Abstract

In this paper we introduce the justified knowledge operator $J$ with the intended meaning of $J \varphi$ as ‘there is a justification for $\varphi$.’ Though justified knowledge appears here in a case study of common knowledge systems, a similar approach is applicable in more general situations. First we consider evidence-based common knowledge systems obtained by augmenting a multi-agent logic of knowledge with a system of evidence assertions $t: \varphi$, reflecting the notion ‘$t$ is an evidence for $\varphi$,’ such that evidence is respected by all agents. Justified common knowledge is obtained by collapsing all evidence terms into one modality $J$. We show that in standard situations, when the base epistemic systems are $T$, $S4$, and $S5$, the resulting justified common knowledge systems are normal modal logics, which places them within the scope of well-developed machinery applicable to modal logic: Kripke-style epistemic models, normalized proofs, automated proof search, etc. In the aforementioned situations, the intended semantics of justified knowledge is supported by a realization theorem stating that for any valid fact about justified knowledge, one could recover its constructive meaning by realizing all occurrences of justified knowledge modalities $J \varphi$ by appropriate evidence terms $t: \varphi$.

1. Introduction

Plato’s celebrated tripartite definition of knowledge as justified true belief is generally regarded as a set of necessary conditions for the possession of knowledge. Due to Hintikka (cf. [23]), the ‘true belief’ components of Plato’s definition of knowledge have been fairly formalized by means of modal logic and its possible world semantics. Building on earlier works [6,18], the papers [9,10] introduced the notion of justification into formal epistemology by combining Hintikka-style epistemic modal logic with justification calculi arising from the logic of proofs [4–6]. Epistemic logic with justification, along with the usual knowledge operator $\square F$ ($F$ is known), contains assertions $t:F$ ($t$ is a justification for $F$). A natural epistemic semantics for epistemic logic with justification was provided by Fitting models which augment Kripke models with a natural treatment of justification assertions $t:F$ based on a special function specifying admissible evidence for each formula in a given world.

Common knowledge is a fundamental feature of multi-agent systems of knowledge which was first discussed in [30] and then studied in [11,21,27,32]. The book [17] 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

$$E \varphi = K_1 \varphi \land K_2 \varphi \land \cdots \land K_n \varphi.$$ 

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

$$C \varphi \iff \varphi \land E \varphi \land E^2 \varphi \land \cdots \land E^n \varphi \ldots .$$

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

$$C \varphi \iff E(\varphi \land C \varphi)$$

along with the Induction Rule

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

capturing the greatest solution of the corresponding fixed-point equation (cf. [17]). This kind of deductive system does not behave well proof-theoretically. In particular, there is no conventional cut-elimination in the common-knowledge systems [1]. This practically rules out automated proof search and 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. [17]), use a very strong, not formalized assumption that the agents possess a common knowledge of the same Kripke-style frame of possible situations.

The language of modal logic captures the most liberal version of the knowledge operator without imposing any conditions on the way this knowledge is attained. As a result, there might be non-constructive versions of knowledge appearing by chance or for some unknown reason or without any particular reasons at all. That is why the notion of justification has received much attention in epistemology (cf., for example, [12,19,20,22,28,29,31,36,41]). This suggests that there is a certain need for formal 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 $\varphi$’) based on the following plausible assumptions:

- all axioms have evidence;
- evidence is undeniable and implies individual knowledge of any agent;
- evidence is checkable;
- evidence is monotone, i.e., new evidence does not spoil existing 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.

Finally, for a given EBK-system, we introduce the justified knowledge operator $J \varphi$ (‘there is a justification for $\varphi$’) obtained by collapsing all evidence terms into one modality $J$:

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

The resulting justified knowledge systems ($T_n^J$, $S4_n^J$, and $S5_n^J$) are normal modal logics with natural epistemic Kripke-style semantics. Moreover, for any valid fact in $T_n^J$, $S4_n^J$, or $S5_n^J$, 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 justified common knowledge operator $J \varphi$ with the common knowledge operator $C \varphi$. Informally

$$J \varphi \Rightarrow \varphi \land E \varphi \land E^2 \varphi \land \cdots \land E^n \varphi \ldots ,$$
but the converse ‘$\iff$’ does not necessarily hold. Such a $J\varphi$ is not necessarily unique, which means that we have a variety of interpretations for a justified knowledge operator. In the epistemic Kripke-style semantics, $J \varphi$ corresponds to any accessibility relation which contains (but does not necessarily coincide with) reachability. Again, there might be many choices for such accessibility relation in a given model.

Justified common knowledge logics postulate $J$ as a normal (here $S4$-like) modality. In particular, this suffices to validate the Fixed-Point Axiom for $J$:

\[ J \varphi \iff E(\varphi \land J \varphi). \]

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

In paradigmatic examples where common knowledge has been used, justified common knowledge is also applicable. Moreover, justified common knowledge may have certain advantages.

1. An axiomatic approach in the form of justified common knowledge seems to avoid foundational loopholes of the standard model-theoretical reasoning about common knowledge. In particular, its application does not rely on a non-formalized assumption of common knowledge of a specific epistemic model; this assumption is normally neither acknowledged nor formalized; nor does it appear to be well-justified. Within an axiomatic approach, one lists all epistemic assumptions formally, which provides a more trustworthy and complete solution. Formalized solutions can be easier to analyze, verify, and optimize.

2. Justified common knowledge systems are simpler than the traditional common knowledge systems. Justified common knowledge is in the scope of well-developed methods in modal logic, e.g., proof theoretic. These include cut-elimination theorems that yield the possibility of automated proof search and verification.¹

3. The common-knowledge operator is a derivative of the agent knowledge operators and carries the features of the latter. The justified common 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

Evidence based knowledge systems, along with the usual knowledge modalities, contain evidence assertions of the format ‘$t$ is evidence for $\varphi$’, 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, $S4$, was found in [5,6] and known as the Logic of Proofs $LP$. Similar systems corresponding to $S5$ were introduced in [7,37,38]. Finally, [13,14] describes systems of terms corresponding to $K$, $K4$, $T$, $D$, and $D4$. These systems of explicit terms share several important features. Among these is the ability to internalize their own proofs as schematized by the Internalization Principle:

\[ \text{if } \vdash \varphi, \text{ then } \vdash p: \varphi \text{ for some proof term } p. \]

the Realization Theorem holds for the aforementioned systems, which asserts 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., $S4$ is the forgetful projection of $LP$. There are also other systems of explicit presentation of knowledge by evidence terms (‘+$’-free fragment of $LP$, [6,18], functional logic of proofs [25,26], etc.), where compatibility of evidence is not required.

Along with the usual choices of $K$, $T$, $K4$, $S4$, and $S5$ 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: $T_{a}LP$, $S_{a}LP$, and $S5_{a}LP$. In all these systems, the evidence logic is $LP$ (which corresponds to $S4$), whereas the base knowledge logics could be weaker ($T$ in $T_{a}LP$), equal to ($S4$ in $S4_{a}LP$), or stronger than ($S5$ in $S5_{a}LP$) the evidence logic. All these $EBK$-systems are supplied with epistemic semantics capturing the notion of admissible evidence.

¹ Some experimentation with automated proof search in justified knowledge systems were made in [16]. In particular, a computer automatically found a proof of the Muddy Children problem (cf. [17, Section 2.3]).
We also consider logics of justified common knowledge (JCK) corresponding to the above EBK-systems: $T_n^J$, $S4_n^J$, and $S5_n^J$ obtained from $T_n$LP, $S4_n$LP, and $S5_n$LP, respectively, by collapsing

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

The intended epistemic semantics of $J \varphi$ is ‘there is a justification for $\varphi$.’ The JCK-systems $T_n^J$, $S4_n^J$, and $S5_n^J$ are normal modal logics with standard Kripke-style semantics. We show that $T_n^J$, $S4_n^J$, and $S5_n^J$ enjoy an important realization property: given a formula $\varphi$ in the JCK-language derivable in $T_n^J$, $S4_n^J$, or $S5_n^J$, one could recover an EBK-formula $\psi$ derivable in the corresponding EBK-system $T_n$LP, $S4_n$LP, or $S5_n$LP, respectively, such that $\varphi$ is a forgetful projection of $\psi$. The realization property opens a possibility of first establishing $\varphi$ in a JCK-system $T_n^J$, $S4_n^J$, or $S5_n^J$, which can be a relatively easy task, and then recovering constructive evidence terms encoded by justified knowledge modality $J$ in $\varphi$, if needed.

We make an easy but fundamental observation that evidence assertions $t: \varphi$, as well as the justified knowledge modality $J$, satisfy the Fixed-Point Axiom above and hence may be regarded as a special constructive 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. As was shown in [3], $T_n^J$ and $S4_n^J$ coincide with McCarthy’s ‘any fool knows’ common knowledge systems (cf. [32]). This observation provides McCarthy’s dummy ‘any fool’ agent with a evidence-based epistemic semantics.

On the technical side, we prove that $T_n^J$ and $S4_n^J$ enjoy cut-elimination theorems and give algorithms of recovering evidence terms in JCK-modalities in all three forgetful systems $T_n^J$, $S4_n^J$, and $S5_n^J$. 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 justified common knowledge $J \varphi$ and common knowledge $C \varphi$. A general answer here is that the JCK approach to common knowledge leads to a stronger modality but leaner axiom systems. Whenever such a comparison is appropriate, JCK-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 JCK-identity is common knowledge-compliant, which justifies using JCK-systems as common knowledge systems. In particular, $(S4_n^J)^* \subset S4_n^C$ but $(S4_n^J)^* \neq S4_n^C$, where $*$ stands for an operation of renaming $J$ to $C$.

Our experience says that JCK-systems are applicable and may provide additional insights whenever the usual common knowledge systems are. As an example, a solution of the Wise Men puzzle (Section 8) is given as a formal derivation in $T_3^J$.

3. Formal systems of evidence-based knowledge

We first introduce the multi-agent logics of the evidence based knowledge series $T_n$LP, $n = 1, 2, 3, \ldots$. In brief, $T_n$LP contains $n$ copies of $T$-style modalities representing knowledge operators of $n$ agents, $K_1, \ldots, K_n$ (cf. systems $T_n$ from [17]); in addition, it contains a system of evidence assertions taken from the logic of proofs LP.

Evidence assertions in $T_n$LP have the form $t: \varphi$, where $\varphi$ is a formula and $t$ is an evidence term (or just evidence) built from evidence constants $a, b, c, \ldots$ and evidence variables $x, y, z, \ldots$ with the help of three operations, application ‘·’, (binary), union ‘+’ (binary), and inspection ‘!’ (unary). Formally, if $t$ is an evidence and $S$ is a sentence variable, the formulas of $T_n$LP are defined by the following grammar:

$$\varphi = \bot | S[\varphi_1 \rightarrow \varphi_2] | \varphi_1 \land \varphi_2 | \varphi_1 \lor \varphi_2 | \neg \varphi | K_1 \varphi | t: \varphi.$$  

We assume also that ‘$t$:’, ‘$K_1$’, and ‘$\neg$’ bind stronger than ‘$\land$, $\lor$’ which, in turn, bind stronger than ‘$\rightarrow$’. $T_n$LP has axioms of both $T_n$ and LP, together with the principle that an evidence assertion yields knowledge of each individual agent: $t: \varphi \rightarrow K_1 \varphi$. This is a schema when there is one growing system of evidence accepted by all the agents.

**Definition 1.** Axioms and rules of $T_n$LP are:

1. **Classical propositional logic**

   The standard set of axioms of the classical propositional logic, e.g., A1–A10 from [24] or a similar system R1. **Modus Ponens**.
Group (III) introduces some combinatorial properties of evidence and explains the meaning of evidence terms. E1 is merely the internalized \emph{modus ponens}, which says that an evidence for \( \varphi \to \psi \) can be applied to an evidence for \( \varphi \) to produce an evidence for \( \psi \). E2 expresses the principle that any evidence \( t \) of \( \varphi \) can be verified by a new evidence \( \dagger t \) (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 \( \varphi \), then \( t \) combined with any other evidence still remains an evidence for \( \varphi \). R3 assigns initial evidence in the form of constants to any axiom of \( T_n \)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 and C1. Naturally, all axioms are in fact schemas in the language of \( T_n \)LP. All rules are applied across sections (I)–(IV).

Consider two more series of principles:

\begin{itemize}
  \item B3i. \( K_i \varphi \to K_i K_i \varphi \) (positive introspection),
  \item B4i. \( \neg K_i \varphi \to K_i \neg K_i \varphi \) (negative introspection).
\end{itemize}

System \( S4_n \)LP is obtained from \( T_n \)LP by adding B3i and \( S5_n \)LP is \( S4_n \)LP plus B4i, \( i = 1, \ldots, n \). Again, all the rules R2i 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 \( \Gamma, \varphi \vdash \psi \Rightarrow \Gamma \vdash \varphi \to \psi \).

**Lemma 1.** For any formula \( \varphi \) and each \( i = 1, 2, \ldots, n \) there is an evidence term \( u_{\varphi}(x) \) such that \( T_n \)LP (hence \( S4_n \)LP and \( S5_n \)LP) proves \( x:\varphi \to u_{\varphi}(x):K_i \varphi \).

**Proof.**
1. \( x:\varphi \to K_i \varphi \), by C1;
2. \( a: (x:\varphi \to K_i \varphi) \), introducing evidence \( a \), by R3;
3. \( \dagger x:x:\varphi \to (a:\dagger x):K_i \varphi \), by E1 and propositional logic;
4. \( x:\varphi \to \dagger x:x:\varphi \), by E2;
5. \( x:\varphi \to (a:\dagger x):K_i \varphi \), by propositional logic.

It suffices now to set \( u_{\varphi}(x) \) equal to \( a:\dagger x \) such that \( a: (x:\varphi \to K_i \varphi) \).

**Proposition 1 (Internalization).** Given \( T_n \)LP \( \vdash \varphi \), there is an evidence term \( p \) such that \( T_n \)LP \( \vdash p: \varphi \). The same holds for \( S4_n \)LP and \( S5_n \)LP.

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

1. \( \varphi \) is obtained from \( \psi \to \varphi \) and \( \psi \) by \emph{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 \( p = s \cdot t \).
2. If \( \varphi \) is obtained by R2i, then \( \varphi \) is \( K_i \psi \) and \( \vdash \psi \). By the Induction Hypothesis, \( \vdash t: \psi \) for some evidence \( t \). Use Lemma 1 to conclude that \( \vdash u_{\psi}(t): K_i \psi \), and \( p \) is \( u_{\psi}(t) \).
3. If \( \varphi \) is obtained by R3, then \( \varphi \) is \( c:A \) for some constant \( c \) and axiom \( A \). Use the axiom E2 to derive \( !c: c: A \), i.e., \( !c: \varphi \).

Here \( p = !c \).

Note that the evidence term \( p \) is ground and built from atomic evidence terms by applications and inspections only.
A similar argument establishes a more general form of internalization: If \( \psi_1, \ldots, \psi_k \vdash \varphi \), then for some evidence term \( p(x_1, \ldots, x_k) \),
\[
\psi_1, \ldots, x_k \vdash p(x_1, \ldots, x_k); \varphi.
\]
Both of the previous formulations of internalization follow from the Proposition 2 which holds for \( T_{n} \text{LP} \), \( S_{4n} \text{LP} \), and \( S_{5n} \text{LP} \).

**Proposition 2 (Lifting).** If \( \psi_1, \ldots, \psi_k, y_1; x_1, \ldots, y_m; x_m \vdash \varphi \), then for some term \( p(x_1, \ldots, x_k, y_1, \ldots, y_m) \),
\[
\psi_1, \ldots, x_k \vdash p(x_1, \ldots, x_k, y_1, \ldots, y_m); \varphi.
\]

**Proof.** Similar to Proposition 1 with two new base clauses. If \( \varphi \) is \( \psi_i \), then \( p \) is \( x_i \). If \( \varphi \) is \( y_j; x \), then \( p \) is \( !y_j \). \( \square \)

The internalization property states that all derived facts have witnesses. Internalization naturally extends to the case when \( T_{n} \text{LP} \), \( S_{4n} \text{LP} \), or \( S_{5n} \text{LP} \) 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 \( \varphi \) and each \( i = 1, 2, \ldots, n \) there is an evidence term \( \downarrow_{\text{down}}(x) \) such that \( T_{n} \text{LP} \) proves \( x; K_i \varphi \rightarrow \downarrow_{\text{down}}(x); \varphi \). The same holds for \( S_{4n} \text{LP} \) and \( S_{5n} \text{LP} \).

**Proof.** \( x; K_i \varphi \rightarrow \varphi \), by E4 and B2;
\( b; (x; K_i \varphi \rightarrow \varphi) \), introducing evidence \( b \), by Proposition 1;
\( \vdash x; K_i \varphi \rightarrow (b; \downarrow_{\text{down}}(x); \varphi) \), by E1 and propositional logic;
\( x; K_i \varphi \rightarrow \downarrow_{\text{down}}(x; K_i \varphi) \), by E2;
\( x; K_i \varphi \rightarrow \downarrow_{\text{down}}(x; K_i \varphi) \), by propositional logic.

It suffices now to set \( \downarrow_{\text{down}}(x) \) equal to \( b; \downarrow_{\text{down}}(x; K_i \varphi) \) such that \( b; (x; K_i \varphi \rightarrow \varphi) \). \( \square \)

A natural assumption about common knowledge is that \( \varphi \) is common knowledge (written \( C \varphi \) iff all agents know that \( \varphi \) and \( C \varphi \). This leads to the Fixed-Point Axiom (cf. [17]):
\[
C \varphi \leftrightarrow E(\varphi \land C \varphi),
\]
where \( E \varphi = K_1 \varphi \land \cdots \land K_n \varphi \). We show that \( t; \varphi \) provably satisfies a similar fixed-point identity in \( T_{n} \text{LP} \), \( S_{4n} \text{LP} \), and \( S_{5n} \text{LP} \).

**Proposition 3.** For each evidence term \( t \), \( t; \varphi \) satisfies the Fixed-Point Axiom in \( T_{n} \text{LP}(S_{4n} \text{LP}, S_{5n} \text{LP}) \).

**Proof.** We prove that \( T_{n} \text{LP} \vdash t; \varphi \leftrightarrow E(\varphi \land t; \varphi) \).
1. \( t; \varphi \rightarrow K_i \varphi \), for all \( i = 1, \ldots, n \); hence \( t; \varphi \rightarrow E \varphi \);
2. \( t; \varphi \rightarrow !t; \varphi \); hence \( t; \varphi \rightarrow K_i!t; \varphi \), for all \( i = 1, \ldots, n \), and \( t; \varphi \rightarrow E t; \varphi \);
3. \( t; \varphi \rightarrow E(\varphi \land t; \varphi) \);

which concludes the left-to-right part of the proof. The right-to-left part \( E(\varphi \land t; \varphi) \rightarrow t; \varphi \) follows from the reflexivity of \( E \). \( \square \)

4. Models

Arithmetical provability semantics for the logic of proofs LP was introduced in [5,6] and extended to evidence-based knowledge systems in [8]. Kripke-style models for modal logics of provability and explicit proofs were considered in [4] and then generalized in [8,34,35,39,42]. Epistemic Kripke-based semantics for the logic of proofs was suggested by Fitting (cf. [18]).

At the heart of this semantics lies the idea, which can be traced back to Mkrtchyan’s [33], of augmenting Boolean (Mkrtchyan) or Kripke-style (Fitting) models by an evidence function which indicates which terms are ‘admissible’ as an evidence to a given statement. 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 \( \varphi \) in \( u \);
2. \( \varphi \) holds in all worlds accessible from \( u \).

Fitting models were adapted for single-agent evidence-based knowledge systems in \([8\text{-}10]\). In this section, we generalize Fitting semantics to capture multi-agent \( EBK \)-systems \( T_n^{LP}, S_4_n^{LP} \), and \( S_5_n^{LP} \); the aforementioned Kripke, Mkrtychev, and Fitting models will be special cases of this semantics. A new feature of these models is a designated evidence accessibility relation.

A \( T_n^{LP} \)-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 \( R_1 \cup \cdots \cup R_n \), 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 \( E \) is a mapping from states and evidence terms to sets of formulas. We can read \( \varphi \in E(u, t) \) as ‘\( \varphi \) 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.** \( E \) is an evidence function on \( (W, R_1, \ldots, R_n, R) \) if for all evidence terms \( s \) and \( t \), for all formulas \( \varphi \) and \( \psi \), and for all \( u, v \in W \):
1. **Monotonicity:** \( uRv \) implies \( E(u, t) \subseteq E(v, t) \).
2. **Application:** \( \varphi \rightarrow \psi \in E(u, s) \) and \( \varphi \in E(u, t) \) implies \( \psi \in E(u, s \cdot t) \).
3. **Inspection:** \( \varphi \in E(u, t) \) implies \( t \varphi \in E(u, t) \).
4. **Sum:** \( E(u, s) \cup E(u, t) \subseteq E(u, s + t) \).

A \( T_n^{LP} \)-model is a structure \( M = (W, R_1, \ldots, R_n, R, E, \mathcal{I}) \), where \( E \) is an evidence function on the frame \( (W, R_1, \ldots, R_n, R) \) and \( \mathcal{I} \) is an arbitrary mapping from sentence variables to subsets of \( W \). Given a model \( M = (W, R_1, \ldots, R_n, R, E, \mathcal{I}) \), a forcing relation \( \models \) is extended from sentence variables to all formulas by the following rules. For each \( u \in W \):
1. \( \models \) respects Boolean connectives at each world.
2. \( u \models K_i \varphi \) iff \( v \models \varphi \) for every \( v \in W \) with \( u R_i v \).
3. \( u \models t \varphi \) iff \( \varphi \in E(u, t) \) and \( v \models \varphi \) 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 \( \varphi \) in \( u \) and \( \varphi \) is true at all worlds \( v \) accessible from \( u \) via a given evidence accessibility relation \( R \). We say \( \varphi \) is true at world \( u \in W \) if \( u \models \varphi \); otherwise, \( \varphi \) is false at \( u \). A formula \( \varphi \) is true in a model if \( \varphi \) is true at each world of the model; \( \varphi \) is valid if \( \varphi \) is true in every model.

A **constant specification** is a map \( CS \) from evidence constants to (possibly empty) sets of axioms. A constant specification \( CS \) 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 \( CS \), a model \( M \) meets \( CS \) if \( M \models c \varphi \) whenever \( \varphi \in CS(a) \). A derivation (in any of \( T_n^{LP}, S_4_n^{LP} \), or \( S_5_n^{LP} \)) meets \( CS \) if whenever rule R3 is used to produce \( c \varphi \), then \( \varphi \in CS(a) \).

\( S_4_n^{LP} \)- and \( S_5_n^{LP} \)-models are defined as \( T_n^{LP} \)-models with only this difference: for \( S_4_n^{LP} \)-models, the accessibility relations \( R_1, \ldots, R_n \) are reflexive and transitive; in \( S_5_n^{LP} \)-models, \( R_1, \ldots, R_n \) are reflexive, transitive, and symmetric. A set \( S \) of formulas is **\( CS-S_4_n^{LP} \)-satisfiable** (\( CS-T_n^{LP} \)-satisfiable, \( CS-S_5_n^{LP} \)-satisfiable) if there is an \( S_4_n^{LP} \)-model (\( T_n^{LP} \)-model, \( S_5_n^{LP} \)-models) \( M \), meeting \( CS \), and containing a world \( u \) such that \( M, u \models \varphi \) for all \( \varphi \in S \).

The usual Kripke models for \( T_n \), \( S_4_n \), and \( S_5_n \) are \( T_n^{LP} \)-, \( S_4_n^{LP} \)-, and \( S_5_n^{LP} \)-models, respectively, where the evidence part (\( R \) and \( E \)) is ignored. Mkrtychev models \(^2\) for \( LP \) are single-world \( T_n^{LP} \)-models. Fitting models \(^3\) for \( LPS4 \) are \( S_4_n^{LP} \)-models with \( R_1 = R \). Kripke models for \( S_4_n^{LP} \) + **weak negative introspection** \( \neg t : \varphi \rightarrow \Box (\neg t : \varphi) \) from \([8]\) are \( S_4_n^{LP} \)-models with \( R = W \times W \).

---

\(^2\) Called pre-models in \([33]\).

\(^3\) Called weak models in \([18]\).
Theorem 1 (Completeness Theorem). Let $CS$ be a constant specification. A formula $\varphi$ is provable in $T_n^{LP}$ ($S_4_n^{LP}$, $S_5_n^{LP}$) meeting $CS$ iff $\varphi$ holds in all $T_n^{LP}$-models (respectively, $S_4_n^{LP}$-models, $S_5_n^{LP}$-models) meeting $CS$.

Proof. We will give a proof for $S_4_n^{LP}$ making note of how to modify this proof for the remaining cases of $T_n^{LP}$ and $S_5_n^{LP}$.

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 $u R v$. Since $R_i \subseteq R$, $v \vdash \varphi$ for all $v$ such that $u R_i v$; 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 $\Gamma$ is consistent if there is no finite subset $\varphi_1, \ldots, \varphi_n$ such that $-(\varphi_1 \land \cdots \land \varphi_n)$ is provable in $S_4_n^{LP}$ meeting $CS$. A consistent set $\Gamma$ is maximal consistent if for any formula $\psi$, either $\psi \in \Gamma$ or $-\psi \in \Gamma$. By the standard Lindenbaum construction, each consistent set can be extended to a maximal consistent set.

We define the canonical model $M = (\mathfrak{W}, R_1, \ldots, R_n, R, E, \vdash)$ for $S_4_n^{LP}$ with a given constant specification $CS$. $W$ is the collection of all maximal consistent sets. If $\Gamma$ is a set of formulas, let $\Gamma^{\sharp i} = \{ \varphi | K_i \varphi \in \Gamma \}$ and $\Gamma^\flat = \{ \varphi | t: \varphi \in \Gamma \}$.

Now define the accessibility relations $R_1, \ldots, R_n, R$ as follows:

$$\Gamma R_i A \text{ iff } \Gamma^{\sharp i} \subseteq A, \ i = 1, \ldots, n; \quad \Gamma R A \text{ iff } \Gamma^\flat \subseteq A.$$ 

Note that $R_i$ are reflexive and transitive (for $S_4_n^{LP}$). For $S_5_n^{LP}$, relations $R_i$ are also symmetric. Suppose $\Gamma R_i A$ and $K_i \varphi \in \Gamma$. 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, $-K_i \varphi \in \Gamma$.

By the axiom $-K_i \varphi \rightarrow K_i -K_i \varphi$, $K_i -K_i \varphi \in \Gamma$. Since $\Gamma R_i A$, $-K_i \varphi \in \Delta$, which contradicts the consistency of $A$.

Note that $R$ is reflexive too. Moreover, $R$ is transitive. Indeed, let $\Gamma R A$ and $\Delta R \Theta$. If $t: \varphi \in \Gamma$, then $\vdash t: t: \varphi \in \Gamma$ and $t: \varphi \in \Delta$. Likewise, $\vdash t: t: \varphi \in \Gamma$ and $t: \varphi \in \Theta$. By reflexivity, $\varphi \in \Theta$. Let us check that $R$ contains all $R_i$’s, $i = 1, \ldots, n$.

Suppose $\Gamma R_i A$ and $t: \varphi \in \Gamma$. Then $\vdash t: t: \varphi \in \Gamma$ and $K_i t: \varphi \in \Gamma$; hence $t: \varphi \in \Delta$. By reflexivity, $\varphi \in \Delta$.

Define the evidence function $E$ as follows:

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

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

Proof. By induction on $\varphi$. 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 $\Delta \vdash K_i \varphi$.

If $K_i \varphi \notin \Gamma$, then $\Gamma' = \Gamma^{\sharp i} \cup \{ \neg \varphi \}$ is consistent. Otherwise $\psi_1 \lor \psi_2 \lor \cdots \lor \psi_t \rightarrow \varphi$ would be provable for some $K_i \psi_1, K_i \psi_2, \ldots, K_i \psi_t \in \Gamma$; hence $K_i \psi_1 \land K_i \psi_2 \land \cdots \land K_i \psi_t \rightarrow K_i \varphi$ would also be provable, which would make $\Gamma$ inconsistent. Let $\Delta$ be a maximal consistent set containing $\Gamma'$. Then $\Gamma R_i \Delta$, $-\varphi \in \Delta$; 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 the Truth Lemma: $\varphi = t: \psi$. Let $t: \psi \in \Gamma$. Then, by the definition of the evidence function, $\psi \in E(\Gamma, t)$. It remains to be shown that $\Delta \vdash \psi$ for all $\Delta$’s such that $\Gamma R \Delta$. Take such a $\Delta$. 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 $\psi \in E(\Gamma, t)$ and $t: \psi \in \Gamma$ by the definition of the evidence function $E$.

It is easy to see now that $M = (\mathfrak{W}, R_1, \ldots, R_n, R, E, \vdash)$ is an $S_4_n^{LP}$-model meeting constant specification $CS$. Indeed, by the definition of a consistent set, $CS \subseteq \Gamma$, for each $\Gamma \in \mathfrak{W}$. By Truth Lemma 3, $\Gamma \vdash CS$.

Let us finish the proof of Theorem 1. If $\varphi$ is not provable in $S_4_n^{LP}$ meeting constant specification $CS$, then $M$ is a countermodel for $\varphi$: consider $\neg \varphi$, which is consistent, and hence contains in a maximal consistent set $\Gamma$. 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 [18].
Proposition 5 (Compactness). For a given constant specification $CS$, a set of formulas $U$ is $CS\text{-}S4_n\text{-}LP$-(CS-$T_n\text{-}LP$-, $CS\text{-}S5_n\text{-}LP$-)satisfiable if any finite subset of $U$ is $CS\text{-}S4_n\text{-}LP$-(CS-$T_n\text{-}LP$-, $CS\text{-}S5_n\text{-}LP$-)satisfiable.

Proof. Suppose any finite subset of $U$ is $CS\text{-}S4_n\text{-}LP$-satisfiable. We will find a world $\Gamma$ in the canonical $CS\text{-}S4_n\text{-}LP$-model such that $\Gamma \models U$. First, note that $U$ is a consistent set. Otherwise, for some $X_1, \ldots, X_m \in U$, $CS\text{-}S4_n\text{-}LP$ proves $\neg(X_1 \land \cdots \land X_m)$, which would make $\{X_1, \ldots, X_m\}$ a finite unsatisfiable subset of $U$, which is impossible. Extend $U$ to a maximal consistent set $\Gamma$, which is hence a world in the canonical $CS\text{-}S4_n\text{-}LP$-model. Since $U \subseteq \Gamma$, by Truth Lemma 3, $\Gamma \models U$. □

The Fully Explanatory property of the canonical models for the logic of proofs was introduced by Fitting in [18]. This property might be summarized as ‘whatever is known is known for a reason.’

Definition 3. An $S4_n\text{-}LP$-(T-$n\text{-}LP$-, $S5_n\text{-}LP$-)model is Fully Explanatory provided that, whenever $v \models \varphi$ for every $v$ such that $uRv$, then for some evidence term $t$ we have $u \models t\varphi$.

Proposition 5 (Fully Explanatory property). For any full constant specification $CS$, the canonical $CS\text{-}S4_n\text{-}LP$-(CS-$T_n\text{-}LP$-, $CS\text{-}S5_n\text{-}LP$-)model is Fully Explanatory.

Proof. The proof follows Fitting’s [18]. We establish the contrapositive. Let $t\varphi \notin \Gamma$ for each evidence term $t$. Consider a set $U = \Gamma^0 \cup \{\neg \varphi\}$. We claim that $U$ is consistent. Otherwise, for some $t_1:\psi_1, t_2:\psi_2, \ldots, t_k:\psi_k \in \Gamma$, $CS\text{-}S4_n\text{-}LP$ proves $\psi_1 \rightarrow (\psi_2 \rightarrow ((\psi_3 \rightarrow \cdots \rightarrow \varphi) \cdots))$. Since $CS$ is a full constant specification, there is an evidence term $s$ such that $CS\text{-}S4_n\text{-}LP$ proves $s:(\psi_1 \rightarrow (\psi_2 \rightarrow ((\psi_3 \rightarrow \cdots \rightarrow \varphi) \cdots))$. Using E1, we establish that $CS\text{-}S4_n\text{-}LP$ proves

$$t_1:\psi_1 \rightarrow (t_2:\psi_2 \rightarrow (t_3:\psi_3 \rightarrow \cdots \rightarrow (s t_1 t_2 \ldots t_k):\varphi) \ldots).$$

Hence $(s t_1 t_2 \ldots t_k):\varphi \in \Gamma$, which contradicts the assumption about $\Gamma$.

Now take $\Delta$ to be a maximal consistent extension of $U$. It is clear that $\Delta$ is a world in a canonical model and that $\Gamma R \Delta$. By Truth Lemma 3, $\Delta \not\models \varphi$. □

6. Justified knowledge

In this section, we introduce a light version of evidence-based knowledge, which we call justified knowledge, in the form of a new modal operator $J \varphi$ (read $\varphi$ 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 justified common knowledge is a modal language with $n+1$ modalities $K_1, \ldots, K_n, J$. Systems $T_n^J, S4_n^J$, and $S5_n^J$ are specified as $T_n, S4_n$, and $S5_n$, with the modalities $K_1, \ldots, K_n$ augmented by $S4$ with the modality $J$, together with the principle: for all $i = 1, \ldots, n$

$$J \varphi \leftrightarrow K_i \varphi.$$

Apparently, the dummy $(n+1)$st agent corresponding to $J$ plays the role of a sceptical $S4$-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. As was noticed in [2,3], $T_n^J$ and $S4_n^J$ correspond to McCarthy’s systems with the ‘any fool knows’ modality (cf. [32]).

Lemma 4. In each of $T_n^J$, $S4_n^J$, and $S5_n^J$,

$$K_i J \varphi \leftrightarrow J \varphi \leftrightarrow J K_i \varphi.$$

Proof. Immediate from $K_i$-reflexivity and the following derivations.

$J \varphi \rightarrow JJ \varphi \rightarrow K_i J \varphi$;

$J \varphi \rightarrow JJ \varphi \rightarrow JK_i \varphi$;
\(K_i J \varphi \rightarrow J \varphi;\)
\(K_i \varphi \rightarrow \varphi, J (K_i \varphi \rightarrow \varphi), J K_i \varphi \rightarrow J \varphi. \square\)

**Proposition 6.** Justified common knowledge \(J\) satisfies the Fixed-Point Axiom in each of \(T_n^J, S4_n^J,\) and \(S5_n^J.\)

**Proof.** Deriving the Fixed-Point identity for \(J\) in \(T_n^J\) (hence in \(S4_n^J\) and \(S5_n^J\))

\(J \varphi \leftrightarrow E(\varphi \land J \varphi)\)

is similar to Proposition 3. \(\square\)

**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. \(S4_n^J\)-models are those where \(R_1, \ldots, R_n\) are reflexive and transitive. \(S5_n^J\)-models are those with reflexive, transitive, and symmetric \(R_1, \ldots, R_n\).

**Proposition 7.** \(T_n^J(S4_n^J, S5_n^J)\) is sound for \(T_n^J\)-models \((S4_n^J\)-models, \(S5_n^J\)-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 \(u R_i v\). Then \(u R v\) also holds, which brings \(u \Vdash \varphi\). Hence, \(u \Vdash K_i \varphi. \square\)

Completeness also occurs. For \(T_n^J\) and \(S4_n^J\), this will follow from Theorem 3 below. The completeness of \(S5_n^J\) will be established in Theorem 5.

**Definition 6.** A **sequent** is a pair of finite sets of \(S4_n^J\)-formulas presented as \(\Gamma \Rightarrow \Delta\). To simplify proofs, we assume a Boolean basis \(\rightarrow, \bot\) and treat the remaining Boolean connectives as definable. Axioms of \(S4_n^J G\) are the sequents \(S \Rightarrow \bot\) and \(\bot \Rightarrow \Rightarrow\), where \(S\) is a propositional variable. The propositional rules of \(S4_n^J G\) are those from the classical propositional Gentzen-style system, including Weakening and Cut (cf. [40]). In addition, there are \(n + 1\) pairs of proper modal rules:

\[
\frac{\varphi, \Gamma \Rightarrow \Delta}{\square \varphi, \Gamma \Rightarrow \Delta} \quad (\square, \Rightarrow) \quad \text{and} \quad \frac{J \Gamma, \square \Delta \Rightarrow \varphi}{J \Gamma, \Rightarrow \square \Delta \Rightarrow \square \varphi} \quad (\Rightarrow, \square),
\]

where \(\square \in \{K_1, \ldots, K_n, J\}\) and \(\square \{\phi_1, \ldots, \phi_m\} = \{\square \phi_1, \ldots, \square \phi_m\} \).

The Gentzen-style version \(T_n^J G\) of \(T_n^J\) has the same rules as \(S4_n^J G\) with the \((\Rightarrow, \square)\) rule replaced by

\[
\frac{J \Gamma, \Rightarrow \varphi}{J \Gamma, \square \Delta \Rightarrow \square \varphi} \quad (\Rightarrow, \square).
\]

**Theorem 2 (Equivalence of Gentzen- and Hilbert-style systems).** \(\Gamma \Rightarrow \Delta\) is provable in \(S4_n^J G(T_n^J G)\) iff \(\Gamma \Rightarrow \Delta\) is provable in \(S4_n^J (T_n^J)\).

**Proof.** The part ‘only if,’ i.e., that \(S4_n^J G \vdash \Gamma \Rightarrow \Delta\) yields \(S4_n^J \vdash \bigwedge \Gamma \rightarrow \bigvee \Delta\), is a standard exercise in modal derivation. Let us check the soundness of the \((\Rightarrow, \square)\)-rule in \(S4_n^J G\). By the Induction Hypothesis,

\[S4_n^J \vdash \bigwedge J \Gamma \land \bigwedge \square \Delta \rightarrow \varphi.\]

By \(S4\)-reasoning,

\[S4_n^J \vdash \bigwedge J \Gamma \rightarrow (\bigwedge \square \Delta \rightarrow \varphi).\]
By Lemma 4,
\[ S^4_n \vdash J \Gamma \rightarrow \Box(\Box \Delta \rightarrow \varphi). \]

Use distribution to establish
\[ S^4_n \vdash J \Gamma \rightarrow (\Box \Delta \rightarrow \Box \varphi). \]

By S4-reasoning,
\[ S^4_n \vdash J \Gamma \rightarrow (\Box \Delta \rightarrow \Box \varphi); \]

hence
\[ S^4_n \vdash J \Gamma \land \Box \Delta \rightarrow \Box \varphi. \]

Let us now check the soundness of the \((\Rightarrow, \Box)-rule\) in \(T_n^\varphi \mathcal{G}\). By the Induction Hypothesis,
\[ T_n^\varphi \vdash J \Gamma \land \Box \Delta \rightarrow \varphi. \]

By T-reasoning,
\[ T_n^\varphi \vdash \Box J \Gamma \rightarrow (\Box \Delta \rightarrow \Box \varphi). \]

By Lemma 4,
\[ T_n^\varphi \vdash J \Gamma \rightarrow (\Box \Delta \rightarrow \Box \varphi). \]

The ‘if’ direction for both \(S^4_n\) and \(T_n^\varphi\) will be established in Corollary 1. □

**Theorem 3 (Completeness Theorem).** The following are equivalent:

1. \( \Gamma \Rightarrow \Delta \) is provable in \( S^4_n \mathcal{G}(T_n^\varphi \mathcal{G}) \) without cut;
2. \( \Gamma \Rightarrow \Delta \) is provable in \( S^4_n \mathcal{G}(T_n^\varphi \mathcal{G}) \);
3. \( \land \Gamma \rightarrow \lor \Delta \) is provable in \( S^4_n(T_n^\varphi) \);
4. \( \land \Gamma \rightarrow \lor \Delta \) is \( S^4_n \)-valid (\( T_n^\varphi \)-valid);
5. \( \land \Gamma \rightarrow \lor \Delta \) is valid in all finite \( S^4_n \)-models (\( T_n^\varphi \)-models).

**Proof.** We will prove the case of \( S^4_n \) in detail. The case of \( T_n^\varphi \) is treated similarly, and we will show what modifications should be made in the \( S^4_n \) proof to make it work for \( T_n^\varphi \) as well. Steps (1) \( \Rightarrow \) (2) and (4) \( \Rightarrow \) (5) are trivial, (2) \( \Rightarrow \) (3) \( \Rightarrow \) (4) has already been covered above. We will concentrate on proving that (5) \( \Rightarrow \) (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 \( S^4_n \mathcal{G} \) without cut, we build a finite \( S^4_n \)-model \( M \), such that at some node of \( M \), all formulas from \( \Gamma_0 \) hold and all formulas from \( \Delta_0 \) do not hold.

To keep the domain of a model finite, we will consider only formulas from a given finite set \( F \) 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 \( S^4_n \mathcal{G} \) 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 \( \psi \in \Delta \);
- \( \square \varphi \in \Gamma \) yields \( \varphi \in \Gamma \) where \( \square \in \{ K_1, \ldots, K_n, J \} \).

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 \( F \), its saturation consists of formulas from \( F \) too.

Define a model \( M = (W, R_1, \ldots, R_n, R, \models) \). \( W \) will be the (finite) set of all consistent saturated sequents. Let \( I^\varphi = \{ J \varphi | J \varphi \in \Gamma \} \) and \( I^\varphi_i = \{ K_i \varphi | K_i \varphi \in \Gamma \} \). Set

\[ (\Gamma \Rightarrow \Delta) R (\Gamma' \Rightarrow \Delta') \text{ iff } I^\varphi \subseteq I', \quad (\Gamma \Rightarrow \Delta) R_i (\Gamma' \Rightarrow \Delta') \text{ iff } I^\varphi \cup I^\varphi_i \subseteq I'. \]
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

\[(\Gamma \Rightarrow \Delta) R_i(\Gamma' \Rightarrow \Delta') \iff \Gamma^i \cup \Gamma'^i \subseteq \Gamma'.\]

Obviously, those \( R_i \)'s are reflexive, but not necessarily transitive. Finally,

\[(\Gamma \Rightarrow \Delta) \models S \iff S \in \Gamma \text{ for a propositional letter } S.\]

**Lemma 5 (Truth Lemma).**

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

**Proof.** It is established by standard induction on \( \varphi \). 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 \( \Gamma^i \cup \Gamma'^i \subseteq \Gamma'; \) hence \( K_i \psi \in \Gamma' \). By the corresponding saturation property, \( \psi \in \Gamma' \). By the Induction Hypothesis, \( (\Gamma' \Rightarrow \Delta') \models \psi; \) hence \( (\Gamma \Rightarrow \Delta) \not\models K_i \psi. \)

Now let \( K_i \psi \in \Delta \). Then \( \Gamma^i \vdash \psi \) is a consistent sequent, otherwise \( \Gamma^i \vdash \psi \) would be derivable in \( S4^J_n \) without cut. By the \((\rightarrow, \Box)-\)rule, \( \Gamma^i \vdash \varphi \) would also be derivable in \( S4^J_n \) without cut. Hence, by Weakening, \( \Gamma \Rightarrow \Delta \) is derivable in \( S4^J_n \) without cut, which contradicts our assumption of the consistency of \( \Gamma \Rightarrow \Delta \). Consider a saturated extension \( \Gamma' \Rightarrow \Delta' \) of \( \Gamma^i \vdash \psi \). Since \( \psi \in \Delta' \), by the Induction Hypothesis, \( (\Gamma' \Rightarrow \Delta') \not\models \psi. \) Obviously, \( (\Gamma \Rightarrow \Delta) \) is accessible from \( (\Gamma \Rightarrow \Delta) \) by \( R_i \); hence \( (\Gamma \Rightarrow \Delta) \not\models K_i \psi. \) For \( T_n^J \) it suffices to take a consistent sequent \( \Gamma^i \vdash \psi \) instead of \( \Gamma^i, \Gamma \Rightarrow \Delta \).

Suppose \( \varphi = J \psi \). If \( J \psi \in \Gamma \), and \( \Gamma' \Rightarrow \Delta' \) is accessible from \( \Gamma \Rightarrow \Delta \) by \( R_i \), then \( \Gamma^i \cup \Gamma'^i \subseteq \Gamma'; \) hence \( J \psi \in \Gamma' \). By the corresponding saturation property, \( \psi \in \Gamma' \). By the Induction Hypothesis, \( (\Gamma' \Rightarrow \Delta') \models \psi; \) hence \( (\Gamma \Rightarrow \Delta) \not\models J \psi. \)

Let \( J \psi \in \Delta \). Then \( \Gamma^i \vdash \psi \) is a consistent sequent, otherwise \( \Gamma^i \vdash \psi \) would be derivable in \( S4^J_n \) without cut. By the \((\rightarrow, \Box)-\)rule, \( \Gamma^i \vdash J \psi \) would also be derivable in \( S4^J_n \) without cut; hence \( \Gamma \Rightarrow \Delta \) would be inconsistent. Consider a saturated extension \( \Gamma' \Rightarrow \Delta' \) of \( \Gamma^i \vdash \psi \). Since \( \psi \in \Delta' \), by the Induction Hypothesis, \( (\Gamma' \Rightarrow \Delta') \not\models \psi \). Since \( (\Gamma \Rightarrow \Delta) \) is accessible from \( (\Gamma \Rightarrow \Delta) \) by \( R_i \), \( (\Gamma \Rightarrow \Delta) \not\models J \psi. \) \qed

Here is the standard conclusion of the proof of Theorem 3. Let \( \Gamma \Rightarrow \Delta \) be a sequent not provable in \( S4^J_n \) 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 \( \Gamma \) hold at \( (\Phi \Rightarrow \Psi) \) and all formulas from \( \Delta \) do not hold at \( (\Phi \Rightarrow \Psi) \). Hence \( (\Phi \Rightarrow \Psi) \not\models \bigwedge \Gamma \rightarrow \vee \Delta \). \qed

**Corollary 1.**

1. Cut-elimination in \( S4^J_n \) and \( T_n^J \).
2. Completeness of \( S4^J_n \) for \( S4^J_n \)-models and \( T_n^J \) for \( T_n^J \)-models.
3. Finite model property of \( S4^J_n \) and \( T_n^J \).
4. Decidability of \( S4^J_n \) and \( T_n^J \).
5. Equivalence of \( S4^J_n \) to \( S4^J_n \) and \( T_n^J \) to \( T_n^J \) (Theorem 2).

Now we are ready to show that \( T_n^J \) and \( S4^J_n \) are exactly the forgetful projections of \( T_n \) and \( S4_n \), respectively, defined by a translation \( (\cdot)^\circ \) which maps \( r \varphi \) to \( J \varphi \) and commutes with all other connectives.

**Proposition 8.** \( (T_n \) \( \varphi \) \( )^\circ \subseteq \) \( T_n^J \) and \( (S4_n \) \( \varphi \) \( )^\circ \subseteq \) \( S4_n^J \).

**Proof.** A straightforward induction on derivations in \( T_n \) and \( S4_n \). It suffices to observe that the forgetful translations of all axioms and rules of \( T_n \) and \( S4_n \) are \( T_n^J \) and \( S4_n^J \)-compliant, respectively. \qed

The converse claim that \( T_n^J \subseteq (T_n \) \( \varphi \) \( )^\circ \) and \( S4_n^J \subseteq (S4_n \) \( \varphi \) \( )^\circ \) is a much trickier.

**Theorem 4 (Realization Theorem).** There is an algorithm that given a \( T_n^J \)-derivation (\( S4_n^J \)-derivation) of a formula \( \varphi \), returns a \( T_n \)-derivation (\( S4_n \)-derivation) of a formula \( \psi \) such that \( (\psi)^\circ = \varphi \).
Proof. First, find a cut-free proof of a given formula in $S4^J_n(T^J_nG)$. Then run the realizability algorithm from [6], 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 $S4^J_n$ only; the case of $T^J_n$ 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 evidence variables. We will speak about a sequent’s $\Gamma \Rightarrow \Delta$ being derivable in $S4^J_n$, meaning $S4^J_n \vdash \Gamma \Rightarrow \Delta$ or, equivalently, $S4^J_nG \vdash \Gamma \Rightarrow \Delta$. Moreover, since $S4^J_n$ enjoys the deduction theorem, $S4^J_n$ derives $\Gamma \Rightarrow \varphi$ iff $\Gamma \vdash \varphi$ in $S4^J_n$, iff $S4^J_nG \vdash \Gamma \Rightarrow \varphi$.

Consider a cut-free derivation $T$ of a sequent $\Rightarrow \varphi$ in $S4^J_nG$. It suffices now to construct a normal realization $r$ such that $S4_nLP \vdash \Gamma \Rightarrow \Delta$ for any sequent $\Gamma \Rightarrow \Delta$ in $T$. Note that in $T$, the rules respect polarities; all occurrences of $J$ introduced by $\iff, \Box$ 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 $T$ are naturally split into disjoint families of related ones. We call a family essential if it contains at least one instance of the ($\iff, J$) 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 evidence variables as provisional variables.

Step 1: For every negative family and non-essential positive family, we replace all occurrences of $J \varphi$ by ‘$x: \varphi$’ for a fresh evidence variable $x$.

Step 2: Pick an essential family $f$, enumerate all the occurrences of rules ($\iff, J$) 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 evidence term

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

where $v_i$’s are fresh provisional variables. The resulting tree $T'$ is labelled by $S4_nLP$-formulas, since all occurrences of the kind $J \varphi$ in $T$ are replaced by $t: \varphi$ for corresponding evidence terms $t$.

Step 3: Replace the provisional variables by evidence terms as follows. Proceed from the leaves of the tree to its root. By induction on the depth of a node in $T'$ we establish that after the process passes a node, the sequent assigned to this node becomes derivable in $S4_nLP$. The axioms $S \Rightarrow S$ and $\bot \Rightarrow \bot$ are derivable in $S4_nLP$. For every rule other than ($\iff, J$), we do not change the realization of formulas and just establish that the concluding sequent is provable in $S4_nLP$, given that the premises are. It is clear that every move down in the tree $T'$ other than ($\iff, J$) is derivable in $S4_nLP$. Let an occurrence of the rule ($\iff, J$) have number $i$ in the numbering of all rules ($\iff, J$) from a given family $f$. The corresponding node in $T'$ is labelled by

$$y_1:B_1, \ldots, y_k:B_k \Rightarrow B$$

where $y_1, \ldots, y_k$ are evidence variables, $u_1, \ldots, u_{n_f}$ are evidence terms, 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 $S4_nLP$. By the Lifting Lemma (Proposition 2), construct an evidence term $t(y_1, \ldots, y_k)$ such that

$$S4_nLP \vdash y_1:B_1, \ldots, y_k:B_k \Rightarrow t(y_1, \ldots, y_k):B.$$  

Since

$$S4_nLP \vdash t:B \Rightarrow (u_1 + \cdots + u_{i-1} + t + u_{i+1} + \cdots + u_{n_f}):B,$$

we have

$$S4_nLP \vdash y_1:B_1, \ldots, y_k:B_k \Rightarrow (u_1 + \cdots + u_{i-1} + t + u_{i+1} + \cdots + u_{n_f}):B.$$  

Now substitute $t(y_1, \ldots, y_k)$ for $u_i$ everywhere in $T'$ (and the corresponding constant specification 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 $T'$. The conclusion of the given rule ($\iff, J$) becomes derivable in $S4_nLP$, and the induction step is complete.

We eventually substitute terms of non-provisional variables for all provisional variables in $T'$ and establish that the root sequent of $T'$ is derivable in $S4_nLP$. The realization $r$ built by this procedure is normal. □
Note that the current version of the realization algorithm can produce evidence terms which are exponential in the size of the original cut-free derivation in $S_{4}^{f}G$. A more efficient modification of the realization algorithm for $S_{4}$ was described by Brezhnev and Kuznets in [15], where the realizing evidence terms are quadratic in the size of the original cut-free derivation. It seems a similar approach could produce an efficient realization algorithm for $S_{4}^{f}G$ and $T_{n}^{f}G$.

Realization for $SS_{5}^{f}$ needs a separate treatment.

**Definition 7.** $SS_{5}^{f}$-models are the models from Definition 5 with reflexive, transitive, and symmetric relations $R_{1}, \ldots, R_{n}$.

**Theorem 5.** $SS_{5}^{f}$ is sound and complete with respect to $SS_{5}^{f}$-models.

**Proof.** The soundness portion is straightforward. In particular, $J \varphi \rightarrow K_{i} \varphi$ is trivially guaranteed by $R_{i} \subseteq R$.

The completeness part is accomplished by the standard maximal consistent set construction. A set $\Gamma$ is consistent if for any finite $A \subseteq \Gamma$, $S_{5}^{f} \not\vdash (\bigwedge A)$. $W$ is a collection of all maximal consistent sets; $\Gamma R_{i} \Delta$ iff $\Gamma^{i} \subseteq \Delta$, $\Gamma R \Delta$ iff $\Gamma^{\sharp} \subseteq \Delta$, where $\Gamma^{\sharp} = \{ \varphi | J \varphi \in \Gamma \}$. 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 $SS_{5}^{f} \vdash J \varphi \rightarrow K_{i} J \varphi$, $J \varphi \rightarrow K_{i} J \varphi \in \Gamma$ 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 the $SS$-axiom $\neg K_{i} \varphi \rightarrow K_{i} \neg K_{i} \varphi$, $\neg K_{i} \varphi \in \Gamma$. Since $\Gamma^{\sharp} \subseteq \Delta$, $\neg K_{i} \varphi \in \Delta$ 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 $SS_{5}^{f}$-model.

The Truth Lemma says that for any formula $\varphi$,

$\Gamma \vdash \varphi$ if $\varphi \in \Gamma$.

Its proof follows from standard induction on $\varphi$. Let us check the case when $\varphi = J \psi$. If $J \psi \in \Gamma$, then $\psi \in \Delta$ for all $\Delta$ such that $\Gamma R \Delta$. By the Induction Hypothesis, $\Delta \vdash \psi$ for all $\Delta$ 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 $\Theta$ of $\Gamma$, $\Theta^{\sharp} \vdash \psi$ and, by modal logic rules, $\Theta \vdash J \psi$; hence $J \psi \in \Gamma$—a contradiction. Take a maximal consistent set $\Delta$ containing $\Gamma^{\sharp} \cup \{ \neg \psi \}$. Apparently, $\psi \notin \Delta$; hence by the Induction Hypothesis, $\Delta \vdash \psi$ and $\Gamma \vdash J \psi$.

Theorem 5 now follows immediately. □

**Theorem 6.** $SS_{5}^{f}$ is the forgetful projection of $SS_{5} LP$, i.e., $(SS_{5} LP)^{o} = SS_{5}^{f}$.

**Proof.** Again, the proof of $(SS_{5} LP)^{o} \subseteq SS_{5}^{f}$ is given by a straightforward induction on derivations in $SS_{5} LP$.

The existence of an $SS_{5} LP$-realization of any theorems of $SS_{5}^{f}$ can be established semantically by methods developed in [18]. The main ingredients of Fitting’s semantical realizability proof are the Fully Explanatory property of $SS_{5} LP$-models with full constant specifications (Proposition 5) and the Compactness property (Proposition 4).

**Definition 8.** By $SS_{5} LP^{-}$ we mean a system $SS_{5} LP$ in a language without ‘+’ and without axiom E3. Models of $SS_{5} LP^{-}$ are the same as for those of $SS_{5} LP$ except that the evidence function is not required to satisfy the $Sum$ condition. We may assume that $SS_{5} LP$-models and $SS_{5} LP^{-}$-models are also models for $SS_{5}^{f}$ 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 $SS_{5} LP^{-}$ and $SS_{5} LP^{-}$-models as well.

Assume an $SS_{5}^{f}$-formula $\varphi$ is fixed for the rest of the proof of Theorem 6. By ‘subformula of $\varphi$’ we will mean an ‘occurrence of a subformula of $\varphi$.’

**Definition 9.** Let $A$ be any assignment of evidence variables to subformulas of $\varphi$ of the form $J \psi$ that are in a negative position. We define two mappings $w_{A}$ and $v_{A}$ of subformulas of $\varphi$ to sets of formulas of $SS_{5} LP$.
and S5LP, respectively.

1. If $P$ is an atomic formula (including $\bot$), then $w_A(P) = v_A(P) = \{P\}$.

2. $w_A(X \rightarrow Y) = \{X' \rightarrow Y' | X' \in w_A(X) \text{ and } Y' \in w_A(Y)\}$.

3. If $K_iX$ is a negative subformula of $\phi$, then $w_A(K_iX) = \{K_iX' | X' \in w_A(X)\}$, $v_A(K_iX) = \{K_iX' | X' \in v_A(X)\}$.

4. If $K_iX$ is a positive subformula of $\phi$, then $w_A(K_iX) = \{K_iX' | X' \in w_A(X)\}$, $v_A(K_iX) = \{K_iX' | X' \in v_A(X)\}$.

5. If $JX$ is a negative subformula of $\phi$, then $w_A(JX) = \{x:X' | A(JX) = x \text{ and } X' \in w_A(X)\}$, $v_A(JX) = \{x:X' | A(JX) = x \text{ and } X' \in v_A(X)\}$.

6. If $JX$ is a positive subformula of $\phi$, then $w_A(JX) = \{t:X' | x \text{ and } X' \in w_A(X)\}$ and $t$ is any evidence term, $v_A(JX) = \{t(X_1 \lor \cdots \lor X_k) | X_i \in v_A(X)\}$ and $t$ is any evidence term.

By $\neg v_A(X)$ we mean $\{X' | X' \notin v_A(X)\}$.

Lemma 6. Let $CS$ be a full constant specification of $S5LP$ and $M$ be a canonical model for $S5LP$ that meets $CS$. Then for each world $\Gamma$ of $M$:

1. If $\psi$ is a positive subformula of $\phi$, then $\Gamma \models \neg v_A(\psi)$ yields $\Gamma \models \neg \psi$.

2. If $\psi$ is a negative subformula of $\phi$, then $\Gamma \models v_A(\psi)$ yields $\Gamma \models \psi$.

Proof. Induction on $\psi$. The atomic case as well as the cases of Boolean connectives are straightforward (cf. [18, Proposition 7.7]).

Suppose $\psi$ is $K_iX$, $\psi$ is a positive subformula of $\phi$, $\Gamma \models \neg v_A(K_iX)$, and the result is known for $X$ (which also occurs positively in $\phi$). We show that $\Gamma^{\sharp_i} \cup \neg v_A(X)$ is consistent. Indeed, otherwise $\Gamma^{\sharp_i} \models X_1 \lor \cdots \lor X_k$ for some $X_1, \ldots, X_k \in v_A(X)$. By the $K_i$-necessitation rule, $\Gamma \models K_i(X_1 \lor \cdots \lor X_k)$. Hence $\Gamma \models K_i(X_1 \lor \cdots \lor X_k)$, which is impossible since $K_i(X_1 \lor \cdots \lor X_k) \in v_A(K_iX)$. Now, extend $\Gamma^{\sharp_i} \cup \neg v_A(X)$ to a maximal consistent $A$, which is therefore a world in $M$ accessible from $\Gamma$ by $R_i$. Since $\neg v_A(X) \subseteq A$, $\Delta \models \neg v_A(X)$. By the Induction Hypothesis, $\Delta \models \neg X$. Therefore, $\Gamma \models \neg K_iX$.

Suppose $\psi$ is $K_iX$, $\psi$ is a negative subformula of $\phi$, $\Gamma \models v_A(K_iX)$, and the result is known for $X$ (which also occurs negatively in $\phi$). In particular, $\Gamma \models K_iX'$ for each $X' \in v_A(X)$. Let $\Delta$ be an arbitrary world such that $\Gamma R_i \Delta$. Then $\Delta \models \neg X'$; hence $\Delta \models v_A(X)$. By the Induction Hypothesis, $\Delta \models X$. Therefore, $\Gamma \models K_iX$.

Suppose $\psi$ is $JX$, $\psi$ is a positive subformula of $\phi$, $\Gamma \models \neg v_A(JX)$, and the result is known for $X$ (which also occurs positively in $\phi$). 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 $S5LP \vdash Y_1 \rightarrow (Y_2 \rightarrow \cdots \rightarrow (Y_m \rightarrow X_1 \lor \cdots \lor X_k) \ldots)$. By internalization, there is an evidence term $s$ such that $S5LP \vdash s:[Y_1 \rightarrow (Y_2 \rightarrow \cdots \rightarrow (Y_m \rightarrow X_1 \lor \cdots \lor X_k) \ldots)]$.

Consider evidence terms $t_1, t_2, \ldots, t_m$ such that $t_1Y_1, t_2Y_2, \ldots, t_mY_m \in \Gamma$. By El and propositional reasoning,

$S5LP \vdash t_1:Y_1 \land t_2:Y_2 \land \cdots \land t_m:Y_m \rightarrow (st_1t_2 \cdots t_m)(X_1 \lor \cdots \lor X_k)$.

Hence $\Gamma \models (st_1t_2 \cdots t_m)(X_1 \lor \cdots \lor X_k)$, which is impossible since $(st_1t_2 \cdots t_m)(X_1 \lor \cdots \lor X_k) \in v_A(JX)$.

Let $\Delta$ be a maximal consistent extension of $\Gamma^{\flat} \cup \neg v_A(X)$. Obviously, $\Gamma R \Delta$ and $\Delta \models \neg v_A(X)$. By the Induction Hypothesis, $\Delta \models X$; hence $\Gamma \models JX$. Therefore, $\Gamma \models JX$. □
Now suppose $S_n^J \vdash \varphi$ but $S_n^{\text{LP}} \not\vdash (\varphi_1 \lor \cdots \lor \varphi_m)$ for all $\varphi_1, \ldots, \varphi_m \in v_A(\varphi)$ with a given full constant specification $CS$. Then every finite subset of $\neg v_A(\varphi)$ is satisfiable. By compactness (Proposition 4) adapted to $S_n^{\text{LP}}$, there is a world $I$ in the canonical model for $S_n^{\text{LP}}$ with $CS$ such that $I \models \neg v_A(\varphi)$. By Lemma 6, $I \models \neg \varphi$. Therefore, since $S_n^J \vdash \varphi$, there are $\varphi_1, \ldots, \varphi_m \in v_A(\varphi)$ such that $S_n^{\text{LP}} \vdash (\varphi_1 \lor \cdots \lor \varphi_m)$.

**Lemma 7.** For every subformula $\psi$ of $\varphi$ and each $\psi_1, \ldots, \psi_m \in v_A(\psi)$, there is a substitution $\sigma$ of evidence terms for evidence variables and a formula $\psi' \in w_A(\psi)$ such that:

1. If $\psi$ is a positive subformula of $\varphi$, $S_n^{\text{LP}} \vdash (\psi_1 \lor \cdots \lor \psi_m) \sigma \rightarrow \psi'$.
2. If $\psi$ is a negative subformula of $\varphi$, $S_n^{\text{LP}} \vdash \psi' \rightarrow (\psi_1 \land \cdots \land \psi_m) \sigma$.

**Proof.** We use the fact that evidence variables assigned to different (occurrences of) subformulas $J\psi$ in $\varphi$ are all different. Induction on $\psi$. Again, the atomic case as well as the cases of Boolean connectives are straightforward (cf. [18, Proposition 7.8]).

Suppose $\psi$ is $K_iX$, $\psi$ is a positive subformula of $\varphi$, and the result is known for $X$ (which also occurs positively in $\varphi$). Let $K_1D_1, \ldots, K_mD_m \in v_A(K_iX)$. Those $D_1, \ldots, D_m$ are disjunctions of formulas from $v_A(X)$. By the Induction Hypothesis, there is a substitution $\sigma$ and $X' \in w_A(X)$ such that $S_n^{\text{LP}} \vdash (D_1 \lor \cdots \lor D_m) \sigma \rightarrow X'$. Consequently, for each $j = 1, \ldots, m$, $S_n^{\text{LP}} \vdash D_j \sigma \rightarrow X'$. By necessitation, $S_n^{\text{LP}} \vdash K_i(D_j \sigma \rightarrow X')$; hence $S_n^{\text{LP}} \vdash K_iD_j \sigma \rightarrow K_iX'$. Therefore, $S_n^{\text{LP}} \vdash (K_iD_1 \lor \cdots \lor K_iD_m \sigma) \rightarrow K_iX'$.

Suppose $\psi$ is $JX$, $\psi$ is a negative subformula of $\varphi$, and the result is known for $X$ (which also occurs negatively in $\varphi$). Let $K_1X_1, \ldots, K_mX_m \in v_A(K_iX)$. By the Induction Hypothesis, there is a substitution $\sigma$ and $X' \in w_A(X)$ such that $S_n^{\text{LP}} \vdash (X_1 \land \cdots \land X_m) \sigma$. By necessitation, $S_n^{\text{LP}} \vdash K_iX' \rightarrow K_i(X_1 \land \cdots \land X_m) \sigma$. Since $K_i$ commutes with $\sigma$ and $\land$, $S_n^{\text{LP}} \vdash K_iX' \rightarrow (K_iX_1 \land \cdots \land K_iX_m) \sigma$.

Suppose $\psi$ is $JX$, $\psi$ is a positive subformula of $\varphi$, and the result is known for $X$ (which also occurs positively in $\psi$). 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 $\sigma$ and $X' \in w_A(X)$ such that $S_n^{\text{LP}} \vdash (D_1 \lor \cdots \lor D_m) \sigma \rightarrow X'$.

Consequently, for each $j = 1, \ldots, m$, $S_n^{\text{LP}} \vdash D_j \sigma \rightarrow X'$. By internalization, there is an evidence term $s_j$ such that $S_n^{\text{LP}} \vdash s_j; (D_j \sigma \rightarrow X')$. Then $S_n^{\text{LP}} \vdash (t_j; D_j) \sigma \rightarrow (s_j; t_j) \sigma \rightarrow X'$. Set $t = (s_1; t_1) \sigma + \cdots + (s_m; t_m) \sigma$. We have $S_n^{\text{LP}} \vdash (t; D_j) \sigma \rightarrow t; X'$, and hence $S_n^{\text{LP}} \vdash (t_1; D_1 \lor \cdots \lor t_m; D_m) \sigma \rightarrow t; X'$.

Suppose $\psi$ is $JX$, $\psi$ is a negative subformula of $\varphi$, and the result is known for $X$ (which also occurs negatively in $\varphi$). 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 $\sigma$ and $X' \in w_A(X)$ such that $S_n^{\text{LP}} \vdash X' \rightarrow (X_1 \land \cdots \land X_m) \sigma$. 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 $\sigma$. From the above, it follows that $S_n^{\text{LP}} \vdash X' \rightarrow X_j \sigma$. By internalization, $S_n^{\text{LP}} \vdash t_j; (X' \rightarrow X_j) \sigma$ for some evidence term $t_j$. Therefore, $S_n^{\text{LP}} \vdash s; (X' \rightarrow X_j) \sigma$ for $s = t_1 \lor \cdots \lor t_m$. Furthermore, $S_n^{\text{LP}} \vdash x; (X' \rightarrow (s; x)) (X_j) \sigma$ for each $j = 1, \ldots, m$. For the substitution $\sigma' = \sigma \cup \{x/(s; x)\}$, $S_n^{\text{LP}} \vdash x; (x: X_1 \land \cdots \land x: X_m) \sigma'$, which completes the proof of Lemma 7.

To conclude the proof of Theorem 6, assume that $S_n^J \vdash \varphi$. Then there are $\varphi_1, \ldots, \varphi_m \in v_A(\varphi)$ such that $S_n^{\text{LP}} \vdash \varphi_1 \lor \cdots \lor \varphi_m$. By Lemma 7, there is a substitution $\sigma$ and $\varphi' \in w_A(\varphi)$ such that $S_n^{\text{LP}} \vdash (\varphi_1 \lor \cdots \lor \varphi_m) \sigma \rightarrow \varphi'$. Since $S_n^{\text{LP}}$ is closed under substitution, $S_n^{\text{LP}} \vdash \varphi'$. □

Theorem 6 yields an algorithm that given $S_n^J$-theorem $\varphi$, retrieves a $S_n^{\text{LP}}$-theorem $\psi$ such that $(\psi)^{\sigma} = \varphi$. Indeed, arrange an enumeration of all $S_n^{\text{LP}}$-realizations of $\varphi$ and their proof searches in $S_n^{\text{LP}}$. By Theorem 6, this process should terminate with success. A question of finding an efficient realization algorithm for $S_n^J$ remains open.

The results of this section show that the justified knowledge $J$ in $JCK$-systems $T_n^J$, $S_n^J$, and $S_n^J$ is indeed the forgetful version of evidence-based knowledge in the corresponding $EBK$-systems. Using $JCK$-systems (i.e., 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 $JCK$-systems. In particular, in light of [3] this provides an $EBK$-semantics for the McCarthy’s ‘any fool’ agent from [32].
Note that models for $T_n,L_P$, $S_4_n,L_P$, and $S_5_n,L_P$ are also models for $T_n$, $S_4_n$, and $S_5_n$, respectively. It suffices to regard the evidence accessibility relation $R$ in models for $T_n,L_P$, $S_4_n,L_P$, and $S_5_n,L_P$ as the accessibility relation for $J$.

7. Justified vs. common knowledge

In this section, we compare justified knowledge systems and common knowledge systems. First of all, we recall that the justified knowledge part in $JCK$-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, justified common knowledge systems cover more situations than the common knowledge systems. When both systems are present, e.g., in the case of $S_4^C_n$ and $S_4^J_n$, it is fair to compare them.

Operators $C$ and $J$ can be compared model-theoretically. Each $S_4^C_n$ model is an $S_4^J_n$ model, but not the other way around, since the evidence accessibility in $S_4^J_n$-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 $S_4^C_n$-model on any $S_4^J_n$-model by adding the reachability relation for the operator $C$, which is done in a unique way for a given $S_4^J_n$-model. The resulting models $M$ support the languages of both $S_4^C_n$ and $S_4^J_n$, thus providing a reasonable context for comparing knowledge operators $C$ and $J$. The logic $S_4^J_n$ is the set of tautologies in the language containing $K_1, \ldots, K_n, J, C$.

**Proposition 9.** Justified common 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 $W = \{a, b\}$, $R_i = \{(a, a), (a, b), (b, b)\}$, $R_j = R_i \cup \{(b, a)\}$. Then the transitive closure of all $R_i$ will be the same $R_j$. 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 $b \not\vDash J(S)$. $\square$

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 $C \varphi \leftrightarrow E(\varphi \land C \varphi)$, 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 10.** Each $JCK$-principle is a common knowledge principle but not vice versa i.e., $(S_4^J_n)^* \subsetneq S_4^C_n$.

**Proof.** For $(S_4^J_n)^* \subseteq S_4^C_n$, it suffices to prove the $*$-translations of all the axioms and rules of $S_4^J_n$ in $S_4^C_n$. Let us check, for example, the necessitation rule for $J$: $S_4^J_n \vdash \psi \Rightarrow S_4^J_n \vdash J \psi$. Suppose $S_4^C_n \vdash \psi^*$, then $S_4^C_n \vdash \top \rightarrow E(\psi^*)$. Use the Induction Rule of $S_4^C_n$ (cf. [17]) to conclude that $S_4^C_n \vdash C \psi^*$, i.e., $S_4^C_n \vdash (J \psi)^*$. The remaining cases can be recovered by inspecting [17].

To show the remaining part of the claim, consider a valid $S_4^C_n$ principle

$$\varphi \land C(\varphi \rightarrow E \varphi) \rightarrow C \varphi,$$

and notice that its $J$ version $IP = \varphi \land J(\varphi \rightarrow E \varphi) \rightarrow J \varphi$ is not valid for $S_4^J_n$. Indeed, consider the same model as in the proof of Proposition 9.2, and pick $\varphi$ such that $a \not\vDash \varphi$, but $b \vDash \varphi$. Then $b \vDash J(\varphi \rightarrow E \varphi)$, since at each node where $\varphi$ 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 $b \vDash \varphi \land J(\varphi \rightarrow E \varphi)$. On the other hand, $a \not\vDash \varphi$, $bRa$; hence $b \not\vDash J \varphi$. $\square$

8. Wise Men puzzle via justified knowledge

In this section, we present a small example of using justified knowledge. Namely, we give a solution to the Wise Men puzzle (cf. [17, p. 12]) as a derivation in the justified knowledge system $T_3^J$. 
There are certain advantages in such kinds of solutions. A high degree of formalization eliminates hidden semantical assumptions typical for a traditional solution of this type of problem by reasoning on a designated Kripke model. The deductive solution to the Wise Men puzzle below is carried within $T^f_3$ which assumes the minimal $T$ as the logic of knowledge for agents, and modest $S_4$ as the justification system, whereas the traditional solutions routinely use much more involved and problematic $S_5$ as agent’s knowledge and $S_5$ extended by the Induction Rule for the Common Knowledge.

The Wise Men story is 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?

The solution below consists of a formal derivation in $T^f_3$. Let atomic propositions $p_i$ stand for ‘wise man $i$ has a red hat’ ($i = 1, 2, 3$). Let also $Kw_i \varphi$ be a shorthand for $K_1 \varphi \vee K_2 \neg \varphi$, i.e., ‘$i$ knows whether $\varphi$.’ The assumption that each wise man observes the other wise men’s hats is represented by the additional axioms ‘Knowing About the Others,’ or ‘K.A.O.’ for short:

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

The rules of the game can be described by the theory

$$L(0) = T^f_3 + J(K.A.O.) + J(\neg 000).$$

The situation after the first and second wise men said they did not know is represented by a theory

$$L(2) = L(0) + J(\neg Kw_1 p_1) + J(\neg Kw_2 p_2).$$

Lemma 8. $L(2) \vdash J p_3.$

Proof. We now have a comfortable choice of methods ranging from model reasoning to all sorts of proof systems, including Gentzen-style ones. We will present a concise Hilbert-style derivation in $L(2)$ which directly formalizes a natural, ‘human’ solution. First, we prove in $L(2)$ that $J(\neg 100)$.

1. $100 \rightarrow K_1(\neg p_2 \land \neg p_3)$, from $J(K.A.O.)$;
2. $K_1(\neg p_2 \land \neg p_3) \rightarrow K_1 p_1$, from $J(\neg 000)$;
3. $100 \rightarrow K_1 p_1$, from 1 and 2;
4. $\neg Kw_1 p_1 \rightarrow \neg 100$, from 3;
5. $J(\neg Kw_1 p_1) \rightarrow J(\neg 100)$, from 4, by $T^f_3$-reasoning;
6. $J(\neg 100)$, from $J(\neg Kw_1 p_1)$ and 5.

Likewise, using $J(\neg Kw_2 p_2)$ we obtain $J(\neg 010).$ Next, we prove in $L(2)$ that $J(\neg 110)$.

1. $110 \rightarrow K_2(110 \lor 100)$, from $J(K.A.O.)$;
2. $110 \rightarrow K_2(110)$, since $J(\neg 100)$;
3. $110 \rightarrow Kw_2 p_2$, by $T^f_3$-reasoning;
4. $\neg Kw_2 p_2 \rightarrow \neg 110$, by propositional logic;
5. $J(\neg Kw_2 p_2) \rightarrow J(\neg 110)$, by $T^f_3$-reasoning;
6. $J(\neg 110)$, from $J(\neg Kw_2 p_2)$ and 5.

Finally, we conclude $J p_3$, since it is common knowledge that all truth combinations with $\neg p_3$ have been ruled out. □

From the above proof it is also clear that the 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 use $Kw_3 p_1$ and $Kw_3 p_2$. 


Acknowledgments

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

References

[38] N.M. Rubtsova, Evidence-based knowledge for S5, Logic Colloq. 2005, accepted for publication.