{S4}_n^J$, or $\mathsf{S5}_n^J$, one could recover an evidence-carrying formula ψ derivable in the corresponding *EBK*-system $\mathsf{T}_n\mathsf{LP}$, $\mathsf{S4}_n\mathsf{LP}$, or $\mathsf{S5}_n\mathsf{LP}$ respectively, such that φ is a forgetful projection of ψ. The Realization Property opens a possibility of first establishing φ in a forgetful *EBK*-system T_n^J, $\mathsf{S4}_n^J$, or $\mathsf{S5}_n^J$, which can be a relatively easy task, and recovering constructive evidence terms behind occurrences of the forgetful modality J in φ later, only if needed.

We make an easy but fundamental observation that evidence assertions $t{:}\varphi$, as well as the forgetful evidence-based knowledge modality J, satisfy the Fixed-Point Axiom above and hence may be regarded as a special sort of common knowledge. In terms of accessibility relations in Kripke-style models, $J\varphi$ corresponds to a transitive and reflexive relation R containing (but not necessarily coinciding with) the reachability relation.

In fact, $\mathsf{S4}_n^J$ coincides with one of systems introduced axiomatically in [McCarthy *et al.*, 1979] with the modality J in $\mathsf{S4}_n^J$ corresponding to the "any fool knows" modality. This provides McCarthy's dummy "any fool" agent with a justification in the form of evidence-based epistemic semantics.

On the technical side, we prove that T_n^J and $\mathsf{S4}_n^J$ enjoy cut-elimination theorems and give algorithms of recovering evidence terms in "any fool knows" modalities in all three forgetful systems T_n^J, $\mathsf{S4}_n^J$, and $\mathsf{S5}_n^J$. We also find epistemic models for each of the three systems above and establish the corresponding completeness theorems.

In Section 7, we give a complete account of the correspondence between evidence-based knowledge $J\varphi$ and common knowledge $C\varphi$. There are more *EBK*-systems than common knowledge systems. When both *EBK* and common knowledge exist, the *EBK*-modality $J\varphi$ is stronger than the common knowledge modality $C\varphi$:

$$J\varphi \ \Rightarrow C\varphi \quad \text{but} \quad C\varphi \ \not\Rightarrow \ J\varphi.$$

Each valid *EBK*-identity is common knowledge-compliant, which justifies using *EBK*-systems as common knowledge systems. In particular,

$$(\mathsf{S4}_n^J)^* \subset \ \mathsf{S4}_n^C \ \text{ but } \ (\mathsf{S4}_n^J)^* \neq \ \mathsf{S4}_n^C \ ,$$

where $*$ stands for an operation of renaming J to C.

As an example, a solution of the wise men puzzle (Section 8) is given as a formal derivation in T_3^J. Hence this solution counts as a common knowledge solution as well.

3 Formal systems of evidence-based knowledge

We first introduce the multi-agent logics of evidence-based knowledge series $\mathsf{T}_n\mathsf{LP}$, $n = 1, 2, 3....$ In brief, $\mathsf{T}_n\mathsf{LP}$ contains n copies of T-style modalities representing knowledge operators of n agents, $K_1, \ldots, K_n$ (cf. systems T_n from [Fagin *et al.*, 1995]); in addition, it contains a system of evidence assertions taken from the logic of proofs LP.

Evidence assertions in $\mathsf{T}_n\mathsf{LP}$ have the form $t{:}\varphi$, where φ is a formula and t is an evidence term (or just *evidence*) built from constants $a, b, c, \ldots$ and variables $x, y, z, \ldots$ with the help of three operations, *application* "$\cdot$" (binary), *union* "+" (binary), and *inspection* "!" (unary). Formally, if t is an evidence and S is a sentence variable, the formulas of $\mathsf{T}_n\mathsf{LP}$ are defined by the following grammar

$$\varphi = \bot \mid S \mid \varphi_1 \rightarrow \varphi_2 \mid \varphi_1 \wedge \varphi_2 \mid \varphi_1 \vee \varphi_2 \mid \neg\varphi \mid K_i\varphi \mid t{:}\varphi.$$

We assume also that "$t{:}$" "K_i," and "$\neg$" bind stronger than "$\wedge, \vee$" which, in turn, bind stronger than "$\rightarrow$." $\mathsf{T}_n\mathsf{LP}$ has axioms of both T_n and LP, together with the principle that an evidence assertion yields knowledge of each individual agent: $t{:}\varphi \rightarrow K_i\varphi$. This is a schema when there is one growing system of evidence accepted by all the agents.

Definition 1. Axioms and rules of $\mathsf{T}_n\mathsf{LP}$ are

I. **Classical propositional logic**
The standard set of axioms of the classical propositional logic, e.g.,

A1-A10 from [Kleene, 1952] or a similar system
R1. *Modus Ponens*

II. **Knowledge principles**
Axioms and rules of T for each individual knowledge operator K_i.

B1$_i$. $K_i(\varphi\rightarrow\psi)\rightarrow(K_i\varphi\rightarrow K_i\psi)$
B2$_i$. $K_i\varphi\rightarrow\varphi$
R2$_i$. $\vdash\varphi\ \Rightarrow\vdash K_i\varphi$ (*knowledge generalization*)

III. **Evidence principles** (axioms and rules of the logic of proofs LP)

E1. $s{:}(\varphi\rightarrow\psi)\ \rightarrow\ (t{:}\varphi\rightarrow(s\cdot t){:}\psi)$ (*application*)
E2. $t{:}\varphi\ \rightarrow !t{:}(t{:}\varphi)$ (*inspection*)
E3. $s{:}\varphi\rightarrow(s+t){:}\varphi,\quad t{:}\varphi\rightarrow(s+t){:}\varphi$ (*union*)
E4. $t{:}\varphi\rightarrow\varphi$ (*reflexivity*)
R3. $\vdash c{:}A$, where A is an axiom from I - IV and c is a proof constant
(*evidence for axioms*)

IV. **Principle connecting evidence and knowledge**

C1. $t{:}\varphi\rightarrow K_i\varphi$ (*undeniability of evidence*).

Group III introduces some combinatorial properties of evidence and explains the meaning of evidence terms. E1 is nothing but the internalized *modus ponens*, which says that an evidence for $\varphi\rightarrow\psi$ can be applied to an evidence for φ to produce an evidence for ψ. E2 expresses the principle that any evidence t of φ can be verified by a new evidence $!t$ (this is similar to a proof checking principle in the logic of proofs). E3 reflects the principle of consistency and monotonicity of evidence: if t is an evidence for φ, then t combined with any other evidence still remains an evidence for φ. R3 assigns initial evidence in the form of constants to any axiom of $\mathsf{T}_n\mathsf{LP}$. This is a formal representation of the basic assumption in evidence-based logics that all axioms have been certified and their justifications have been accepted by all the agents. Finally, E4 is redundant and immediately follows from B2$_i$ and C1.

Naturally, all axioms are in fact schemas in the language of $\mathsf{T}_n\mathsf{LP}$. All rules are applied across sections I-IV.

Consider two more series of principles:

B3$_i$. $K_i\varphi\rightarrow K_iK_i\varphi$ *Positive Introspection*,
B4$_i$. $\neg K_i\varphi\rightarrow K_i\neg K_i\varphi$ *Negative Introspection.*

System $\mathsf{S4}_n\mathsf{LP}$ is obtained from $\mathsf{T}_n\mathsf{LP}$ by adding B3$_i$ and $\mathsf{S5}_n\mathsf{LP}$ is $\mathsf{S4}_n\mathsf{LP}$ plus B4$_i$, $i=1,\ldots,n$. Again, all the rules R2$_i$ are extended to these new axioms as well.

All these systems are closed under substitutions of evidence terms for evidence variables and formulas for propositional variables, and enjoy the deduction theorem $\Gamma,\varphi\vdash\psi\ \Rightarrow\Gamma\vdash\varphi\rightarrow\psi$.

Lemma 1. *For any formula φ and each $i=1,2,\ldots,n$ there are evidence terms $\mathtt{up}_i(x)$ a such that $\mathsf{T}_n\mathsf{LP}$ (hence $\mathsf{S4}_n\mathsf{LP}$ and $\mathsf{S5}_n\mathsf{LP}$) proves $x{:}\varphi\rightarrow\mathtt{up}_i(x){:}K_i\varphi$.*

Proof.

$x{:}\varphi \to K_i\varphi$, by C1;
$a{:}(x{:}\varphi \to K_i\varphi)$, introducing evidence a, by R3;
$!x{:}x{:}\varphi \to (a\cdot!x){:}K_i\varphi$, by E1 and propositional logic;
$x{:}\varphi \to !x{:}x{:}\varphi$, by E2;
$x{:}\varphi \to (a\cdot!x){:}K_i\varphi$, by propositional logic.

It suffices now to put $\mathtt{up}_i(x)$ equal to $a\cdot!x$ such that $a{:}(x{:}\varphi \to K_i\varphi)$. □

Proposition 1. [Internalization] *Given* $\mathsf{T}_n\mathsf{LP} \vdash \varphi$*, there is an evidence term* p *such that* $\mathsf{T}_n\mathsf{LP} \vdash p{:}\varphi$*. The same holds for* $\mathsf{S4}_n\mathsf{LP}$ *and* $\mathsf{S5}_n\mathsf{LP}$*.*

Proof. Induction on a derivation of φ. *Base*: φ is an axiom. Then use R3. In this case, p is an atomic evidence (a constant). *Induction step*:

1. φ is obtained from $\psi \to \varphi$ and ψ by *modus ponens*. By the induction hypothesis, $\vdash s{:}(\psi \to \varphi)$ and $\vdash t{:}\psi$ for some evidence terms s and t. Hence by E1, $\vdash (s\cdot t){:}\varphi$, so p is $s\cdot t$.

2. If φ is obtained by $\mathrm{R2}_i$, then φ is $K_i\psi$ and $\vdash \psi$. By the induction hypothesis, $\vdash t{:}\psi$ for some evidence t. Use Lemma 1 to conclude that $\vdash \mathtt{up}_i(t){:}K_i\psi$, and p is $\mathtt{up}_i(t)$.

3. If φ is obtained by R3, then φ is $c{:}A$ for some constant c and axiom A. Use the evidence inspection axiom E2 to derive $!c{:}c{:}A$, i.e., $!c{:}\varphi$. Here p is $!c$.

Note that the evidence term p is ground and built from atomic evidence terms by applications and inspections only. Moreover, the whole derivation of $p{:}\varphi$ is carried out inside the LP-part of $\mathsf{S4}_n\mathsf{LP}$, in particular, $\mathrm{R2}_i$ has not been used. □

A similar argument establishes a more general form of internalization: *If* $\psi_1, \ldots, \psi_k \vdash \varphi$*, then for some evidence* $p(x_1, \ldots, x_k)$*,*

$$x_1{:}\psi_1, \ldots, x_k{:}\psi_k \vdash p(x_1, \ldots, x_k){:}\varphi.$$

Both of the previous formulations of internalization follow from

Proposition 2. [Lifting]
If $\psi_1, \ldots, \psi_k, y_1{:}\chi_1, \ldots, y_n{:}\chi_n \vdash \varphi$*, then for some evidence* $p(x_1, \ldots, x_k, y_1, \ldots, y_n)$*,*

$$x_1{:}\psi_1, \ldots, x_k{:}\psi_k, y_1{:}\chi_1, \ldots, y_n{:}\chi_n \vdash p(x_1, \ldots, x_k, y_1, \ldots, y_n){:}\varphi.$$

Proof. Similar to Proposition 1 with two new base clauses. If φ is ψ_i, then p is x_i. If φ is $y_j{:}\chi$, then p is $!y_j$. □

The internalization property states that all derived facts have witnesses. Internalization naturally extends to the case when $\mathsf{T}_n\mathsf{LP}$, $\mathsf{S4}_n\mathsf{LP}$, or $\mathsf{S5}_n\mathsf{LP}$ are augmented by new axioms, each of which has witnessing evidence (e.g., has the form $t{:}\psi$ for some evidence t).

Lemma 2. *For any formula* φ *and each* $i = 1, 2, \ldots, n$ *there are evidence terms* $\mathtt{down}_i(x)$ *such that* $\mathsf{T}_n\mathsf{LP}$ *(hence* $\mathsf{S4}_n\mathsf{LP}$ *and* $\mathsf{S5}_n\mathsf{LP}$*) proves* $x{:}K_i\varphi \rightarrow \mathtt{down}_i(x){:}\varphi$.

Proof.

$x{:}K_i\varphi \rightarrow \varphi$,	by E4 and B2$_i$;
$b{:}(x{:}K_i\varphi \rightarrow \varphi)$,	introducing evidence b, by Proposition 1;
$!x{:}x{:}K_i\varphi \rightarrow (b\cdot !x){:}\varphi$,	by E1 and propositional logic;
$x{:}K_i\varphi \rightarrow\, !x{:}x{:}K_i\varphi$,	by E2;
$x{:}K_i\varphi \rightarrow (b\cdot !x){:}\varphi$,	by propositional logic.

It suffices now to put $\mathtt{down}_i(x)$ equal to $b\cdot !x$ such that $b{:}(x{:}K_i\varphi \rightarrow \varphi)$ □

A natural assumption about common knowledge is that φ is common knowledge (written $C\varphi$) iff all agents know that φ and $C\varphi$. This leads to the *Fixed-Point Axiom* (cf. [Fagin *et al.*, 1995]):

$$C\varphi \leftrightarrow E(\varphi \wedge C\varphi),$$

where $E\varphi = K_1\varphi \wedge \ldots \wedge K_n\varphi$. We show that $t{:}\varphi$ provably satisfies a similar fixed-point identity in $\mathsf{T}_n\mathsf{LP}$, $\mathsf{S4}_n\mathsf{LP}$, and $\mathsf{S5}_n\mathsf{LP}$.

Proposition 3. *For each evidence term* t, $t{:}\varphi$ *satisfies the Fixed-Point Axiom for common knowledge in* $\mathsf{T}_n\mathsf{LP}$ *(* $\mathsf{S4}_n\mathsf{LP}$*,* $\mathsf{S5}_n\mathsf{LP}$*).*

Proof. We prove that $\mathsf{T}_n\mathsf{LP} \vdash t{:}\varphi \leftrightarrow E(\varphi \wedge t{:}\varphi)$.

1. $t{:}\varphi \rightarrow K_i\varphi$, for all $i = 1, \ldots, n$, hence $t{:}\varphi \rightarrow E\varphi$
2. $t{:}\varphi \rightarrow\, !t{:}t{:}\varphi$, hence $t{:}\varphi \rightarrow K_i t{:}\varphi$, for all $i = 1, \ldots, n$, and $t{:}\varphi \rightarrow Et{:}\varphi$
3. $t{:}\varphi \rightarrow E(\varphi \wedge t{:}\varphi)$,

which concludes the left-to-right part of the proof. The right-to-left part $E(\varphi \wedge t{:}\varphi) \rightarrow t{:}\varphi$ is straightforward. □

4 Models

Kripke-style models for modal logics with justifications were introduced in [Artemov, 1994] and then generalized in [Nogina, 1994; Nogina, 1996; Sidon, 1997; Yavorskaya (Sidon), 2002; Artemov and Nogina, 2004]. In the last of those papers, Kripke semantics was adopted for $\mathsf{S4}_1\mathsf{LP}$. Special models capturing evidence were developed in [Mkrtychev, 1997; Fitting, 2003] for the logic of proofs LP, [Artemov and Nogina, 2004; Fitting, 2004] for $\mathsf{S4}_1\mathsf{LP}$.

In this section, we will introduce models for all three systems $\mathsf{T}_n\mathsf{LP}$, $\mathsf{S4}_n\mathsf{LP}$, and $\mathsf{S5}_n\mathsf{LP}$ that contain the aforementioned Kripke, Mkrtychev, and Fitting models as special cases.

At the heart of this semantics lies the idea, which can be traced back to [Mkrtychev, 1997; Fitting, 2003], of augmenting Boolean (Mkrtychev) or Kripke-style (Fitting) models by an evidence function which assigns "admissible" evidence terms to a statement, regardless of its truth value. The statement $t{:}\varphi$ holds in a given world u iff both of the following conditions are met: 1) t is an admissible evidence for φ in u; 2) φ holds in all worlds accessible from u.

An important new feature of these models is a new *evidence* accessibility relation R. This innovation provides an additional flexibility in the choice of R so that we may capture all of the above-mentioned models and provide an adequate epistemic semantics for a wide range of systems.

A $\mathsf{T}_n\mathsf{LP}$-*frame* is a structure $(W, R_1, \ldots, R_n, R)$, where W is a non-empty set of *states* (*possible worlds*); $R_1, \ldots, R_n$ are binary relations on W called *accessibility* relations, associated with agents $1, \ldots, n$ respectively; and R is a binary *evidence accessibility* relation on W. The relations $R_1, \ldots, R_n$ are reflexive, R is reflexive and transitive, and R contains all R_i's. Hence R contains the transitive closure of $R_1 \cup \ldots \cup R_n$ but does not necessarily coincide with it. In other words, if v is reachable from u by a finite number of $R_1, \ldots, R_n$-edges, then uRv but the converse is not necessarily true.

Given a frame $(W, R_1, \ldots, R_n, R)$, a *possible evidence* function $\mathcal{E}$ is a mapping from states and evidence terms to sets of formulas. We can read $\varphi \in \mathcal{E}(u,t)$ as "φ is one of the formulas for which t serves as possible evidence in state u." An evidence function must obey conditions that respect the intended meanings of the operations on evidence terms.

Definition 2. $\mathcal{E}$ is an evidence function on $(W, R_1, \ldots, R_n, R)$ if for all evidence terms s and t, for all formulas φ and ψ, and for all $u, v \in W$:

1. *Monotonicity*: uRv implies $\mathcal{E}(u,t) \subseteq \mathcal{E}(v,t)$.
2. *Application*: $\varphi \to \psi \in \mathcal{E}(u,s)$ and $\varphi \in \mathcal{E}(u,t)$ implies $\psi \in \mathcal{E}(u, s \cdot t)$.
3. *Inspection*: $\varphi \in \mathcal{E}(u,t)$ implies $t{:}\varphi \in \mathcal{E}(u, !t)$.
4. *Sum*: $\mathcal{E}(u,s) \cup \mathcal{E}(u,t) \subseteq \mathcal{E}(u, s+t)$.

A $\mathsf{T}_n\mathsf{LP}$-model is a structure $\mathcal{M} = (W, R_1, \ldots, R_n, R, \mathcal{E}, \Vdash)$, where $(W, R_1, \ldots, R_n, R)$ is a frame with an evidence function $\mathcal{E}$ on $(W, R_1, \ldots, R_n, R)$ and $\Vdash$ is an arbitrary mapping from sentence variables to subsets of W.

Given a model $\mathcal{M} = (W, R_1, \ldots, R_n, R, \mathcal{E}, \Vdash)$, a forcing relation $\Vdash$ is extended from sentence variables to all formulas by the following rules. For each $u \in W$:

1. $\Vdash$ respects Boolean connectives at each world.
2. $u \Vdash K_i\varphi$ iff $v \Vdash \varphi$ for every $v \in W$ with uR_iv.
3. $u \Vdash t{:}\varphi$ iff $\varphi \in \mathcal{E}(u,t)$ and $v \Vdash \varphi$ for every $v \in W$ with uRv.

Informally speaking, $t{:}\varphi$ is true at a given world u iff t is an acceptable evidence for φ in u and φ is true at all worlds v accessible from u via a given evidence accessibility relation R. We say φ is *true* at world $u \in W$ if $u \Vdash \varphi$; otherwise, φ is *false* at u. A formula φ is *true* in a model if φ is true at each world of the model; φ is valid if φ is true in every model.

A *constant specification* is a map $\mathcal{CS}$ from evidence constants to (possibly empty) sets of axioms. A constant specification $\mathcal{CS}$ is *full*, if it entails internalization (Proposition 1). The proof of Proposition 1 demonstrates that for a constant specification to be full, it is sufficient to have a constant for each axiom.

Given a constant specification $\mathcal{CS}$, a model $\mathcal{M}$ meets $\mathcal{CS}$ if $\mathcal{M} \Vdash a{:}\varphi$ whenever $\varphi \in \mathcal{CS}(a)$. A derivation (in any of $\mathsf{T}_n\mathsf{LP}$, $\mathsf{S4}_n\mathsf{LP}$, or $\mathsf{S5}_n\mathsf{LP}$) meets $\mathcal{CS}$ if whenever rule R3 is used to produce $a{:}\varphi$, then $\varphi \in \mathcal{CS}(a)$.

$\mathsf{S4}_n\mathsf{LP}$- and $\mathsf{S5}_n\mathsf{LP}$-models are defined as $\mathsf{T}_n\mathsf{LP}$-models with only this difference: for $\mathsf{S4}_n\mathsf{LP}$-models, the accessibility relations $R_1, \ldots, R_n$ are reflexive and transitive; in $\mathsf{S5}_n\mathsf{LP}$-models, $R_1, \ldots, R_n$ are reflexive, transitive, and symmetric.

A set S of formulas is $\mathcal{CS}$-$\mathsf{S4}_n\mathsf{LP}$-*satisfiable* ($\mathcal{CS}$-$\mathsf{T}_n\mathsf{LP}$-*satisfiable*, $\mathcal{CS}$-$\mathsf{S5}_n\mathsf{LP}$-*satisfiable*) if there is an $\mathsf{S4}_n\mathsf{LP}$-model ($\mathsf{T}_n\mathsf{LP}$-model, $\mathsf{S5}_n\mathsf{LP}$-models) $\mathcal{M}$, meeting $\mathcal{CS}$, and a world u in it such that $\mathcal{M}, u \Vdash \varphi$ for all $\varphi \in S$.

The usual Kripke models for T_n, $\mathsf{S4}_n$, and $\mathsf{S5}_n$ are $\mathsf{T}_n\mathsf{LP}$-, $\mathsf{S4}_n\mathsf{LP}$-, and $\mathsf{S5}_n\mathsf{LP}$-models respectively, where the evidence part (R and $\mathcal{E}$) is ignored. Mkrtychev models[2] for LP are single-world $\mathsf{T}_n\mathsf{LP}$-models. Fitting models[3] for $\mathsf{LPS4}$ are $\mathsf{S4}_1\mathsf{LP}$-models with $R_1 = R$. Kripke models for $\mathsf{S4}_1\mathsf{LP}$ + *weak negative introspection* $\neg t{:}\varphi \rightarrow \Box(\neg t{:}\varphi)$ from [Artemov and Nogina, 2004] are $\mathsf{S4}_1\mathsf{LP}$-models with $R = W \times W$.

Theorem 1. [Completeness Theorem] *Let $\mathcal{CS}$ be a constant specification. A formula φ is proved in $\mathsf{T}_n\mathsf{LP}$ ($\mathsf{S4}_n\mathsf{LP}$, $\mathsf{S5}_n\mathsf{LP}$) meeting $\mathcal{CS}$ iff φ holds in all $\mathsf{T}_n\mathsf{LP}$-models (respectively, $\mathsf{S4}_n\mathsf{LP}$-models, $\mathsf{S5}_n\mathsf{LP}$-models) meeting $\mathcal{CS}$.*

Proof. We will give a proof for $\mathsf{S4}_n\mathsf{LP}$ making note of how to modify this proof for the remaining cases of $\mathsf{T}_n\mathsf{LP}$ and $\mathsf{S5}_n\mathsf{LP}$.

Soundness is straightforward; we will check $t{:}\varphi \rightarrow K_i\varphi$ (axiom C1) only. Suppose $u \Vdash t{:}\varphi$, then $v \Vdash \varphi$ for all v such that uRv. Since $R_i \subseteq R$, $v \Vdash \varphi$ for all v such that uR_iv, hence $u \Vdash K_i\varphi$.

Completeness is proved using a maximal consistent set construction properly adapted for evidence-based multi-agent systems. A set of formulas Γ is *consistent* if there is no finite subset $\varphi_1, \ldots, \varphi_n$ such that $\neg(\varphi_1 \wedge \ldots \wedge \varphi_n)$ is provable in $\mathsf{S4}_n\mathsf{LP}$ meeting $\mathcal{CS}$. A consistent set Γ is *maximal consistent* if for any formula ψ, either $\psi \in \Gamma$ or $\neg\psi \in \Gamma$. By the standard Lindenbaum construction, each consistent set can be extended to a maximal consistent set. We define the *canonical model* $\mathcal{M} = (W, R_1, \ldots, R_n, R, \mathcal{E}, \Vdash)$ for $\mathsf{S4}_n\mathsf{LP}$ with a given constant specification $\mathcal{CS}$.

1. W is the collection of all maximal consistent sets.

2. If Γ is a set of formulas, let $\Gamma^{\sharp_i} = \{\varphi \mid K_i\varphi \in \Gamma\}$ and $\Gamma^{\flat} = \{\varphi \mid t{:}\varphi \in \Gamma\}$. Now define the accessibility relations $R_1, \ldots, R_n$ as follows:

$$\Gamma R_i \Delta \ \text{ iff } \ \Gamma^{\sharp_i} \subseteq \Delta, \ i = 1, \ldots, n.$$

Note that R_i are reflexive and transitive (for $\mathsf{S4}_n\mathsf{LP}$). For $\mathsf{S5}_n\mathsf{LP}$, relations R_i are also symmetric. Suppose $\Gamma R_i \Delta$ and $K_i\varphi \in \Delta$. We claim that $K_i\varphi \in \Gamma$, hence $\varphi \in \Gamma$ and $\Delta R_i \Gamma$. Indeed, suppose $K_i\varphi \notin \Gamma$, then by maximality, $\neg K_i\varphi \in \Gamma$. By the axiom $\neg K_i\varphi \rightarrow K_i\neg K_i\varphi$, $K_i\neg K_i\varphi \in \Gamma$. Since $\Gamma R_i \Delta$, $\neg K_i\varphi \in \Delta$, which contradicts the consistency of Δ.

Define the evidence accessibility relation R as follows:

$$\Gamma R \Delta \ \text{ iff } \ \Gamma^{\flat} \subseteq \Delta.$$

Note that R is reflexive. Moreover, R is transitive. Indeed, let $\Gamma R \Delta$ and $\Delta R \Theta$. If $t{:}\varphi \in \Gamma$, then $!t{:}t{:}\varphi \in \Gamma$ and $t{:}\varphi \in \Delta$. Likewise, $!t{:}t{:}\varphi \in \Delta$ and $t{:}\varphi \in \Theta$. By reflexivity, $\varphi \in \Theta$.

[2] Called *pre-models* in [Mkrtychev, 1997].

[3] Called *weak models* in [Fitting, 2004].

Let us check that R contains all R_i's, $i = 1, \ldots, n$. Suppose $\Gamma R_i \Delta$ and $t{:}\varphi \in \Gamma$. Then $!t{:}t{:}\varphi \in \Gamma$ and $K_i t{:}\varphi \in \Gamma$ hence, $t{:}\varphi \in \Delta$. By reflexivity, $\varphi \in \Delta$.

3. Define the evidence function $\mathcal{E}$ as follows

$$\mathcal{E}(\Gamma, t) = \{\varphi \mid t{:}\varphi \in \Gamma\}.$$

To show that $\mathcal{E}$ is an evidence function, we must prove that it satisfies conditions of Definition 2. *Application*, *Inspection*, and *Sum* are straightforward. For *Monotonicity*, assume $\varphi \in \mathcal{E}(\Gamma, t)$, i.e., $t{:}\varphi \in \Gamma$, and $\Gamma R \Delta$. Again, $!t{:}t{:}\varphi \in \Gamma$ hence, $t{:}\varphi \in \Delta$, i.e., $\varphi \in \mathcal{E}(\Delta, t)$.

4. Finally, the forcing relation is defined canonically, i.e., for each sentence variable S we stipulate $\Gamma \Vdash S$ iff $S \in \Gamma$.

Lemma 3. [Truth Lemma]

$$\Gamma \Vdash \varphi \quad \text{iff} \quad \varphi \in \Gamma \ .$$

Proof. By induction on φ. The base and Boolean cases are standard. Consider modalities $K_1, \ldots, K_n$.

If $K_i\varphi \in \Gamma$, and $\Gamma R_i \Delta$, then $\varphi \in \Delta$. By the induction hypothesis, $\Delta \Vdash \varphi$, hence $\Gamma \Vdash K_i\varphi$.

If $K_i\varphi \notin \Gamma$, then $\Gamma' = \Gamma^{\sharp_i} \cup \{\neg\varphi\}$ is consistent. Otherwise $\psi_1 \wedge \psi_2 \wedge \ldots \wedge \psi_l \rightarrow \varphi$ would be provable for some $K_i\psi_1, K_i\psi_2, \ldots, K_i\psi_l \in \Gamma$ and hence $K_i\psi_1 \wedge K_i\psi_2 \wedge \ldots \wedge K_i\psi_l \rightarrow K_i\varphi$ would also be provable, which would make Γ inconsistent. Let Δ be a maximal consistent set containing Γ'. Then $\Gamma R_i \Delta$, $\neg\varphi \in \Delta$, hence $\varphi \notin \Delta$ and, by the induction hypothesis, $\Delta \not\Vdash \varphi$, which yields $\Gamma \not\Vdash K_i\varphi$.

Now consider the last remaining case of Truth Lemma: $\varphi = t{:}\psi$. Let $t{:}\psi \in \Gamma$. Then, by the definition of the evidence function, $\psi \in \mathcal{E}(\Gamma, t)$. It remains to show that $\Delta \Vdash \psi$ for all Δ's such that $\Gamma R \Delta$. Take such a Δ. By *Monotonicity* of the evidence function, $t{:}\psi \in \Delta$. By reflexivity, $\psi \in \Delta$. By the induction hypothesis, $\Delta \Vdash \psi$. Conversely, if $\Gamma \Vdash t{:}\psi$, then $\psi \in \mathcal{E}(\Gamma, t)$ and $t{:}\psi \in \Gamma$ by the definition of the evidence function $\mathcal{E}$. □

It is easy to see now that $\mathcal{M} = (W, R_1, \ldots, R_n, R, \mathcal{E}, \Vdash)$ is an $\mathsf{S4}_n\mathsf{LP}$-model meeting constant specification $\mathcal{CS}$. Indeed, by the definition of a consistent set, $\mathcal{CS} \subseteq \Gamma$, for each $\Gamma \in W$. By Truth Lemma 3, $\Gamma \Vdash \mathcal{CS}$.

Let us finish the proof of Theorem 1. If φ is not provable in $\mathsf{S4}_n\mathsf{LP}$ meeting constant specification $\mathcal{CS}$, then $\mathcal{M}$ is a countermodel for φ: consider $\{\neg\varphi\}$, which is consistent, and hence contained in a maximal consistent set Γ. By Truth Lemma 3, $\Gamma \not\Vdash \varphi$. □

5 Compactness and Fully Explanatory property

The above models satisfy the following *compactness* property, first noticed for canonical models of LP in [Fitting, 2003; Fitting, 2005].

Proposition 4. [Compactness] *For a given constant specification $\mathcal{CS}$, a set of formulas U is $\mathcal{CS}$-$\mathsf{S4}_n\mathsf{LP}$-($\mathcal{CS}$-$\mathsf{T}_n\mathsf{LP}$-, $\mathcal{CS}$-$\mathsf{S5}_n\mathsf{LP}$-)satisfiable iff any finite subset of U is $\mathcal{CS}$-$\mathsf{S4}_n\mathsf{LP}$-($\mathcal{CS}$-$\mathsf{T}_n\mathsf{LP}$-, $\mathcal{CS}$-$\mathsf{S5}_n\mathsf{LP}$-)satisfiable.*

Proof. Suppose any finite subset of U is $\mathcal{CS}$-$\mathsf{S4}_n\mathsf{LP}$-satisfiable. We will find a world Γ in the canonical $\mathcal{CS}$-$\mathsf{S4}_n\mathsf{LP}$-model such that $\Gamma \Vdash U$. First, note that U is a consistent set. Otherwise, for some $X_1, \ldots, X_m \in U$, $\mathcal{CS}$-$\mathsf{S4}_n\mathsf{LP} \vdash \neg(X_1 \wedge \ldots \wedge X_m)$, which would make $\{X_1, \ldots, X_m\}$ a finite unsatisfiable subset of U, which is impossible. Extend U to a maximal consistent set Γ, which is hence a world in the canonical $\mathcal{CS}$-$\mathsf{S4}_n\mathsf{LP}$-model. Since $U \subseteq \Gamma$, by Truth Lemma 3, $\Gamma \Vdash U$. □

Fully Explanatory property of the canonical models for the logic of proofs was discovered in [Fitting, 2003; Fitting, 2005]. This property might be summarized as "whatever is known, is known for a reason."

Definition 3. An $\mathsf{S4}_n\mathsf{LP}$-($\mathsf{T}_n\mathsf{LP}$-, $\mathsf{S5}_n\mathsf{LP}$-)model is *Fully Explanatory* provided that, whenever $v \Vdash \varphi$ for every v such that uRv, then for some proof polynomial t we have $u \Vdash t{:}\varphi$.

Proposition 5. [Fully Explanatory property] *For any full constant specification $\mathcal{CS}$, the canonical $\mathcal{CS}$-$\mathsf{S4}_n\mathsf{LP}$-($\mathcal{CS}$-$\mathsf{T}_n\mathsf{LP}$-, $\mathcal{CS}$-$\mathsf{S5}_n\mathsf{LP}$-)model is Fully Explanatory.*

Proof. We prove the contrapositive. Let $t{:}\varphi \notin \Gamma$ for each proof polynomial t. Consider a set $U = \Gamma^\flat \cup \{\neg\varphi\}$. We claim that U is consistent. Otherwise, for some $t_1{:}\psi_1, t_2{:}\psi_2, \ldots, t_k{:}\psi_k \in \Gamma$, $\mathcal{CS}$-$\mathsf{S4}_n\mathsf{LP}$ proves $\psi_1 \rightarrow (\psi_2 \rightarrow (\psi_3 \rightarrow \ldots \rightarrow \varphi) \ldots)$. Since $\mathcal{CS}$ is a full constant specification, there is a proof polynomial s such that $\mathcal{CS}$-$\mathsf{S4}_n\mathsf{LP}$ proves $s{:}(\psi_1 \rightarrow (\psi_2 \rightarrow (\psi_3 \rightarrow \ldots \rightarrow \varphi) \ldots)$. Using E1, we establish that $\mathcal{CS}$-$\mathsf{S4}_n\mathsf{LP}$ proves

$$t_1{:}\psi_1 \rightarrow (t_2{:}\psi_2 \rightarrow (t_3{:}\psi_3 \rightarrow \ldots \rightarrow (st_1t_2 \ldots t_k){:}\varphi) \ldots).$$

Hence $(st_1t_2 \ldots t_k){:}\varphi \in \Gamma$, which contradicts the assumption about Γ.

Now take Δ to be a maximal consistent extension of U. It is clear that Δ is a world in a canonical model and that $\Gamma R \Delta$. By Truth Lemma 3, $\Delta \nVdash \varphi$. □

6 Forgetful evidence-based knowledge

In this section, we introduce a light version of evidence-based knowledge, which we call *forgetful evidence-based knowledge*, in the form of a new modal operator $J\varphi$ (read φ *is justified*) which is the forgetful projection of evidence assertions $t{:}\varphi$. In the spirit of this paper, we consider an axiomatic description first.

Definition 4. The language of forgetful evidence-based knowledge is a modal language with $n+1$ modalities $K_1, \ldots, K_n, J$. Systems T_n^J, $\mathsf{S4}_n^J$, and $\mathsf{S5}_n^J$ are specified as T_n, $\mathsf{S4}_n$, and $\mathsf{S5}_n$, with the modalities $K_1, \ldots, K_n$ augmented by $\mathsf{S4}$ with the modality J, together with the forgetful version of the *undeniability of evidence* principle

$$J\varphi \rightarrow K_i\varphi,$$

for all $i = 1, \ldots, n$.

Apparently, the dummy $(n+1)$st agent corresponding to J plays the role of a sceptical and not logically omniscient $\mathsf{S4}$-agent who accepts facts only if they are supplied with checkable evidence. On the other hand, this agent is trusted by all other agents and is capable of internalizing and inspecting any fact actually proven in the system.

Comment 1. $\mathsf{S4}_n^J$ corresponds to one of $\mathsf{S4}$-based systems with the "any fool knows" modality considered in [McCarthy *et al.*, 1979] (with different axiom system). In the latter paper, the modality "any fool knows" is assumed to be the same type as the modalities of real agents. In this paper, we consider evidence-based knowledge operators independently from the real agents knowledge operators.

Lemma 4. *In each of* T_n^J*,* $\mathsf{S4}_n^J$*, and* $\mathsf{S5}_n^J$*,*

$$K_iJ\varphi \leftrightarrow J\varphi \leftrightarrow JK_i\varphi \ .$$

Proof. Immediate from K_i-reflexivity and the following derivations.
$J\varphi \rightarrow JJ\varphi \rightarrow K_iJ\varphi$;
$J\varphi \rightarrow JJ\varphi \rightarrow JK_i\varphi$;
$K_iJ\varphi \rightarrow J\varphi$;
$K_i\varphi \rightarrow \varphi$, $J(K_i\varphi \rightarrow \varphi)$, $JK_i\varphi \rightarrow J\varphi$. □

Proposition 6. *Forgetful evidence-based knowledge J satisfies the Fixed-Point Axiom in each of* T_n^J*,* $\mathsf{S4}_n^J$*, and* $\mathsf{S5}_n^J$*.*

Proof. Deriving the Fixed-Point identity for J in T_n^J (hence in $\mathsf{S4}_n^J$ and $\mathsf{S5}_n^J$)

$$J\varphi \quad \leftrightarrow \quad E(\varphi \wedge J\varphi)$$

is similar to Proposition 3. □

Definition 5. T_n^J-models are Kripke models for $(n+1)$-agent modal logics with a *frame* $(W, R_1, \ldots, R_n, R)$, where W is a non-empty set of possible worlds, $R_1, \ldots, R_n$ are reflexive accessibility relations on W associated to operators $K_1, \ldots, K_n$ respectively, R is a reflexive transitive relation on W, and $R_i \subseteq R$ for all $i = 1, \ldots, n$. As usual, a forcing relation $\Vdash$ is an arbitrary mapping from propositional letters to subsets of W, which is extended from propositional letters to all formulas by the usual modal rules.
$\mathsf{S4}_n^J$-models are those where $R_1, \ldots, R_n$ are reflexive and transitive.
$\mathsf{S5}_n^J$-models are those with reflexive, transitive, and symmetric $R_1, \ldots, R_n$.

Proposition 7. T_n^J *(*$\mathsf{S4}_n^J$*,* $\mathsf{S5}_n^J$*) is sound with respect to* T_n^J*-models (*$\mathsf{S4}_n^J$*-models,* $\mathsf{S5}_n^J$*-models).*

Proof. The usual modal axioms are valid by our choice of accessibility relations. $J\varphi \rightarrow K_i\varphi$ is trivially guaranteed by $R_i \subseteq R$. Indeed, let $u \Vdash J\varphi$ and uR_iv. Then uRv also holds, which brings $u \Vdash \varphi$. Hence, $u \Vdash K_i\varphi$. □

Completeness also occurs. For T_n^J and $\mathsf{S4}_n^J$, this will follow from Theorem 3 below. The completeness of $\mathsf{S5}_n^J$ will be established in Theorem 6.

Definition 6. A *sequent* is a pair of finite sets of $\mathsf{S4}_n^J$-formulas presented as $\Gamma \Rightarrow \Delta$. To simplify proofs, we assume a Boolean basis $\rightarrow, \bot$ and treat the remaining Boolean connectives as definable ones.

Axioms of $\mathsf{S4}_n^J\mathsf{G}$ are the sequents $S \Rightarrow S$ and $\bot \Rightarrow$, where S is a propositional variable. The propositional rules of $\mathsf{S4}_n^J\mathsf{G}$ are those from the classical propositional Gentzen-style system, including Weakening and Cut (cf. [Troelstra and Schwichtenberg, 1996]). In addition, there are $n+1$ pairs of proper modal rules:

$$\frac{\varphi, \Gamma \Rightarrow \Delta}{\Box\varphi, \Gamma \Rightarrow \Delta}\ (\Box, \Rightarrow) \qquad \text{and} \qquad \frac{J\Gamma, \Box\Delta \Rightarrow \varphi}{J\Gamma, \Box\Delta \Rightarrow \Box\varphi}\ (\Rightarrow, \Box).$$

where $\Box \in \{K_1, \ldots, K_n, J\}$ and $\Box\{\phi_1, \ldots, \phi_m\} = \{\Box\phi_1, \ldots, \Box\phi_m\}$.

The Gentzen-style version $\mathsf{T}_n^J\mathsf{G}$ of T_n^J has the same rules as $\mathsf{S4}_n^J\mathsf{G}$ with the $(\Rightarrow, \Box)$ rule replaced by

$$\frac{J\Gamma, \Delta \Rightarrow \varphi}{J\Gamma, \Box\Delta \Rightarrow \Box\varphi}\ (\Rightarrow, \Box).$$

Theorem 2. [Equivalence of Gentzen- and Hilbert-style systems]

$\Gamma \Rightarrow \Delta$ *is provable in* $\mathsf{S4}_n^J\mathsf{G}$ *(*$\mathsf{T}_n^J\mathsf{G}$*) iff* $\bigwedge\Gamma \rightarrow \bigvee\Delta$ *is provable in* $\mathsf{S4}_n^J$ *(*T_n^J*).*

Proof. The part "only if," i.e., that $\mathsf{S4}_n^J\mathsf{G} \vdash \Gamma \Rightarrow \Delta$ yields $\mathsf{S4}_n^J \vdash \bigwedge\Gamma \rightarrow \bigvee\Delta$, is a standard exercise in modal derivation. Let us check the soundness of the $(\Rightarrow, \Box)$-rule in $\mathsf{S4}_n^J\mathsf{G}$. By the induction hypothesis,

$$\mathsf{S4}_n^J \vdash \bigwedge J\Gamma \wedge \bigwedge\Box\Delta \rightarrow \varphi \ .$$

By $\mathsf{S4}$-reasoning,

$$\mathsf{S4}_n^J \vdash \bigwedge J\Gamma \rightarrow (\bigwedge\Box\Delta \rightarrow \varphi) \ .$$

By Lemma 4,

$$\mathsf{S4}_n^J \vdash \bigwedge J\Gamma \rightarrow \Box(\bigwedge\Box\Delta \rightarrow \varphi) \ .$$

Use distribution to establish

$$\mathsf{S4}_n^J \vdash \bigwedge J\Gamma \rightarrow (\bigwedge\Box\Box\Delta \rightarrow \Box\varphi) \ .$$

By $\mathsf{S4}$-reasoning,

$$\mathsf{S4}_n^J \vdash \bigwedge J\Gamma \rightarrow (\bigwedge\Box\Delta \rightarrow \Box\varphi) \ ,$$

hence

$$\mathsf{S4}_n^J \vdash \bigwedge J\Gamma \wedge \bigwedge\Box\Delta \rightarrow \Box\varphi \ .$$

Let us now check the soundness of the $(\Rightarrow, \Box)$-rule in $\mathsf{T}_n^J\mathsf{G}$. By the induction hypothesis,

$$\mathsf{T}_n^J \vdash \bigwedge J\Gamma \wedge \bigwedge\Delta \rightarrow \varphi \ .$$

By T-reasoning,

$$\mathsf{T}_n^J \vdash \bigwedge\Box J\Gamma \rightarrow (\bigwedge\Box\Delta \rightarrow \Box\varphi) \ .$$

By Lemma 4,

$$\mathsf{T}_n^J \vdash \bigwedge J\Gamma \rightarrow (\bigwedge \Box\Delta \rightarrow \Box\varphi) \ .$$

The "if" direction for both $\mathsf{S4}_n^J$ and T_n^J will be established later in Corollary 1. □

Below we prove completeness, cut-elimination, and adequacy theorem for $\mathsf{S4}_n^J$ (T_n^J) and $\mathsf{S4}_n^J\mathsf{G}$ ($\mathsf{T}_n^J\mathsf{G}$).

Theorem 3. [Consolidated completeness theorem] *The following are equivalent:*

1. $\Gamma \Rightarrow \Delta$ *is provable in* $\mathsf{S4}_n^J\mathsf{G}$ *(*$\mathsf{T}_n^J\mathsf{G}$*) without cut;*
2. $\Gamma \Rightarrow \Delta$ *is provable in* $\mathsf{S4}_n^J\mathsf{G}$ *(*$\mathsf{T}_n^J\mathsf{G}$*);*
3. $\bigwedge\Gamma \rightarrow \bigvee\Delta$ *is provable in* $\mathsf{S4}_n^J$ *(*T_n^J*) ;*
4. $\bigwedge\Gamma \rightarrow \bigvee\Delta$ *is* $\mathsf{S4}_n^J$*-valid (*T_n^J*-valid);*
5. $\bigwedge\Gamma \rightarrow \bigvee\Delta$ *is valid in all finite* $\mathsf{S4}_n^J$*-models (*T_n^J*-models).*

Proof. We will prove the case of $\mathsf{S4}_n^J$ in detail. The case of T_n^J is treated similarly, and we will show what modifications should be made in the $\mathsf{S4}_n^J$ proof to make it work for T_n^J as well.

Steps (1) $\Longrightarrow$ (2) and (4) $\Longrightarrow$ (5) are trivial, (2) $\Longrightarrow$ (3) $\Longrightarrow$ (4) has already been covered above. We will concentrate on proving that (5) $\Longrightarrow$ (1). As usual for this sort of proof, we assume not (1) and establish not (5), i.e., given that $\Gamma_0 \Rightarrow \Delta_0$ is not provable in $\mathsf{S4}_n^J\mathsf{G}$ without cut, we build a finite $\mathsf{S4}_n^J$-model $\mathcal{M}$, such that at some node of $\mathcal{M}$, all formulas from Γ_0 hold and all formulas from Δ_0 do not hold.

To keep the domain of a model finite, we will consider only formulas from a given finite set $\mathcal{F}$ of formulas closed under subformulas and containing all formulas from the given sequent $\Gamma_0 \Rightarrow \Delta_0$.

We call a sequent $\Gamma \Rightarrow \Delta$ *consistent* if $\Gamma \Rightarrow \Delta$ is not provable in $\mathsf{S4}_n^J\mathsf{G}$ without cut. A sequent $\Gamma \Rightarrow \Delta$ is called *saturated* if the following conditions hold:

- $\bot \in \Delta$;
- $\varphi \rightarrow \psi \in \Gamma$ yields $\psi \in \Gamma$ or $\varphi \in \Delta$;
- $\varphi \rightarrow \psi \in \Delta$ yields $\varphi \in \Gamma$ and $\psi \in \Delta$;
- $\Box\varphi \in \Gamma$ yields $\varphi \in \Gamma$ where $\Box \in \{K_1, \ldots, K_n, J\}$.

It is easy to see that any consistent sequent $\Gamma \Rightarrow \Delta$ can be extended to a saturated consistent sequent by an obvious terminating saturation procedure. If the original sequent $\Gamma \Rightarrow \Delta$ contains only formulas from $\mathcal{F}$, its saturation consists of formulas from $\mathcal{F}$ too.

Define a model $\mathcal{M} = (W, R_1, \ldots, R_n, R, \Vdash)$. W will be the (finite) set of all consistent saturated sequents.

Let $\Gamma^\natural = \{J\varphi \mid J\varphi \in \Gamma\}$ and $\Gamma^{\natural_i} = \{K_i\varphi \mid K_i\varphi \in \Gamma\}$. Set

$$(\Gamma \Rightarrow \Delta)R(\Gamma' \Rightarrow \Delta') \text{ if } \Gamma^\natural \subseteq \Gamma',$$
$$(\Gamma \Rightarrow \Delta)R_i(\Gamma' \Rightarrow \Delta') \text{ if } \Gamma^\natural \cup \Gamma^{\natural_i} \subseteq \Gamma'.$$

From this definition, all $R_1, \ldots, R_n, R$ are reflexive and transitive, and $R_i \subseteq R$ for all $i = 1, \ldots n$.

For T_n^J we define

$$(\Gamma \Rightarrow \Delta) R_i (\Gamma' \Rightarrow \Delta') \text{ if } \Gamma^\natural \cup \Gamma^{\sharp_i} \subseteq \Gamma'.$$

Obviously, those R_i's are reflexive, but not necessarily transitive.

Finally,

$$(\Gamma \Rightarrow \Delta) \Vdash S \text{ iff } S \in \Gamma \text{ for a propositional letter } S.$$

Lemma 5. [Truth Lemma]

1. *If $\varphi \in \Gamma$, then $(\Gamma \Rightarrow \Delta) \Vdash \varphi$;*
2. *If $\varphi \in \Delta$, then $(\Gamma \Rightarrow \Delta) \not\Vdash \varphi$.*

Proof. It is established by a standard induction on φ. The base and the cases of Boolean connectives are trivial.

Suppose $\varphi = K_i\psi$. If $K_i\psi \in \Gamma$, and $\Gamma' \Rightarrow \Delta'$ is accessible from $\Gamma \Rightarrow \Delta$ by R_i, then $\Gamma^{\natural_i} \subseteq \Gamma'$, hence $K_i\psi \in \Gamma'$. By the corresponding saturation property, $\psi \in \Gamma'$. By the induction hypothesis, $(\Gamma' \Rightarrow \Delta') \Vdash \psi$, hence $(\Gamma \Rightarrow \Delta) \Vdash K_i\psi$.

Now let $K_i\psi \in \Delta$. Then $\Gamma^{\natural_i}, \Gamma^\natural \Rightarrow \psi$ is a consistent sequent, otherwise $\Gamma^{\natural_i}, \Gamma^\natural \Rightarrow \psi$ would be derivable in $\mathsf{S4}_n^J\mathsf{G}$ without cut. By the $(\Rightarrow, \Box)$-rule, $\Gamma^{\natural_i}, \Gamma^\natural \Rightarrow K_i\psi$ would also be derivable in $\mathsf{S4}_n^J\mathsf{G}$ without cut. Hence, by Weakening, $\Gamma \Rightarrow \Delta$ is derivable in $\mathsf{S4}_n^J\mathsf{G}$ without cut, which contradicts our assumption of the consistency of $\Gamma \Rightarrow \Delta$. Consider a saturated extension $\Gamma' \Rightarrow \Delta'$ of $\Gamma^{\natural_i}, \Gamma^\natural \Rightarrow \psi$. Since $\psi \in \Delta'$, by the induction hypothesis, $(\Gamma' \Rightarrow \Delta') \not\Vdash \psi$. Obviously, $(\Gamma' \Rightarrow \Delta')$ is accessible from $(\Gamma \Rightarrow \Delta)$ by R_i, hence $(\Gamma \Rightarrow \Delta) \not\Vdash K_i\psi$. For T_n^J it suffices to take a consistent sequent $\Gamma^{\sharp_i}, \Gamma^\natural \Rightarrow \psi$ instead of $\Gamma^{\natural_i}, \Gamma^\natural \Rightarrow \psi$.

Suppose $\varphi = J\psi$. If $J\psi \in \Gamma$, and $\Gamma' \Rightarrow \Delta'$ is accessible from $\Gamma \Rightarrow \Delta$ by R, then $\Gamma^\natural \subseteq \Gamma'$, hence $J\psi \in \Gamma'$. By the corresponding saturation property, $\psi \in \Gamma'$. By the induction hypothesis, $(\Gamma' \Rightarrow \Delta') \Vdash \psi$, hence $(\Gamma \Rightarrow \Delta) \Vdash J\psi$.

Let $J\psi \in \Delta$. Then $\Gamma^\natural \Rightarrow \psi$ is a consistent sequent, since otherwise $\Gamma^\natural \Rightarrow \psi$ would be derivable in $\mathsf{S4}_n^J\mathsf{G}$ without cut. By the $(\Rightarrow, \Box)$-rule, $\Gamma^\natural \Rightarrow J\psi$ would also be derivable in $\mathsf{S4}_n^J\mathsf{G}$ without cut, hence $\Gamma \Rightarrow \Delta$ would be inconsistent. Consider a saturated extension $\Gamma' \Rightarrow \Delta'$ of $\Gamma^\natural \Rightarrow \psi$. Since $\psi \in \Delta'$, by the induction hypothesis, $(\Gamma' \Rightarrow \Delta') \not\Vdash \psi$. Since $(\Gamma' \Rightarrow \Delta')$ is accessible from $(\Gamma \Rightarrow \Delta)$ by R, $(\Gamma \Rightarrow \Delta) \not\Vdash J\psi$. □

Here is the standard conclusion of the proof of Theorem 3. Let $\Gamma \Rightarrow \Delta$ be a sequent not provable in $\mathsf{S4}_n^J\mathsf{G}$ without cut, hence consistent. Consider its saturated consistent extension $(\Phi \Rightarrow \Psi)$, which is an element of W. Since $\Gamma \subseteq \Phi$ and $\Delta \subseteq \Psi$, by Lemma 5, all formulas from Γ hold at $(\Phi \Rightarrow \Psi)$ and all formulas from Δ do not hold at $(\Phi \Rightarrow \Psi)$. Hence $(\Phi \Rightarrow \Psi) \not\Vdash \bigwedge\Gamma \to \bigvee\Delta$. □

Corollary 1.

1. *Cut-elimination theorem in $\mathsf{S4}_n^J\mathsf{G}$ and $\mathsf{T}_n^J\mathsf{G}$.*
2. *Completeness of $\mathsf{S4}_n^J$ with respect to $\mathsf{S4}_n^J$-models and T_n^J with respect to T_n^J-models.*
3. *Finite model property of $\mathsf{S4}_n^J$ and T_n^J.*
4. *Decidability of $\mathsf{S4}_n^J$ and T_n^J.*
5. *Equivalence of $\mathsf{S4}_n^J$ to $\mathsf{S4}_n^J\mathsf{G}$ and T_n^J to $\mathsf{T}_n^J\mathsf{G}$ (Theorem 2).*

Now we are ready to show that T_n^J and $\mathsf{S4}_n^J$ are exactly the forgetful projections of $\mathsf{T}_n\mathsf{LP}$ and $\mathsf{S4}_n\mathsf{LP}$ respectively, defined by a translation $(\ )^\circ$ which maps $t{:}\varphi$ to $J\varphi$ and commutes with all other connectives.

Theorem 4.

$$(\mathsf{T}_n\mathsf{LP})^o \subseteq \mathsf{T}_n^J \quad \textit{and} \quad (\mathsf{S4}_n\mathsf{LP})^o \subseteq \mathsf{S4}_n^J.$$

Proof. A straightforward induction on derivations in $\mathsf{T}_n\mathsf{LP}$ and $\mathsf{S4}_n\mathsf{LP}$. It suffices to observe that the forgetful translations of all axioms and rules of $\mathsf{T}_n\mathsf{LP}$ and $\mathsf{S4}_n\mathsf{LP}$ are T_n^J- and $\mathsf{S4}_n^J$-compliant, respectively. □

The converse claim that $\mathsf{T}_n^J \subseteq (\mathsf{T}_n\mathsf{LP})^o$ and $\mathsf{S4}_n^J \subseteq (\mathsf{S4}_n\mathsf{LP})^o$, is a much trickier *Realization Theorem.*

Theorem 5. [Realization Theorem] *There is an algorithm that given a T_n^J-derivation ($\mathsf{S4}_n^J$-derivation) of a formula φ, retrieves a $\mathsf{T}_n\mathsf{LP}$-derivation ($\mathsf{S4}_n\mathsf{LP}$-derivation) of a formula ψ such that $(\psi)^o = \varphi$.*

Proof. First, find a cut-free proof of a given formula in $\mathsf{S4}_n^J$ ($\mathsf{T}_n^J\mathsf{G}$). Then run the realizability algorithm from [Artemov, 2001], Theorem 9.4, to retrieve evidence terms at every occurrence of the modality J in this derivation. Here is a brief exposition of how the realization algorithm works. We consider $\mathsf{S4}_n^J$ only; the case of T_n^J is quite similar.

We call a realization r of modality J in a given formula or sequent *normal* if all negative occurrences of J are realized by proof variables.

We will speak about a sequent's $\Gamma \Rightarrow \Delta$ being derivable in $\mathsf{S4}_n^J$ meaning $\mathsf{S4}_n^J \vdash \bigwedge\Gamma \rightarrow \bigvee\Delta$, or, equivalently, $\mathsf{S4}_n^J\mathsf{G} \vdash \Gamma \Rightarrow \Delta$. Moreover, since $\mathsf{S4}_n^J$ enjoys the deduction theorem, $\mathsf{S4}_n^J$ derives $\Gamma \Rightarrow \varphi$ iff $\Gamma \vdash \varphi$ in $\mathsf{S4}_n^J$ iff $\mathsf{S4}_n^J\mathsf{G} \vdash \Gamma \Rightarrow \varphi$.

Consider a cut-free derivation $\mathcal{T}$ of a sequent $\Rightarrow \varphi$ in $\mathsf{S4}_n^J\mathsf{G}$. It suffices now to construct a normal realization r such that $\mathsf{S4}_n\mathsf{LP} \vdash \bigwedge\Gamma^r \rightarrow \bigvee\Delta^r$ for any sequent $\Gamma \Rightarrow \Delta$ in $\mathcal{T}$. Note that in $\mathcal{T}$, the rules respect polarities; all occurrences of J introduced by $(\Rightarrow, \Box)$ are positive, and all negative occurrences are introduced by $(\Box, \Rightarrow)$ or by Weakening. Occurrences of J are *related* if they occur in related formulas of premises and conclusions of rules; we extend this relationship by transitivity. All occurrences of J in $\mathcal{T}$ are naturally split into disjoint *families* of related ones. We call a family *essential* if it contains at least one instance of the $(\Rightarrow, J)$ rule where the modality J of this family has been introduced.

The desired r will be constructed by steps $1-3$ described below. We reserve a sufficiently large set of proof variables as *provisional variables.*

Step 1. For every negative family and nonessential positive family, we replace all occurrences of $J\phi$ by "$x{:}\phi$" for a fresh proof variable x.

Step 2. Pick an essential family f, enumerate all the occurrences of rules $(\Rightarrow, J)$ which introduce the modality J of this family. Let n_f be the total number of such rules for the family f. Replace all boxes of the family f by the polynomial

$$v_1 + \ldots + v_{n_f},$$

where v_i's are fresh provisional variables. The resulting tree $\mathcal{T}'$ is labelled by $\mathsf{S4}_n\mathsf{LP}$-formulas, since all occurrences of the kind $J\phi$ in $\mathcal{T}$ are replaced by $t{:}\phi$ for corresponding proof polynomials t.

Step 3. Replace the provisional variables by proof polynomials as follows. Proceed from the leaves of the tree to its root. By induction on the depth of a node in $\mathcal{T}'$ we establish that after the process passes a node, the sequent assigned to this node becomes derivable in $\mathsf{S4}_n\mathsf{LP}$. The axioms $S \Rightarrow S$ and $\bot \Rightarrow$ are derivable in $\mathsf{S4}_n\mathsf{LP}$. For every rule other than $(\Rightarrow, J)$, we do not change the realization of formulas and just establish that the concluding sequent is provable in $\mathsf{S4}_n\mathsf{LP}$, given that the premises are. It is clear that every move down in the tree $\mathcal{T}'$ other than $(\Rightarrow, J)$ is derivable in $\mathsf{S4}_n\mathsf{LP}$.

Let an occurrence of the rule $(\Rightarrow, J)$ have number i in the numbering of all rules $(\Rightarrow, J)$ from a given family f. The corresponding node in $\mathcal{T}'$ is labelled by

$$\frac{y_1{:}B_1, \ldots, y_k{:}B_k \Rightarrow B}{y_1{:}B_1, \ldots, y_k{:}B_k \Rightarrow (u_1 + \ldots + u_{n_f}){:}B}\ ,$$

where $y_1, \ldots, y_k$ are proof variables, $u_1, \ldots, u_{n_f}$ are proof polynomials, and u_i is a provisional variable. By the induction hypothesis, the premise sequent $y_1{:}B_1, \ldots, y_k{:}B_k \Rightarrow B$ is derivable in $\mathsf{S4}_n\mathsf{LP}$. By the Lifting Lemma (Proposition 2), construct a proof polynomial $t(y_1, \ldots, y_k)$ such that

$$\mathsf{S4}_n\mathsf{LP} \vdash y_1{:}B_1, \ldots, y_k{:}B_k \Rightarrow t(y_1, \ldots, y_k){:}B.$$

Since

$$\mathsf{S4}_n\mathsf{LP} \vdash t{:}B \rightarrow (u_1 + \ldots + u_{i-1} + t + u_{i+1} + \ldots + u_{n_f}){:}B$$

we have

$$\mathsf{S4}_n\mathsf{LP} \vdash y_1{:}B_1, \ldots, y_k{:}B_k \Rightarrow (u_1 + \ldots + u_{i-1} + t + u_{i+1} + \ldots + u_{n_f}){:}B.$$

Now substitute $t(y_1, \ldots, y_k)$ for u_i everywhere in $\mathcal{T}'$ (and the corresponding constant specification $\mathcal{CS}$).

Note that $t(y_1, \ldots, y_k)$ has no provisional variables, and that there is one less provisional variable (namely u_i) in $\mathcal{T}'$. The conclusion of the given rule $(\Rightarrow, J)$ becomes derivable in $\mathsf{S4}_n\mathsf{LP}$, and the induction step is complete.

Eventually, we substitute polynomials of non-provisional variables for all provisional variables in $\mathcal{T}'$ and establish that the root sequent of $\mathcal{T}'$ is derivable in $\mathsf{S4}_n\mathsf{LP}$. The realization r built by this procedure is normal. $\square$

Note that the current version of the realization algorithm can produce proof polynomials which are exponential in the size of the original cut-free derivation in $\mathsf{S4}^J_n\mathsf{G}$. A more efficient realization algorithm has been described in [Brezhnev and Kuznets, 2005], where the realizing proof polynomials are quadratic in the size of the original cut-free derivation in $\mathsf{S4}^J_n\mathsf{G}$.

The case of $\mathsf{S5}^J_n$ needs a separate treatment.

Definition 7. $\mathsf{S5}^J_n$-models are the models from Definition 5 with reflexive, transitive, and symmetric relations $R_1, \ldots, R_n$.

Theorem 6. $\mathsf{S5}^J_n$ *is sound and complete with respect to* $\mathsf{S5}^J_n$*-models.*

Proof. The soundness part is straightforward. In particular, $J\varphi \to K_i\varphi$ is trivially guaranteed by $R_i \subseteq R$.

The completeness part is done by the standard maximal consistent set construction. A set Γ is consistent if for any finite $\Delta \subseteq \Gamma$, $\mathsf{S5}_n^J \nvdash \neg(\bigwedge\Delta)$. W is a collection of all maximal consistent sets, $\Gamma R_i \Delta$ iff $\Gamma^{\sharp_i} \subseteq \Delta$, $\Gamma R \Delta$ iff $\Gamma^\sharp \subseteq \Delta$, where $\Gamma^\sharp = \{\varphi \mid J\varphi \in \Gamma\}$. All R_i and R are reflexive and transitive. Let us check the inclusions $R_i \subseteq R$. Suppose $\Gamma R_i \Delta$ and $J\varphi \in \Gamma$. Since $\mathsf{S5}_n^J \vdash J\varphi \to K_i J\varphi$, $J\varphi \to K_i J\varphi \in \Gamma$ and $K_i J\varphi \in \Gamma$, hence $J\varphi \in \Delta$ and $\varphi \in \Delta$. Therefore $\Gamma R \Delta$.

Moreover, each of R_i, $i = 1, \ldots, n$ is symmetric (hence each is an equivalence relation). Indeed, let $\Gamma R_i \Delta$ and $K_i\varphi \in \Delta$. It suffices to show that $K_i\varphi \in \Gamma$ (hence $\varphi \in \Gamma$). Suppose $K_i\varphi \notin \Gamma$. Then $\neg K_i\varphi \in \Gamma$. By $\mathsf{S5}$-axiom $\neg K_i\varphi \to K_i\neg K_i\varphi$, $K_i\neg K_i\varphi \in \Gamma$. Since $\Gamma^{\sharp_i} \subseteq \Delta$, $\neg K_i\varphi \in \Delta$ as well, a contradiction.

As usual, $\Gamma \Vdash S$ iff $S \in \Gamma$ for any sentence variable S. We have shown that the resulting construction $(W, R_1, \ldots, R_n, R, \Vdash)$ is an $\mathsf{S5}_n^J$-model.

The Truth Lemma says that for any formula φ

$$\Gamma \Vdash \varphi \quad \text{iff} \quad \varphi \in \Gamma.$$

The proof follows from a standard induction on φ. Let us check the case when $\varphi = J\psi$. If $J\psi \in \Gamma$, then $\psi \in \Delta$ for all Δ such that $\Gamma R \Delta$. By the induction hypothesis, $\Delta \Vdash \psi$ for all Δ such that $\Gamma R \Delta$. Hence $\Gamma \Vdash J\psi$. If $J\psi \notin \Gamma$, then $\Gamma^\sharp \cup \{\neg\psi\}$ is a consistent set. Otherwise, for some finite subset Θ of Γ, $\Theta^\sharp \vdash \psi$ and, by modal logic rules, $\Theta \vdash J\psi$, hence $J\psi \in \Gamma$, a contradiction. Take a maximal consistent set Δ containing $\Gamma^\sharp \cup \{\neg\psi\}$. Apparently, $\psi \notin \Delta$, hence by the induction hypothesis, $\Delta \nVdash \psi$ and $\Gamma \nVdash J\psi$.

Theorem 6 now follows immediately. □

Theorem 7. $\mathsf{S5}_n^J$ *is the forgetful projection of* $\mathsf{S5}_n\mathsf{LP}$*, i.e.,* $(\mathsf{S5}_n\mathsf{LP})^o = \mathsf{S5}_n^J$.

Proof. Again, the proof of $(\mathsf{S5}_n\mathsf{LP})^o \subseteq \mathsf{S5}_n^J$ is given by a straightforward induction on derivations in $\mathsf{S5}_n\mathsf{LP}$.

The existence of an $\mathsf{S5}_n\mathsf{LP}$-realization of any theorems of $\mathsf{S5}_n^J$ can be established semantically by methods developed in [Fitting, 2005]. The main ingredients of Fitting's semantical realizability proof are the Fully Explanatory property of $\mathsf{S5}_n\mathsf{LP}$-models with full constant specifications (Proposition 5) and the Compactness property (Proposition 4).

Definition 8. By $\mathsf{S5}_n\mathsf{LP}^-$ we mean a system $\mathsf{S5}_n\mathsf{LP}$ in a language without '+' and without axioms E3. Models of $\mathsf{S5}_n\mathsf{LP}^-$ are the same as for $\mathsf{S5}_n\mathsf{LP}$ except that the evidence function is not required to satisfy the *Sum* condition. We may assume that $\mathsf{S5}_n\mathsf{LP}$-models and $\mathsf{S5}_n\mathsf{LP}^-$-models are also models for $\mathsf{S5}_n^J$ with R being an accessibility relation for the modality J.

Note that such features as internalization and the Fully Explanatory property of the canonical model hold for $\mathsf{S5}_n\mathsf{LP}^-$ and $\mathsf{S5}_n\mathsf{LP}^-$-models as well.

Assume an $\mathsf{S5}_n^J$-formula φ is fixed for the rest of the proof of Theorem 7. By "subformula of φ" we will mean an "occurrence of a subformula of φ."

Definition 9. Let A be any assignment of proof variables to subformulas of φ of the form $J\phi$ that are in a negative position. We define two mappings w_A and v_A of subformulas of φ to sets of formulas of $\mathsf{S5}_n\mathsf{LP}$ and $\mathsf{S5}_n\mathsf{LP}^-$, respectively.

1. If P is an atomic formula (including $\bot$), then $w_A(P) = v_A(P) = \{P\}$.
2. $w_A(X \to Y) = \{X' \to Y' \mid X' \in w_A(X) \text{ and } Y' \in w_A(Y)\}$.
   $v_A(X \to Y) = \{X' \to Y' \mid X' \in v_A(X) \text{ and } Y' \in v_A(Y)\}$.
3. If K_iX is a negative subformula of φ, then
   $w_A(K_iX) = \{K_iX' \mid X' \in w_A(X)\}$,
   $v_A(K_iX) = \{K_iX' \mid X' \in v_A(X)\}$.
4. If K_iX is a positive subformula of φ, then
   $w_A(K_iX) = \{K_iX' \mid X' \in w_A(X)\}$,
   $v_A(K_iX) = \{K_i(X_1 \vee \ldots \vee X_k) \mid X_1, \ldots, X_k \in v_A(X)\}$.
5. If JX is a negative subformula of φ, then
   $w_A(JX) = \{x{:}X' \mid A(JX) = x \text{ and } X' \in w_A(X)\}$,
   $v_A(JX) = \{x{:}X' \mid A(JX) = x \text{ and } X' \in v_A(X)\}$.
6. If JX is a positive subformula of φ, then
   $w_A(JX) = \{t{:}X' \mid X' \in w_A(X) \text{ and } t \text{ is any proof polynomial}\}$,
   $v_A(JX) = \{t{:}(X_1 \vee \ldots \vee X_k) \mid X_1, \ldots, X_k \in v_A(X) \text{ and } t \text{ is any proof polynomial}\}$.

By $\neg v_A(X)$ we mean $\{\neg X' \mid X' \in v_A(X)\}$ (which has nothing to do with $v_A(\neg X)$).

Lemma 6. *Let $\mathcal{CS}$ be a full constant specification of $\mathsf{S5}_n\mathsf{LP}^-$ and $\mathcal{M}$ be a canonical model for $\mathsf{S5}_n\mathsf{LP}^-$ that meets $\mathcal{CS}$. Then for each world Γ of the model:*

1. *If ψ is a positive subformula of φ then $\Gamma \Vdash \neg v_A(\psi)$ yields $\Gamma \Vdash \neg\psi$.*
2. *If ψ is a negative subformula of φ then $\Gamma \Vdash v_A(\psi)$ yields $\Gamma \Vdash \psi$.*

Proof. Induction on ψ. The atomic case as well as the cases of Boolean connectives are straightforward (cf. Proposition 7.7 in [Fitting, 2005]).

Suppose ψ is K_iX, ψ is a positive subformula of φ, $\Gamma \Vdash \neg v_A(K_iX)$, and the result is known for X (which also occurs positively in φ). We show that $\Gamma^{\sharp_i} \cup \neg v_A(X)$ is consistent. Indeed, otherwise in $\Gamma^{\sharp_i} \vdash X_1 \vee \ldots \vee X_k$ for some $X_1, \ldots, X_k \in v_A(X)$. By the K_i-necessitation rule, $\Gamma \vdash K_i(X_1 \vee \ldots \vee X_k)$. Hence $\Gamma \Vdash K_i(X_1 \vee \ldots \vee X_k)$, which is impossible since $K_i(X_1 \vee \ldots \vee X_k) \in v_A(K_iX)$. Now, extend $\Gamma^{\sharp_i} \cup \neg v_A(X)$ to a maximal consistent Δ, which is therefore a world in $\mathcal{M}$ accessible from Γ by R_i. Since $\neg v_A(X) \subseteq \Delta$, $\Delta \Vdash \neg v_A(X)$. By the induction hypothesis, $\Delta \Vdash \neg X$. Therefore, $\Gamma \Vdash \neg K_iX$.

Suppose ψ is K_iX, ψ is a negative subformula of φ, $\Gamma \Vdash v_A(K_iX)$, and the result is known for X (which also occurs negatively in φ). In particular, $\Gamma \Vdash K_iX'$, for each $X' \in v_A(X)$. Let Δ be an arbitrary world such that $\Gamma R_i \Delta$. Then $\Delta \Vdash X'$, hence $\Delta \Vdash v_A(X)$. By the induction hypothesis, $\Delta \Vdash X$. Therefore $\Gamma \Vdash K_iX$.

Suppose ψ is JX, ψ is a positive subformula of φ, $\Gamma \Vdash \neg v_A(JX)$, and the result is known for X (which also occurs positively in φ). We show that $\Gamma^{\flat} \cup \neg v_A(X)$ is consistent. Indeed, otherwise, by compactness, $\{Y_1, \ldots, Y_m, \neg X_1, \ldots, \neg X_k\}$ is inconsistent for some $Y_1, \ldots, Y_m \in \Gamma^{\flat}$ and $X_1, \ldots, X_k \in v_A(X)$. This means that

$$\mathsf{S5}_n\mathsf{LP}^- \vdash Y_1 \to (Y_2 \to \ldots \to (Y_m \to X_1 \vee \ldots \vee X_k) \ldots) \ .$$

By internalization, there is a proof polynomial s such that

$$\mathsf{S5}_n\mathsf{LP}^- \vdash s{:}[Y_1 \to (Y_2 \to \ldots \to (Y_m \to X_1 \vee \ldots \vee X_k) \ldots)] \ .$$

Consider proof polynomials $t_1, t_2, \ldots, t_m$ such that $t_1{:}Y_1, t_2{:}Y_2, \ldots, t_m{:}Y_m \in \Gamma$. By E1 and propositional reasoning,

$$\mathsf{S5}_n\mathsf{LP}^- \vdash t_1{:}Y_1 \wedge t_2{:}Y_2 \wedge \ldots \wedge t_m{:}Y_m \to (st_1t_2\ldots t_m){:}[X_1 \vee \ldots \vee X_k] \ .$$

Therefore

$$\Gamma \Vdash (st_1t_2\ldots t_m){:}[X_1 \vee \ldots \vee X_k],$$

which is impossible since $(st_1t_2\ldots t_m){:}[X_1 \vee \ldots \vee X_k] \in v_A(JX)$.

Let Δ be a maximal consistent extension of $\Gamma^\flat \cup \neg v_A(X)$. Obviously, $\Gamma R \Delta$ and $\Delta \Vdash \neg v_A(X)$. By the induction hypothesis, $\Delta \Vdash \neg X$, hence $\Gamma \Vdash \neg JX$.

Suppose ψ is JX, ψ is a negative subformula of φ, $\Gamma \Vdash v_A(JX)$, and the result is known for X (which also occurs negatively in φ). Let X' be an arbitrary element of $v_A(X)$. Then $\Gamma \Vdash x{:}X'$, where x is a proof variable assigned to this occurrence JX by the mapping A. For any world Δ such that $\Gamma R \Delta$, $\Delta \Vdash X'$. By the induction hypothesis, $\Delta \Vdash X$. Therefore $\Gamma \Vdash JX$. □

Now suppose $\mathsf{S5}_n^J \vdash \varphi$ but $\mathsf{S5}_n\mathsf{LP}^- \nvdash (\varphi_1 \vee \ldots \vee \varphi_m)$ for all $\varphi_1, \ldots \varphi_m \in v_A(\varphi)$ with a given full constant specification $\mathcal{CS}$. Then every finite subset of $\neg v_A(\varphi)$ is satisfiable. By compactness (Proposition 4) adapted to $\mathsf{S5}_n\mathsf{LP}^-$, there is a world Γ in the canonical model for $\mathsf{S5}_n\mathsf{LP}^-$ with $\mathcal{CS}$ such that $\Gamma \Vdash \neg v_A(\varphi)$. By Lemma 6, $\Gamma \Vdash \neg\varphi$. Therefore, since $\mathsf{S5}_n^J \vdash \varphi$, there are $\varphi_1, \ldots \varphi_m \in v_A(\varphi)$ such that $\mathsf{S5}_n\mathsf{LP}^- \vdash (\varphi_1 \vee \ldots \vee \varphi_m)$.

Lemma 7. *For every subformula ψ of φ and each $\psi_1, \ldots \psi_m \in v_A(\psi)$, there is a substitution σ of proof polynomials for proof variables and a formula $\psi' \in w_A(\psi)$ such that:*

1. *If ψ is a positive subformula of φ, $\mathsf{S5}_n\mathsf{LP} \vdash (\psi_1 \vee \ldots \vee \psi_m)\sigma \to \psi'$.*
2. *If ψ is a negative subformula of φ, $\mathsf{S5}_n\mathsf{LP} \vdash \psi' \to (\psi_1 \wedge \ldots \wedge \psi_m)\sigma$.*

Proof. We use the fact that proof variables assigned to different (occurrences of) subformulas $J\psi$ in φ are all different.

Induction on ψ. Again, the atomic case as well as the cases of Boolean connectives are straightforward (cf. Proposition 7.8 in [Fitting, 2005]).

Suppose ψ is K_iX, ψ is a positive subformula of φ, and the result is known for X (which also occurs positively in φ). Let $K_iD_1, \ldots, K_iD_m \in v_A(K_iX)$. Those $D_1, \ldots, D_m$ are disjunctions of formulas from $v_A(X)$. By the induction hypothesis, there is a substitution σ and $X' \in w_A(X)$ such that $\mathsf{S5}_n\mathsf{LP} \vdash (D_1 \vee \ldots \vee D_m)\sigma \to X'$. Consequently, for each $j = 1, \ldots, m$, $\mathsf{S5}_n\mathsf{LP} \vdash D_j\sigma \to X'$. By necessitation, $\mathsf{S5}_n\mathsf{LP} \vdash K_i(D_j\sigma \to X')$, hence $\mathsf{S5}_n\mathsf{LP} \vdash K_iD_j\sigma \to K_iX'$. Therefore,

$$\mathsf{S5}_n\mathsf{LP} \vdash (K_iD_1 \vee \ldots \vee K_iD_m)\sigma \to K_iX'.$$

Suppose ψ is K_iX, ψ is a negative subformula of φ, and the result is known for X (which also occurs negatively in φ). Let $K_iX_1, \ldots, K_iX_m \in v_A(K_iX)$. By the induction hypothesis, there is a substitution σ and $X' \in w_A(X)$ such that $\mathsf{S5}_n\mathsf{LP} \vdash X' \to (X_1 \wedge \ldots \wedge X_m)\sigma$. By necessitation, $\mathsf{S5}_n\mathsf{LP} \vdash K_iX' \to K_i(X_1 \wedge \ldots \wedge X_m)\sigma$. Since K_i commutes with σ and $\wedge$, $\mathsf{S5}_n\mathsf{LP} \vdash K_iX' \to (K_iX_1 \wedge \ldots \wedge K_iX_m)\sigma$.

Suppose ψ is JX, ψ is a positive subformula of φ, and the result is known for X (which also occurs positively in φ). In this case $\psi_1, \ldots \psi_m \in v_A(\psi)$ are of the form $t_1{:}D_1, \ldots, t_m{:}$

D_m, where each of $D_1, \ldots, D_m$ is a disjunction of formulas from $v_A(X)$. By the induction hypothesis, there is a substitution σ and $X' \in w_A(X)$ such that

$$\mathsf{S5}_n\mathsf{LP} \vdash (D_1 \vee \ldots \vee D_m)\sigma \rightarrow X'.$$

Consequently, for each $j = 1, \ldots, m$, $\mathsf{S5}_n\mathsf{LP} \vdash D_j\sigma \rightarrow X'$. By internalization, there is a proof polynomial s_j such that $\mathsf{S5}_n\mathsf{LP} \vdash s_j{:}(D_j\sigma \rightarrow X')$. Then $\mathsf{S5}_n\mathsf{LP} \vdash (t_j{:}D_j)\sigma \rightarrow (s_j \cdot t_j\sigma){:}X'$. Set $t = (s_1 \cdot t_1\sigma) + \ldots + (s_m \cdot t_m\sigma)$. We have $\mathsf{S5}_n\mathsf{LP} \vdash (t_j{:}D_j)\sigma \rightarrow t{:}X'$, and hence

$$\mathsf{S5}_n\mathsf{LP} \vdash (t_1{:}D_1 \vee \ldots \vee t_m{:}D_m)\sigma \rightarrow t{:}X'.$$

Suppose ψ is JX, ψ is a negative subformula of φ, and the result is known for X (which also occurs negatively in φ). In this case $\psi_1, \ldots \psi_m \in v_A(\psi)$ are of the form $x{:}X_1, \ldots, x{:}X_m$, where each of $X_1, \ldots, X_m$ is from $v_A(X)$. By the induction hypothesis, there is a substitution σ and $X' \in w_A(X)$ such that $\mathsf{S5}_n\mathsf{LP} \vdash X' \rightarrow (X_1 \wedge \ldots \wedge X_m)\sigma$. Since the variable x is not assigned by A to any of subformulas of X, we may assume that x is not in the domain of σ. From the above, it follows that $\mathsf{S5}_n\mathsf{LP} \vdash X' \rightarrow X_j\sigma$. By internalization, $\mathsf{S5}_n\mathsf{LP} \vdash t_j{:}(X' \rightarrow X_j\sigma)$ for some proof polynomial t_j. Therefore, $\mathsf{S5}_n\mathsf{LP} \vdash s{:}(X' \rightarrow X_j\sigma)$ for $s = t_1 + \ldots + t_m$. Furthermore, $\mathsf{S5}_n\mathsf{LP} \vdash x{:}X' \rightarrow (s \cdot x){:}(X_j\sigma)$ for each $j = 1, \ldots, m$. Consider a new substitution $\sigma' = \sigma \cup \{x/(s \cdot x)\}$. Obviously, $\mathsf{S5}_n\mathsf{LP} \vdash x{:}X' \rightarrow (x{:}X_1 \wedge \ldots \wedge x{:}X_m)\sigma'$, which completes the proof of Lemma 7. □

To conclude the proof of Theorem 7, assume that $\mathsf{S5}^J_n \vdash \varphi$. Then there are $\varphi_1, \ldots \varphi_m \in v_A(\varphi)$ such that $\mathsf{S5}_n\mathsf{LP}^- \vdash \varphi_1 \vee \ldots \vee \varphi_m$. By Lemma 7, there is a substitution σ and $\varphi' \in w_A(\varphi)$ such that $\mathsf{S5}_n\mathsf{LP} \vdash (\varphi_1 \vee \ldots \vee \varphi_m)\sigma \rightarrow \varphi'$. Since $\mathsf{S5}_n\mathsf{LP}$ is closed under substitution, $\mathsf{S5}_n\mathsf{LP} \vdash \varphi'$. □

Theorem 7 yields an algorithm that given $\mathsf{S5}^J_n$-theorem φ, retrieves a $\mathsf{S5}_n\mathsf{LP}$-theorem ψ such that $(\psi)^o = \varphi$. Indeed, arrange an enumeration of all $\mathsf{S5}_n\mathsf{LP}$-realizations of φ and their proof searches in $\mathsf{S5}_n\mathsf{LP}$. By Theorem 7, this process should terminate with success. A question of finding an efficient realization algorithm for $\mathsf{S5}^J_n$ remains open.

The results of this section show that J may be regarded as the *forgetful version of evidence-based knowledge*. Using forgetful *EBK*-systems instead of the original *EBK*-systems makes sense, since the former are conventional multi-modal logics which are easier to work with. On the other hand, *EBK*-systems have a solid justification, which can be extended to the corresponding forgetful *EBK*-systems. In particular, this provides an *EBK*-semantics for the dummy "any fool" agent for $\mathsf{S4}^J_n$ from [McCarthy *et al.*, 1979].

Note that models for $\mathsf{T}_n\mathsf{LP}$, $\mathsf{S4}_n\mathsf{LP}$, and $\mathsf{S5}_n\mathsf{LP}$ are also models for T^J_n, $\mathsf{S4}^J_n$, and $\mathsf{S5}^J_n$ respectively, as well. It suffices to regard the evidence accessibility relation R in models for $\mathsf{T}_n\mathsf{LP}$, $\mathsf{S4}_n\mathsf{LP}$, and $\mathsf{S5}_n\mathsf{LP}$ as the accessibility relation for J.

7 Evidence-Based Knowledge vs. Common Knowledge

In this section, we compare evidence-based knowledge systems and common knowledge systems. First of all, we recall that the evidence part in *EBK*-systems can be chosen independently of the knowledge system for individual agents, whereas the common knowledge operators are determined by the individual knowledge systems for the agents. Therefore, evidence-based knowledge systems cover more situations than the common knowledge systems. When both systems are present, e.g., in the case of $\mathsf{S4}_n^J$ and $\mathsf{S4}_n^C$, it is fair to compare them.

Operators C and J can be compared model theoretically. Each $\mathsf{S4}_n^C$-model is an $\mathsf{S4}_n^J$-model, but not the other way around, since the evidence accessibility in $\mathsf{S4}_n^J$-models contains (but not necessarily coincides with) the reachability on the frame $(W, R_1, \ldots, R_n)$. We could, however, impose a structure of an $\mathsf{S4}_n^C$-model on any $\mathsf{S4}_n^J$-model by adding the reachability relation for the operator C, which is done in a unique way for a given $\mathsf{S4}_n^J$-model. The resulting models $\mathcal{M}$ support the languages of both $\mathsf{S4}_n^C$ and $\mathsf{S4}_n^J$, thus providing a reasonable context for comparing knowledge operators C and J. The logic $\mathsf{S4}_n^{JC}$ is the set of tautologies in the langauge containing $K_1, \ldots, K_n, J, C$.

Proposition 8. *Evidence-based knowledge is stronger than common knowledge, i.e.,*

1. $J\varphi \rightarrow C\varphi$ *is valid;*
2. $C\varphi \rightarrow J\varphi$ *is not valid.*

Proof. 1. This obviously follows, since the common knowledge accessibility is a subset of the evidence accessibility.

2. For a counterexample, take a two-element model $W = \{a, b\}$, $R_i = \{(a,a), (a,b), (b,b)\}$, $R_J = R_i \cup \{(b,a)\}$. Then the transitive closure of all R_i will be the same R_i. Consider a forcing relation such that $a \not\Vdash S$ and $b \Vdash S$ for some sentence variable S. In this setup, $b \Vdash C(S)$, but $b \not\Vdash J(S)$. □

This baby example demonstrates, however, the main model-theoretical difference between common knowledge and evidence-based knowledge: the former captures the greatest solution of the Fixed-Point common knowledge equation $C\varphi \leftrightarrow E(\varphi \wedge C\varphi)$, whereas the latter considers all of its solutions.

To compare valid principles of common knowledge and evidence-based knowledge, consider a syntactic transformation $*$ that converts all occurrences of J into C.

Proposition 9. *Each evidence-based principle is a common knowledge principle, i.e.,*

$$(\mathsf{S4}_n^J)^* \subseteq \mathsf{S4}_n^C,$$

but not vice versa.

Proof. For $(\mathsf{S4}_n^J)^* \subseteq \mathsf{S4}_n^C$, it suffices to prove the $*$-translations of all the axioms and rules of $\mathsf{S4}_n^J$ in $\mathsf{S4}_n^C$. Let us check, for example, the necessitation rule for J: $\mathsf{S4}_n^J \vdash \psi \;\Rightarrow\; \mathsf{S4}_n^J \vdash J\psi$. Suppose $\mathsf{S4}_n^C \vdash \psi^*$, then $\mathsf{S4}_n^C \vdash \top \rightarrow E(\psi^*)$. Use the Induction Rule of $\mathsf{S4}_n^C$ (cf. [Fagin *et al.*, 1995]) to conclude that $\mathsf{S4}_n^C \vdash C\psi^*$, i.e., $\mathsf{S4}_n^C \vdash (J\psi)^*$. The remaining cases can be recovered by inspecting [Fagin *et al.*, 1995].

To show the remaining part of the claim, consider a valid $\mathsf{S4}_n^C$ principle[4]

$$\iota_C = \varphi \wedge C(\varphi \rightarrow E\varphi) \rightarrow C\varphi,$$

such that its J version

$$\iota_J = \varphi \wedge J(\varphi \rightarrow E\varphi) \rightarrow J\varphi$$

[4]This example was offered independently by Evan Goris and Eric Pacuit.

is not valid for $\mathsf{S4}_n^J$. Indeed, consider the same model as in the proof of Proposition 9.2, and pick φ such that $a \not\Vdash \varphi$, but $b \Vdash \varphi$. Then $b \Vdash J(\varphi \to E\varphi)$, since at each node where φ holds (b only), $E\varphi$ also does (b is the only node accessible from b by agent's relations $R_1, \ldots, R_n$). Hence $b \Vdash \varphi \wedge J(\varphi \to E\varphi)$. On the other hand, $a \not\Vdash \varphi$, bRa, hence $b \not\Vdash J\varphi$. □

8 Solution of the wise men puzzle

In this section, we will use evidence-based systems to give a solution to the wise men puzzle from [Fagin *et al.*, 1995], p.12. The story goes as follows:

> There are three wise men. It is common knowledge that there are three red hats and two white hats. The king puts a hat on the head of each of the three wise men, and asks them (sequentially) if they know the color of the hat on their head. The first wise man says that he does not know, the second wise man says that he does not know, then the third wise man says that he knows.
>
> (a) What color is the third wise man's hat?
>
> (b) Suppose the third wise man is blind and that it is common knowledge that the first two wise men can see. Can the third wise man still figure out the color of his hat?

We pick the smallest of *EBK* systems above, $\mathsf{T_3S4}$, describe the puzzle in its language, and present a solution consisting of a formal derivation in the corresponding theory.

Let atomic propositions p_i stand for "wise man i has a red hat" ($i = 1, 2, 3$),
"**000**" for $\neg p_1 \wedge \neg p_2 \wedge \neg p_3$;
"**001**" for $(\neg p_1 \wedge \neg p_2 \wedge p_3)$;
...
"**111**" for $(p_1 \wedge p_2 \wedge p_3)$.

Let also $K_i\$\varphi$ be a shorthand for $K_i\varphi \vee K_i\neg\varphi$, i.e., "$i$ knows whether φ." Note that $K_i p_i$ is equivalent in $\mathsf{T_3S4}$ to

$$(p_i \to K_j p_i) \wedge (\neg p_i \to K_j \neg p_i).$$

The basic assumption that each wise man observes the other wise men's hats is represented by the additional axiom "*KNOWING ABOUT THE OTHERS*," or "*K.A.O.*" for short:

$$K.A.O. = \bigwedge_{j \neq i} K_j\$p_i.$$

The rules of the game can be described by the theory

$$\mathsf{W(0)} = \mathsf{T_3S4} + J(K.A.O.) + J(\neg\mathbf{000}).$$

The situation after the first and second wise men said they didn't know is represented by a theory

$$\mathsf{W(2)} = \mathsf{W(0)} + J(\neg K_1\$p_1) + J(\neg K_2\$p_2).$$

Theorem 8. $\mathsf{W}(2) \vdash Jp_3$

Proof. We have now a comfortable choice of methods ranging from model reasoning to all sorts of proof systems. We will choose the latter and present a concise Hilbert style derivation in $\mathsf{W}(2)$.

First, we prove $J(\neg\mathbf{100})$.

1. $\mathbf{100} \rightarrow \neg p_2 \wedge \neg p_3$, by propositional logic;
2. $\mathbf{100} \rightarrow K_1(\neg p_2) \wedge K_1(\neg p_3)$, from $J(K.A.O.)$;
3. $\mathbf{100} \rightarrow K_1(\neg p_2 \wedge \neg p_3)$, by modal logic reasoning;
4. $K_1(\neg p_2 \wedge \neg p_3) \rightarrow K_1 p_1$, from $J(\neg\mathbf{000})$;
5. $\mathbf{100} \rightarrow K_1 p_1$, from 3. and 4.;
6. $\neg K_1\$p_1 \rightarrow \neg\mathbf{100}$, from 5.;
7. $J(\neg K_1\$p_1) \rightarrow J(\neg\mathbf{100})$, from 6., by J reasoning;
8. $J(\neg\mathbf{100})$, from $J(\neg K_1\$p_1)$ and 7.

Likewise, using $J(\neg K_2\$p_2)$ we obtain $J(\neg\mathbf{010})$.

Next, we prove $J(\neg\mathbf{110})$. Indeed,

1. $p_1 \wedge \neg p_3 \rightarrow \mathbf{110} \vee \mathbf{100}$, by propositional logic;
2. $p_1 \wedge \neg p_3 \rightarrow \mathbf{110}$, from $J(\neg\mathbf{100})$;
3. $p_1 \wedge \neg p_3 \rightarrow p_2$, by propositional logic;
4. $K_2 p_1 \wedge K_2 \neg p_3 \rightarrow K_2 p_2$, by modal logic reasoning;
5. $p_1 \wedge \neg p_3 \rightarrow K_2\$p_2$, from $J(K.A.O.)$;
6. $\mathbf{110} \rightarrow K_2\$p_2$, by propositional logic;
7. $\neg K_2\$p_2 \rightarrow \neg\mathbf{110}$, by propositional logic;
8. $J(\neg K_2\$p_2) \rightarrow J(\neg\mathbf{110})$, by J reasoning;
9. $J(\neg\mathbf{110})$, from $J(\neg K_2\$p_2)$ and 8.

Since all truth combinations of p_1, p_2 with $\neg p_3$ have been ruled out in $\mathsf{W}(2)$, this theory proves p_3, hence Jp_3. $\square$

Corollary 2. *Wise man 3 wears a red hat, and he will know this after the answers of 1 and 2, even without seeing their hats.*

Indeed, the above reasoning does not make use of the facts that $K_3\$p_1$ and $K_3\$p_2$.

One could also wonder whether conditions formalized in $W(2)$ are consistent. Here is a $\mathsf{T}_3\mathsf{S4}$-model for $W(2)$: W consists of three nodes (001), (011) and (101); R_1, R_2, R_3 are reflexive, $(001)R_1(101)$ and $(001)R_2(011)$, $R = W \times W$; $(001) \Vdash p_3$, $(011) \Vdash p_2, p_3$, $(101) \Vdash p_1, p_3$.

Formal methods are also good for verifying proofs and for analyzing assumptions made. In particular, here we can see that one does not need to assume the whole power of $\mathsf{S5}$ reasoning to solve this puzzle. A modest T as the agent knowledge logic and $\mathsf{S4}$ as the evidence component do the job just fine.

9 Discussion

The inspection axiom for evidence $t\!:\!\varphi \rightarrow\, !t\!:\!t\!:\!\varphi$ requires that each evidence assertion t is supported by some other evidence $!t$. This principle is nothing but an explicit version of a *transitivity* assumption commonly accepted as a property of knowledge, including the common knowledge. Since such a verification evidence could be of a very general character, we believe the inspection axiom holds for a wide range of situations involving evidence.

Monotonicity of evidence,

$$t{:}\varphi \rightarrow [(s+t){:}\varphi \wedge (t+s){:}\varphi],$$

requires that a given evidence t of φ remains such in the presence of any other piece of evidence s. This principle imposes certain restrictions on the class of situations covered by this kind of evidence-based approach. However, monotonicity has been a hidden assumption in the modal approach to knowledge in general. One needs this or a similar principle to provide all modal theorems with explicit evidence reading. Evidence-based knowledge systems just made this assumption explicit.

The language with explicit evidence also provides an opportunity to express principles which lie off the scope of the standard logic of knowledge. Here is an example from [Artemov and Nogina, 2004]. The weak principle of negative introspection $\neg x{:}\varphi \rightarrow \Box(\neg x{:}\varphi)$ holds, for example, for formal provability. It makes sense to consider extending S4-based systems by this principle without committing to S5.

10 Acknowledgements

I am very grateful to Walter Dean, Melvin Fitting, Evan Goris, Karen Kletter, Vladimir Krupski, Roman Kuznets, Elena Nogina, Eric Pacuit, Bryan Renne, Johan van Benthem, and Tatiana Yavorskaya whose advice and support helped me with this paper.

References

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

[Artemov and Nogina, 2004] S. Artemov and E. Nogina. Logic of knowledge with justifications from the provability perspective. Technical Report TR-2004011, CUNY Ph.D. Program in Computer Science, 2004.

[Artemov *et al.*, 1999] S. Artemov, E. Kazakov, and D. Shapiro. On logic of knowledge with justifications. Technical Report CFIS 99-12, Cornell University, 1999.

[Artemov, 1994] S. Artemov. Logic of proofs. *Annals of Pure and Applied Logic*, 67(1):29–59, 1994.

[Artemov, 1995] S. Artemov. Operational modal logic. Technical Report MSI 95-29, Cornell University, 1995.

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

[Aumann, 1976] R.J. Aumann. Agreeing to disagree. *Annals of Statistics*, 4(6):1236–1239, 1976.

[Brezhnev and Kuznets, 2005] V. Brezhnev and R. Kuznets. Making knowledge explicit: How hard it is. Technical Report TR-2005003, CUNY Ph.D. Program in Computer Science, 2005.

[Brezhnev, 2000] V. Brezhnev. On explicit counterparts of modal logics. Technical Report CFIS 2000-05, Cornell University, 2000.

[Brezhnev, 2001] V. Brezhnev. On the logic of proofs. In *Proceedings of the Sixth ESSLLI Student Session, Helsinki*, pages 35–46, 2001. `http://www.helsinki.fi/esslli/`.

[Fagin *et al.*, 1995] R. Fagin, J. Halpern, Y. Moses, and M. Vardi. *Reasoning About Knowledge.* MIT Press, 1995.

[Fitting, 2003] M. Fitting. A semantics for the logic of proofs. Technical Report TR-2003012, CUNY Ph.D. Program in Computer Science, 2003.

[Fitting, 2004] M. Fitting. Semantics and tableaus for LPS4. Technical Report TR-2004016, CUNY Ph.D. Program in Computer Science, 2004.

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

[Halpern and Moses, 1990] J.Y. Halpern and Y. Moses. Knowledge and common knowledge in a distributed environment. *Journal of ACM*, 37(3):549–587, 1990. A preliminary version appeared in *Proc. 3d ACM Symposium on Principles of Distributed Computing, 1984*.

[Kleene, 1952] S. Kleene. *Introduction to Metamathematics.* Van Norstrand, 1952.

[Krupski, 2002] V. Krupski. The single-conclusion proof logic and inference rules specification. *Annals of Pure and Applied Logic*, 113(1-3):181–201, 2002.

[Krupski, 2005] V. Krupski. Referential logic of proofs. *Theoretical Computer Science*, (accepted), 2005.

[Lehman, 1984] D.J. Lehman. Knowledge, common knowledge, and related puzzles. In *Proc. 3rd ACM Symp. on Principles of Distributed Computing*, pages 62–67, 1984.

[Lewis, 1969] D. Lewis. *Convention. A philosophical study.* Harvard University Press, Cambridge, Mass., 1969.

[McCarthy *et al.*, 1979] J. McCarthy, M. Sato, T. Hayashi, and S. Igarishi. On the model theory of knowledge. Technical Report STAN-CS-79-725, Stanford University, 1979.

[Mkrtychev, 1997] A. Mkrtychev. Models for the logic of proofs. In S. Adian and A. Nerode, editors, *Logical Foundations of Computer Science' 97, Yaroslavl'*, volume 1234 of *Lecture Notes in Computer Science*, pages 266–275. Springer, 1997.

[Nogina, 1994] E. Nogina. Logic of proofs with the strong provability operator. Technical Report ILLC Prepublication Series ML-94-10, Institute for Logic, Language and Computation, University of Amsterdam, 1994.

[Nogina, 1996] E. Nogina. Grzegorczyk logic with arithmetical proof operators. *Fundamental and Applied Mathematics*, 2(2):483–499, 1996. In Russian, URL `http://mech.math.msu.su/~fpm/eng/96/962/96206.htm`.

[Sidon, 1997] T. Sidon. Provability logic with operations on proofs. In S. Adian and A. Nerode, editors, *Logical Foundations of Computer Science' 97, Yaroslavl'*, volume 1234 of *Lecture Notes in Computer Science*, pages 342–353. Springer, 1997.

[Troelstra and Schwichtenberg, 1996] A. Troelstra and H. Schwichtenberg. *Basic Proof Theory.* Cambridge University Press, Amsterdam, 1996.

[van Benthem, 1991] J. van Benthem. Reflections on epistemic logic. *Logique & Analyse*, 133-134:5–14, 1991.

[Yavorskaya (Sidon), 2002] T. Yavorskaya (Sidon). Logic of proofs and provability. *Annals of Pure and Applied Logic*, 113(1-3):345–372, 2002.