

Basic systems of epistemic logic with justification

Sergei Artemov

Graduate Center CUNY

Ph.D. Program in Computer Science

365 Fifth Avenue

New York, NY 10016, U.S.A.

SArtemov@gc.cuny.edu

Elena Nogina

BMCC CUNY

Department of Mathematics

199 Chambers Street

New York, NY 10007, U.S.A.

ENogina@bmcc.cuny.edu

February 21, 2005

Abstract

An issue of an epistemic logic with justification has been discussed since the early 1990s. Such a logic, along with the usual knowledge operator $\Box F$ (*F is known*), should contain assertions $t:F$ (*t is a justification for F*), which gives a more nuanced and realistic model of knowledge. In this paper, we build two systems of epistemic logic with justification: the minimal one—**S4LP**—which is an extension of the basic epistemic logic **S4** by an appropriate calculus of justification corresponding to the logic of proofs **LP**, and **S4LPN**—which is **S4LP** augmented by the explicit negative introspection principle $\neg(t:F) \rightarrow \Box\neg(t:F)$. Epistemic semantics for both systems are suggested. Completeness and specific properties of **S4LP** and **S4LPN**, reflecting the explicit character of those systems, are established.

1 Introduction

The need for a logic of knowledge with justifications has been discussed by van Benthem in [33]. Such a logic should contain assertions of the form $\Box F$ (*F is known*) along with those of the form $t:F$ (*t is a justification for F*) bringing together the implicit and explicit components of our informal notion of knowledge.

The explicit character of judgments significantly expands the expressive power of epistemic logics. The traditional Hintikka-style modal logic approach to knowledge has the well-known defect of *logical omniscience*, caused by an unrealistic stipulation that an agent knows all logical consequences of his/her assumptions ([15, 27, 30, 31]). Hence, the usual epistemic modality $\Box F$ should be regarded as “potential knowledge” or “knowability” (cf. [18]) rather than actual knowledge. Evidence operators $t:F$ provide a more nuanced and realistic model of knowledge. This new language enables us to formulate new logical

principles about knowledge. For example, in the context of mathematical provability, the modal principle of negative introspection $\neg\Box F \rightarrow \Box\neg\Box F$ is not valid. A purely explicit version of negative introspection $\neg(x:F) \rightarrow t(x):\neg(x:F)$ does not hold in the logic of proofs LP either. However, negative introspection in a mixed explicit-implicit form $\neg(t:F) \rightarrow \Box\neg(t:F)$ is valid in the provability semantics, providing a good reason for considering this principle in the context of logic of knowledge in general.

The ability of terms to encode the complexity of justifications could be useful in dealing with the logical omniscience problem, since an evidence term t in $t:F$ carries information about how hard it was to justify F from given assumptions.

In this paper, we derive basic principles of knowledge and justification from the laws of proofs and provability. The provability semantics is a representative case of the epistemic reading of modal logic, and as such sheds light on the logic of knowledge with justification in general.

The idea of the logic of proofs as an explicit counterpart of S4 first appeared in Gödel's [20]. The formal system LP of the logic of proofs was introduced in [3, 4]. LP describes all valid principles of proof operators $t:F$

$$t \text{ is a proof of } F \text{ in Peano arithmetic} \tag{1}$$

with an appropriate set of operations on proofs sufficient to realize the modal logic S4 explicitly [4]. A similar explicit counterpart of S5 was found in [8]. A semantical approach to the logic of proofs as a general calculus of evidence in the epistemic framework has been developed by Mkrtychev and Fitting in [16, 18, 26].

Joint logics of proofs and provability, studied in [2, 9, 28, 29, 32, 34], are of special interest for this paper since they serve as a prototype of the logic of knowledge with justification.

In this paper, we introduce and study basic epistemic logics with justification. We construct two systems. The basic one, S4LP, consists of S4 combined with LP as a calculus of justifications and the principle $t:F \rightarrow \Box F$ connecting implicit and explicit knowledge operators. S4LP may be regarded as the generic epistemic logic with justification where no specific assumptions are made concerning explicit knowledge. The other system, S4LPN, is S4LP augmented by the principle of *explicit negative introspection* $\neg(t:F) \rightarrow \Box\neg(t:F)$ which first came up in the logics of proofs and provability. Alternatively, S4LPN can be axiomatized over S4LP by the principle of *decidability of evidence assertions* $\Box t:F \vee \Box\neg(t:F)$.

In our technical report [9] we used Fitting models [16], originally designed for the logic of proofs LP, as a semantics for S4LP. In [17], Fitting showed that S4LP is complete with respect to this semantics. In [6], the first author augmented Fitting semantics by a new feature, an evidence accessibility relation. This more general class of models, *AF-models*, already covers all the above systems: LP, S4LP, and S4LPN (as well as a broad class of the so-called evidence-based common knowledge systems from [6]).

In this paper we give soundness and completeness theorems for S4LP and S4LPN with respect to AF-semantics. Furthermore, both S4LP and S4LPN are

shown to enjoy the arithmetical provability semantics when $\Box F$ is interpreted as the so-called *strong provability operator* (cf. [13]):

F is true and provable in Peano arithmetic .

S4LP and S4LPN provide a framework for reasoning about knowledge and justification and hence answer a question concerning such logics raised by van Benthem in [33]. Epistemic logics with justification and similar systems were used in recent work [6] on the common knowledge phenomenon.

2 The logic of proofs as a general calculus of justification

The logic of proofs LP was inspired by the classical works of the 1930s by Kolmogorov [23] and Gödel ([19, 20]) and found in [3, 4] (see also surveys [5, 7, 14]). LP naturally extends classical propositional logic by adding symbolically represented proofs into the language of the system. Internal proof terms in LP are called proof polynomials. A new formula formation rule is postulated, stating that $t:F$ is a formula whenever t is a proof polynomial and F is an arbitrary formula, hence the language of LP is a general propositional proof-carrying language. According to the completeness theorems from [3, 4], LP captures exactly the set of all valid logical principles concerning propositions and mathematical proofs with a fixed, sufficiently rich set of operations on proofs. Moreover, by the realization theorem ([3, 4]), proof polynomials suffice to recover the explicit provability content in all S4-theorems (and hence all intuitionistic propositional theorems) by realizing modalities in the latter with appropriate proof terms. In a more general setting LP may be regarded as a device that makes reasoning about knowledge explicit and keeps track of the evidence.

Here are some formal definitions.

Definition 1. *Proof polynomials* are terms built from *proof variables* x, y, z, \dots and *proof constants* a, b, c, \dots by means of three operations: *application* “.” (binary), *union* “+” (binary), and *proof checker* “!” (unary).

Definition 2. Using t to stand for any proof polynomial and S for any sentence variable, the formulas are defined by the grammar

$$A = S \mid A_1 \rightarrow A_2 \mid A_1 \wedge A_2 \mid A_1 \vee A_2 \mid \neg A \mid t.A \ .$$

We assume also that “ t .” and “ \neg ” bind stronger than “ \wedge ” and “ \vee ,” which bind stronger than “ \rightarrow .”

Definition 3. The *logic of proofs* LP has the following Hilbert-style axioms and rules:

I. The standard set of axioms A1-A10 from [22] (or a similar system)

- R1. *Modus Ponens*
- II. LP1. $s:(F \rightarrow G) \rightarrow (t:F \rightarrow (s \cdot t):G)$ (application)
 LP2. $t:F \rightarrow !t:(t:F)$ (proof checker)
 LP3. $s:F \rightarrow (s+t):F, \quad t:F \rightarrow (s+t):F$ (union)
 LP4. $t:F \rightarrow F$ (explicit reflexivity)
 R2. $\vdash c:A$, where A is an axiom from I - II and c is a proof constant
 (constant specification rule)

The principle LP1 specifies the basic operation of application: a justification of an implication $F \rightarrow G$ applied to any justification of the premise F returns a justification of the conclusion G . LP2 is the verifiability property of evidence: for any evidence t of a fact F , the result of applying a checker to t , $!t$, provides a justification of $t:F$. LP3 reflects the monotonicity principle: a justification for F remains a justification after adding any additional evidence. Finally, LP4 is the reflexivity property, which is provably valid.

A *constant specification* \mathcal{CS} is a set $\{c_1:A_1, c_2:A_2, \dots\}$ of formulas in which each A_i is an axiom from I-II and each c_i is a proof constant. By default, with each derivation in LP we associate a constant specification \mathcal{CS} that consists of formulas introduced in this derivation by the rule of *constant specification*. The claim that F is derivable in LP is equivalent to the existence of a derivation with a constant specification \mathcal{CS} associated with this derivation, i.e.:

$$F \text{ is derivable given } c_1:A_1, \dots, c_n:A_n .$$

LP is closed under substitutions of proof polynomials for proof variables and formulas for propositional variables, and LP enjoys the deduction theorem.

In addition to the arithmetical completeness theorem, LP enjoys two fundamental properties: *internalization* and *realizability*.

Proposition 1. (Internalization) *If $A_1, \dots, A_k \vdash F$ then for some proof polynomial $p(x_1, \dots, x_k)$*

$$x_1:A_1, \dots, x_k:A_k \vdash p(x_1, \dots, x_k):F .$$

Proposition 2. (Realizability) *There is an effective procedure that constructs a realization r , which substitutes proof polynomials for all modalities in a given S4-derivation of formula F and thereby produces formula F^r derivable in LP.*

The logic of proofs LP may be regarded as the explicit version of S4. A paper [8] introduced a variant of the logic of proofs corresponding to S5. Logics of proofs corresponding to the modal logics K, K4, D, D4, and T were described in [10, 11].

3 Basic epistemic logic with justification

We introduce the basic epistemic logic with justifications, S4LP, consisting of S4 as the “knowledge component” and LP as the “justification component” together with the principle $t:F \rightarrow \Box F$ connecting explicit and implicit knowledge.

Definition 4. *Proof polynomials* for S4LP are the same as proof polynomials for LP, i.e. they are terms built from *variables* x, y, z, \dots and *constants* a, b, c, \dots by means of three operations, *application* “.” (binary), *union* “+” (binary), and *evidence checker* “!” (unary). Formulas of the language of S4LP are defined by the grammar

$$A = S \mid A_1 \rightarrow A_2 \mid A_1 \wedge A_2 \mid A_1 \vee A_2 \mid \neg A \mid \Box A \mid t:A .$$

We assume also that “t:,” “□,” and “¬” bind stronger than “∧” and “∨,” which bind stronger than “→.”

Definition 5. The system S4LP has the following axioms and rules:

I. Classical propositional logic

- A1-A10. (the standard set of axioms, e.g., from [22])
 R1. *Modus ponens*

II. Logic of Proofs LP

- LP1. $s:(F \rightarrow G) \rightarrow (t:F \rightarrow (s \cdot t):G)$ (*application*)
 LP2. $t:F \rightarrow !t:(t:F)$ (*inspection*)
 LP3. $s:F \rightarrow (s+t):F, \quad t:F \rightarrow (s+t):F$ (*union*)
 LP4. $t:F \rightarrow F$ (*reflexivity of explicit knowledge*)
 R2. $\vdash c:A$, where A is an axiom from I-IV and c is a proof constant
 (*constant specification*)

III. Basic Epistemic Logic S4

- E1. $\Box(F \rightarrow G) \rightarrow (\Box F \rightarrow \Box G)$
 E2. $\Box F \rightarrow \Box \Box F$
 E3. $\Box F \rightarrow F$
 R3. $\vdash F \Rightarrow \vdash \Box F$

IV. Principle connecting explicit and implicit knowledge

- C1. $t:F \rightarrow \Box F$ (*explicit-implicit connection*)

Obviously, S4LP contains both LP and S4. The principle LP4 is redundant but we keep it listed for convenience. S4LP is closed under substitutions of proof polynomials for proof variables and formulas for sentence variables. S4LP also enjoys the deduction theorem.

Consider a *constant specification* $\mathcal{CS} = \{c_1:A_1, c_2:A_2, \dots\}$ (where each A_i is an axiom from I-IV and each c_i is a proof constant). By $\text{S4LP}_{\mathcal{CS}}$ we mean a subsystem of S4LP where R2 is restricted to producing formulas from a given \mathcal{CS} only. In particular, S4LP_\emptyset is the subsystem of S4LP without R2.

Lemma 1. *The principle of positive introspection*

$$t:F \rightarrow \Box t:F$$

is provable in S4LP_\emptyset (hence in $\text{S4LP}_{\mathcal{CS}}$ for any constant specification \mathcal{CS}).

Proof.

$$\begin{array}{ll}
t:F \rightarrow !t:(t:F) & \text{by LP2} \\
!t:(t:F) \rightarrow \Box t:F & \text{by C1} \\
t:F \rightarrow \Box t:F & \text{by propositional logic}
\end{array}$$

□

Lemma 2. $S4LP_{CS} \vdash F \Leftrightarrow S4LP_{\emptyset} \vdash \bigwedge CS \rightarrow F$.

Proof. The direction “ \Leftarrow ” is straightforward. “ \Rightarrow ” is proven by induction on the derivation of F in $S4LP_{CS}$. The only interesting case is the rule of necessitation R3. If F is obtained by the necessitation rule R3, i.e., F is $\Box G$ and $S4LP_{CS} \vdash G$, then by the induction hypothesis, $S4LP_{\emptyset} \vdash \bigwedge CS \rightarrow G$. By S4 reasoning,

$$S4LP_{\emptyset} \vdash \Box \bigwedge CS \rightarrow \Box G .$$

By positive introspection (Lemma 1) and some trivial S4 reasoning,

$$S4LP_{\emptyset} \vdash \bigwedge CS \rightarrow \Box \bigwedge CS ,$$

hence $S4LP_{\emptyset} \vdash \bigwedge CS \rightarrow F$.

□

Lemma 3. *For any formula F , there are proof polynomials $\text{up}_F(x)$ and $\text{down}_F(x)$ such that S4LP proves*

1. $x:F \rightarrow \text{up}_F(x):\Box F$
2. $x:\Box F \rightarrow \text{down}_F(x):F$

Proof.

1. $x:F \rightarrow \Box F$ by C1
 $a:(x:F \rightarrow \Box F)$ specifying constant a , by R2
 $!x:(x:F) \rightarrow (a!\cdot x):\Box F$ by LP1 and propositional logic
 $x:F \rightarrow !x:(x:F)$ by LP2
 $x:F \rightarrow (a!\cdot x):\Box F$ by propositional logic

It suffices now to set $\text{up}_F(x)$ to $a!\cdot x$ with $a:(x:F \rightarrow \Box F)$.

2. $\Box F \rightarrow F$ by E3
 $b:(\Box F \rightarrow F)$ specifying constant b , by R2
 $x:\Box F \rightarrow (b \cdot x):F$ by LP1 and propositional logic

It suffices now to set $\text{down}_F(x)$ to $b \cdot x$ with $b:(\Box F \rightarrow F)$.

□

Proposition 3. (Constructive necessitation) *If $S4LP \vdash F$, then $S4LP \vdash p:F$ for some proof polynomial p .*

Proof. Induction on a derivation of F . Base: F is an axiom. Then use constant specification rule. In this case, p is an arbitrary proof constant and $p:F$ is included in the constant specification corresponding to this derivation.

Induction step: Let F be obtained from $X \rightarrow F$ and X by *modus ponens*. By the induction hypothesis, $\vdash s:(X \rightarrow F)$ and $\vdash t:X$, hence by LP1, $\vdash (s \cdot t):F$ and hence p is $s \cdot t$. If F is obtained by R2, then F is $c:A$ for some constant c and axiom A . Use the axiom LP2 to derive $!c:(c:A)$, i.e., $!c:F$. Here p is $!c$. If F is obtained by R3, then $F = \Box G$ and $\vdash G$. By the induction hypothesis, $\vdash t:G$ for some proof polynomial t . Use Lemma 3.1 to conclude that $\vdash \mathbf{up}_G(t):\Box G$, and put $p = \mathbf{up}_G(t)$.

Note that the proof polynomial p is always a ground term and built from proof constants by applications and proof checker operations only. Moreover, the presented derivation of $p:F$ does not use rule R3. \square

The necessitation rule R3 is derivable from the rest of S4LP. Indeed, if $\vdash F$ then, by Proposition 3, $\vdash p:F$ for some proof polynomial p . By C1, $\vdash \Box F$. However, the rule of necessitation is not redundant in S4LP_{CS} for any finite constant specification \mathcal{CS} . To emulate the rule of necessitation one needs to apply constructive necessitation to the unbounded set of theorems of S4LP_{CS}, which requires an unbounded set of constant specifications.

The following property of S4LP is a generalization of constructive necessitation (Proposition 3). It is the explicit analogue of the rule

$$\frac{A_1, \dots, A_k, \Box B_1, \dots, \Box B_n \vdash F}{\Box A_1, \dots, \Box A_k, \Box B_1, \dots, \Box B_n \vdash \Box F}$$

which holds in any normal modal logic containing K4.

Proposition 4. (Lifting) *If $A_1, \dots, A_k, y_1:B_1, \dots, y_n:B_n \vdash F$, then for some proof polynomial $p(x_1, \dots, x_k, y_1, \dots, y_n)$*

$$x_1:A_1, \dots, x_k:A_k, y_1:B_1, \dots, y_n:B_n \vdash p(x_1, \dots, x_k, y_1, \dots, y_n):F .$$

Proof. Similar to Proposition 3 with two new base clauses. If F is A_i , then x_i can be taken as p . If F is $y_j:B_j$, then p is equal to $!y_j$. \square

Proposition 5. (Internalization) *If $A_1, \dots, A_k \vdash F$, then for some proof polynomial $p(x_1, \dots, x_k)$*

$$x_1:A_1, \dots, x_k:A_k \vdash p(x_1, \dots, x_k):F.$$

Proof. A special case of Proposition 4. \square

The internalization property states that any derivation in S4LP can be internalized as a proof polynomial and verified in S4LP itself.

Note that axiom C1 in S4LP can be replaced by the *explicit positive introspection* principle $t:F \rightarrow \Box t:F$. The new system will coincide with S4LP modulo replacement of some constants by ground proof polynomials.

4 Introducing explicit negative introspection

As was noticed earlier, the *explicit negative introspection* principle

$$\neg t:F \rightarrow \Box \neg t:F$$

holds when we interpret \Box as mathematical provability, thus suggesting it as an important epistemic principle.

Definition 6. The system S4LPN has the same syntax, axioms, and rules as S4LP with one additional axiom:

$$\text{C2. } \neg t:F \rightarrow \Box \neg t:F \quad (\text{explicit negative introspection})$$

S4LPN_{CS} is S4LPN with the rule R2 limited to a given constant specification \mathcal{CS} . S4LPN _{\emptyset} is S4LPN_{CS} with the empty constant specification.

Analogues of Lemmas 1, 2, and 3 as well as Propositions 3, 4, and 5 hold for S4LPN as well.

Lemma 4. *The principle of decidability of explicit knowledge*

$$\Box t:F \vee \Box \neg t:F$$

is provable in S4LPN (hence in S4LPN_{CS} for any constant specification \mathcal{CS}).

Proof.

$t:F \rightarrow \Box t:F$	positive introspection
$\neg t:F \rightarrow \Box \neg t:F$	by negative introspection
$(t:F \vee \neg t:F) \rightarrow (\Box t:F \vee \Box \neg t:F)$	by propositional logic
$\Box t:F \vee \Box \neg t:F$	by propositional logic

□

5 Models

Kripke-style models for modal logics with justification were introduced in [2] and then generalized in [9, 28, 29, 34]. A new type of models capturing evidence was developed in [16, 26] for the logic of proofs LP and in [9, 17] for S4LP. A general class of models covering all of the above and providing semantics for S4LPN as well as for the so-called evidence-based common knowledge systems was introduced in [6].

At the heart of this semantics lies the idea, which can be traced back to Mkrtychev and Fitting ([16, 26]), of augmenting Boolean or Kripke models with an evidence function, which assigns “admissible evidence” terms to a 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 φ in u ;

2) φ holds in all worlds accessible from u .

The second idea came from the paper [6] which introduced an “evidence accessibility” relation different from the knowledge accessibility relation, thus semantically separating explicit knowledge from usual knowledge.

A *frame* is a structure (W, R, R^e) , where W is a non-empty set of *states* (*possible worlds*), R is a binary *accessibility* relation on W , and R^e is a binary *evidence accessibility* relation on W . For our purposes, the relations R and R^e can be taken as reflexive and transitive. R^e should contain R but not necessarily coincide with R .

Given a frame (W, R, R^e) , a *possible evidence function* \mathcal{E} is a mapping from worlds and justification terms to sets of formulas. We can read $F \in \mathcal{E}(u, t)$ as “ F is one of the formulas for which t serves as possible evidence in world u .” An evidence function must obey conditions that respect the intended meanings of the operations on justification terms (i.e. proof polynomials).

Definition 7. \mathcal{E} is an evidence function on (W, R, R^e) if for all proof polynomials s and t , for all formulas F and G , and for all $u, v \in W$, each of the following hold:

1. *Monotonicity:* uR^ev implies $\mathcal{E}(u, t) \subseteq \mathcal{E}(v, t)$.
2. *Application:* $F \rightarrow G \in \mathcal{E}(u, s)$ and $F \in \mathcal{E}(u, t)$ implies $G \in \mathcal{E}(u, s \cdot t)$.
3. *Inspection:* $F \in \mathcal{E}(u, t)$ implies $t:F \in \mathcal{E}(u, !t)$.
4. *Sum:* $\mathcal{E}(u, s) \cup \mathcal{E}(u, t) \subseteq \mathcal{E}(u, s + t)$.

A model is a structure $\mathcal{M} = (W, R, R^e, \mathcal{E}, \Vdash)$ where (W, R, R^e) is a frame with an evidence function \mathcal{E} on (W, R, R^e) and \Vdash is an arbitrary mapping from sentence variables to subsets of W .

Given a model $\mathcal{M} = (W, R, R^e, \mathcal{E}, \Vdash)$, the 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 ($u \Vdash F \wedge G$ iff $u \Vdash F$ and $u \Vdash G$, $u \Vdash \neg F$ iff $u \not\Vdash F$, etc.).
2. $u \Vdash \Box F$ iff $v \Vdash F$ for every $v \in W$ with uRv .
3. $u \Vdash t:F$ iff $F \in \mathcal{E}(u, t)$ and $v \Vdash F$ for every $v \in W$ with uR^ev .

We say F is *true* at a world $u \in W$ if $u \Vdash F$; otherwise, F is *false* at u . Informally speaking, $t:F$ is true in a given world u iff t is an acceptable evidence term for F in u and F is true in all worlds v accessible from u via the evidence accessibility relation R^e . A formula F is *true* in a model if F is true at each world of the model; F is *valid* if F is true in every model. Given a constant specification \mathcal{CS} , a model \mathcal{M} *meets* \mathcal{CS} if $\mathcal{M} \Vdash c:A$ whenever $c:A \in \mathcal{CS}$.

The following lemma is a straightforward corollary of the definitions:

Lemma 5. $u \Vdash t:F$ and uR^ev yield $v \Vdash t:F$.

The above models with singleton W 's are called *Mkrtychev models* (M-models, for short). M-models were introduced in [26] under the name of pre-models. The logic of proofs LP was shown in [26] to be sound and complete with respect to M-models.

We call models with $R = R^e$ *Fitting models* (F-models). They were first introduced in [16] under the name *weak models* as an epistemic semantics for the logic of proofs LP. In [9, 17], it was shown that F-models work for S4LP as well.

Finally, we call arbitrary models of the above class *AF-models*. AF-models were introduced in [6] in a general setting for several agents where the need to separate knowledge and explicit knowledge was apparent. AF-models work for a wide class of systems, including the ones mentioned above (LP, S4LP, and S4LPN).

Theorem 1. *For any given constant specification \mathcal{CS} , the logic $\text{S4LP}_{\mathcal{CS}}$ is sound and complete with respect to AF-models that meet \mathcal{CS} .*

Proof. Soundness is straightforward. S4-axioms and rules hold because an AF-model with respect to the modal language is the usual Kripke model for S4. LP axioms and rules are guaranteed by the properties of the evidence function \mathcal{E} . Let us check the connection axiom $t:F \rightarrow \Box F$. Suppose $u \Vdash t:F$ and uRv . Then $uR^e v$, since $R \subseteq R^e$, and $v \Vdash F$. Hence, $u \Vdash \Box F$.

Completeness is established by the standard maximal consistent set construction. First of all, we define the canonical model $(W, R, \mathcal{E}, \Vdash)$ for $\text{S4LP}_{\mathcal{CS}}$. Call a set S of formulas in the language of $\text{S4LP}_{\mathcal{CS}}$ *consistent* if for no $F_1, \dots, F_n \in S$, $\neg(F_1 \wedge \dots \wedge F_n)$ is provable in $\text{S4LP}_{\mathcal{CS}}$. Consistent sets extend to *maximal consistent* sets by the standard Lindenbaum construction. W is the collection of all maximal consistent sets. Define Γ^\sharp as $\{\Box F \mid \Box F \in \Gamma\}$ and Γ^b as $\{t:F \mid t:F \in \Gamma\}$. The accessibility relation R , the evidence accessibility relation R^e and the evidence function \mathcal{E} are defined by

$$\Gamma R \Delta \quad \text{iff} \quad \Gamma^\sharp \subseteq \Delta ,$$

$$\Gamma R^e \Delta \quad \text{iff} \quad \Gamma^b \subseteq \Delta ,$$

and

$$F \in \mathcal{E}(\Gamma, t) \quad \text{iff} \quad t:F \in \Gamma .$$

Obviously, R^e extends R . Indeed, let $\Gamma R \Delta$ and $t:F \in \Gamma$ hold. Then $\Box t:F \in \Gamma$, since $\text{S4LP}_{\mathcal{CS}} \vdash t:F \rightarrow \Box t:F$. By $\Gamma R \Delta$, we conclude that $\Box t:F \in \Delta$. Since $\text{S4LP}_{\mathcal{CS}} \vdash \Box t:F \rightarrow t:F$, $t:F \in \Delta$ as well. So, $R \subseteq R^e$. Hence, (W, R, R^e) is an S4LP-frame.

Let us check the evidence function properties.

Monotonicity: $F \in \mathcal{E}(\Gamma, t)$ yields $t:F \in \Gamma$. If $\Gamma R^e \Delta$, then $t:F \in \Delta$, by the definition of R^e . By the definition of \mathcal{E} , $F \in \mathcal{E}(\Delta, t)$. *Application:* $F \rightarrow G \in \mathcal{E}(\Gamma, s)$ and $F \in \mathcal{E}(G, t)$ implies $s:(F \rightarrow G) \in \Gamma$ and $t:F \in \Gamma$. Since $s:(F \rightarrow G) \rightarrow (t:F \rightarrow (s \cdot t):G) \in \Gamma$ and Γ is closed under *modus ponens* (as a

maximal consistent set of formulas), $(st):G \in \Gamma$. Hence, $G \in \mathcal{E}(u, st)$. A similar argument proves *inspection* and *sum*.

Finally, for each propositional letter p ,

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

Lemma 6. (Truth Lemma) *For each formula F and each $\Gamma \in W$,*

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

Proof. Induction on F . The base case is given by the definitions and the cases of boolean connectives are standard.

Case: F is $\Box X$.

If $\Box X \in \Gamma$, then $\Box X \in \Delta$ for each Δ such that $\Gamma R \Delta$. Since $\text{S4LP}_{CS} \vdash \Box X \rightarrow X$, $X \in \Delta$. By the induction hypothesis, $\Delta \Vdash X$, hence, $\Gamma \Vdash \Box X$.

If $\Box X \notin \Gamma$, then $\Gamma^\sharp \cup \{\neg X\}$ is a consistent set. It it were not consistent, then $\text{S4LP}_{CS} \vdash \Box Y_1 \wedge \Box Y_2 \wedge \dots \wedge \Box Y_n \rightarrow X$ for some $\Box Y_1, \Box Y_2, \dots, \Box Y_n \in \Gamma$. By S4 reasoning, $\text{S4LP}_{CS} \vdash \Box Y_1 \wedge \Box Y_2 \wedge \dots \wedge \Box Y_n \rightarrow \Box X$, hence $\Box X \in \Gamma$, a contradiction. So, $\Gamma^\sharp \cup \{\neg X\}$ is consistent. Take Δ to be a maximal consistent extension of $\Gamma^\sharp \cup \{\neg X\}$. It is apparent that $\Delta \in W$, $\Gamma R \Delta$ and $X \notin \Delta$. By the definition of a model, $\Delta \not\Vdash X$, hence $\Gamma \not\Vdash \Box X$.

Case: F is $t:X$.

Let $t:X \in \Gamma$. Then $X \in \mathcal{E}(\Gamma, t)$. By the definition of R^e , $t:X \in \Delta$ for each Δ such that $\Gamma R^e \Delta$. Since $\text{S4LP}_{CS} \vdash t:X \rightarrow X$, $X \in \Delta$ as well. By the induction hypothesis, $\Delta \Vdash X$. By the definition of forcing at node Γ , $\Gamma \Vdash t:X$.

If $\Gamma \not\Vdash t:X$, then $X \in \mathcal{E}(\Gamma, t)$, hence $t:X \in \Gamma$, by the definition of \mathcal{E} . \square

To conclude the proof of Theorem 1, suppose $\text{S4LP}_{CS} \not\Vdash F$. Then $\{\neg F\}$ is a consistent set. Take its maximal consistent extension Γ . Then $F \notin \Gamma$ and, by the Truth Lemma, $\Gamma \not\Vdash F$ in the canonical model. \square

Theorem 1 also follows from the completeness of S4LP with respect to M-models proven in [17], where the canonical model W , R , \mathcal{E} and \Vdash were chosen as above and R^e was defined as $R^e = R$. The completeness proof is essentially the same as above with the following two minor deviations.

1. To establish the *monotonicity* property of \mathcal{E} , assume $F \in \mathcal{E}(\Gamma, t)$. Then $t:F \in \Gamma$ and $\Box t:F \in \Gamma$ by positive introspection in S4LP_{CS} . By the definition of R^e as R , $\Box t:F \in \Delta$ for each Δ such that $\Gamma R^e \Delta$. By reflexivity, $t:F \in \Delta$. Hence, $F \in \mathcal{E}(\Delta, t)$.

2. The case $t:X \in G$ in the Truth Lemma. By the definition of \mathcal{E} , $X \in \mathcal{E}(\Gamma, t)$. Take Δ such that $\Gamma R^e \Delta$, i.e., $\Gamma R \Delta$. By positive introspection, $\Box t:X \in \Gamma$, hence $\Box t:X \in \Delta$. By reflexivity, $t:X \in \Delta$ and $X \in \Delta$. By the induction hypothesis, $\Delta \Vdash X$. By the definition of forcing, $\Gamma \Vdash t:X$.

Theorem 2. *For any constant specification CS, S4LPN_{CS} is sound and complete with respect to AF-models with symmetric R^e meeting CS.*

Proof. Let $(W, R, R^e, \mathcal{E}, \Vdash)$ be an AF-model from the formulation of the theorem. By the definitions, $R \subseteq R^e$, R is reflexive and transitive, whereas R^e is reflexive, symmetric, and transitive, i.e., R^e is an equivalence relation on W that extends R .

The soundness part can be established by a straightforward induction on derivations in $\mathbf{S4LPN}_{cs}$. All the cases but C2 follow from AF-soundness of $\mathbf{S4LP}$, cf. Theorem 1. Let us check C2. Suppose $u \Vdash \neg t:F$, and pick v such that uRv . Suppose $v \not\Vdash \neg t:F$, i.e., $v \Vdash t:F$. Since $vR^e u$, by Lemma 5, $u \Vdash t:F$ —a contradiction. Actually, we have shown the stability property of AF-models of the above kind: each formula $t:F$ either holds at all worlds of a given equivalence class with respect to R^e , or it does not hold in all worlds of this class.

The completeness part is proved by the maximal consistent sets construction. Define the canonical model for $\mathbf{S4LPN}_{cs}$. Call a set S of formulas in the language of $\mathbf{S4LPN}_{cs}$ *consistent* if for no $F_1, \dots, F_n \in S$, is $\neg(F_1 \wedge \dots \wedge F_n)$ provable in $\mathbf{S4LPN}_{cs}$. Consistent sets extend to *maximal consistent* sets by the Lindenbaum construction. W is the collection of all maximal consistent sets. As before, $\Gamma^\sharp = \{\Box F \mid \Box F \in \Gamma\}$ and $\Gamma^\flat = \{t:F \mid t:F \in \Gamma\}$. Define R , R^e , and \mathcal{E} by

$$\begin{aligned} \Gamma R \Delta & \quad \text{iff} \quad \Gamma^\sharp \subseteq \Delta, \\ \Gamma R^e \Delta & \quad \text{iff} \quad \Gamma^\flat = \Delta^\flat, \end{aligned}$$

and

$$F \in \mathcal{E}(\Gamma, t) \quad \text{iff} \quad t:F \in \Gamma.$$

Finally, for each propositional letter p ,

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

Let us check that (W, R, R^e) is an $\mathbf{S4LPN}$ -frame. Clearly, R is reflexive and transitive, and R^e is an equivalence relation. Furthermore, $R \subseteq R^e$. Indeed, let $\Gamma R \Delta$ and $t:F \in \Gamma$. Since $\mathbf{S4LPN}_{cs} \vdash t:F \rightarrow \Box t:F$, the latter formula is in Γ , hence $\Box t:F \in \Gamma$ as well. Since $\Gamma R \Delta$, $\Box t:F \in \Delta$. Since $\mathbf{S4LPN}_{cs} \vdash \Box t:F \rightarrow t:F$, $t:F \in \Delta$. So, $\Gamma R^e \Delta$.

Let us check the properties of the evidence accessibility relation.

Monotonicity: Let $F \in \mathcal{E}(\Gamma, t)$ and $\Gamma R^e \Delta$. By the definition of $\mathcal{E}(\Gamma, t)$, $t:F \in \Gamma$. Hence $t:F \in \Delta$, since $\Delta^\flat = \Delta^\flat$. So, $F \in \mathcal{E}(\Delta, t)$. *Application, inspection,* and *sum* follow immediately from the definitions.

Lemma 7. (Truth Lemma) *For each formula F ,*

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

Proof. Induction on F . The base case is given by the definitions and the cases of Boolean connectives are standard.

Case: F is $\Box X$ is treated similarly to Lemma 6.

Case: F is $t:X$.

If $t:X \in \Gamma$, then $X \in \mathcal{E}(\Gamma, t)$. Let $\Gamma R^e \Delta$. By the definition of R^e , $t:X \in \Delta$. Since $\text{S4LPN}_{cs} \vdash t:X \rightarrow X$, $X \in \Delta$. So, by the induction hypothesis, $\Delta \Vdash X$. By the definitions, $\Gamma \Vdash t:X$.

If $\Gamma \Vdash t:X$, then $X \in \mathcal{E}(\Gamma, t)$, hence $t:X \in \Gamma$, by the definition of \mathcal{E} . \square

A standard argument concludes the proof of the theorem. Suppose $\text{S4LPN}_{cs} \not\vdash F$. Then the set $\{\neg F\}$ is consistent and let Γ be its maximal consistent extension. Then $F \notin \Gamma$ and, by Lemma 7, $\Gamma \not\vdash F$. \square

The following stronger form of the completeness theorem holds:

Theorem 3. *For each F such that $\text{S4LPN}_{cs} \not\vdash F$, there is an AF-model \widehat{M} that meets \mathcal{CS} such that the evidence accessibility relation in \widehat{M} is total and F is false in \widehat{M} .*

Proof. Take the canonical model for S4LPN_{cs} and a world Γ_0 such that $\Gamma_0 \not\vdash F$. Consider the equivalence class \widehat{W} with respect to R^e such that $\Gamma_0 \in \widehat{W}$. Since $R \subseteq R^e$, \widehat{W} is closed under accessibility relation R : if $\Gamma \in \widehat{W}$ and $\Gamma R \Delta$, then $\Delta \in \widehat{W}$. Let \widehat{R} , \widehat{R}^e , $\widehat{\mathcal{E}}$, and $\widehat{\Vdash}$ be R , R^e , \mathcal{E} , and \Vdash restricted to \widehat{W} , respectively. The resulting structure is an AF-model

$$\widehat{M} = (\widehat{W}, \widehat{R}, \widehat{R}^e, \widehat{\mathcal{E}}, \widehat{\Vdash})$$

with the evidence accessibility relation \widehat{R}^e total on its domain \widehat{W} . Indeed, we have already checked that $(\widehat{W}, \widehat{R}, \widehat{R}^e)$ is an AF-frame. The properties of the evidence function $\widehat{\mathcal{E}}$ are nothing but the special cases of the corresponding properties for \mathcal{E} .

Lemma 8. *For each formula X and each $\Gamma \in \widehat{W}$,*

$$\Gamma \Vdash F \text{ iff } F \in \Gamma .$$

Proof. Induction on F . The cases of atomic formulas and Boolean connectives are immediate.

Case: X is $\Box Y$.

If $\Gamma \Vdash \Box Y$, then $\Delta \Vdash Y$ for each Δ such that $\Gamma R \Delta$. In particular, $\Delta \Vdash Y$ for each $\Delta \in \widehat{W}$ such that $\Gamma R \Delta$. By the induction hypothesis, $\Delta \Vdash Y$, for each $\Delta \in \widehat{W}$ such that $\Gamma R \Delta$. Since R coincides with \widehat{R} on \widehat{W} , $\Delta \Vdash Y$, $\Delta \in \widehat{W}$ such that $\Gamma \widehat{R} \Delta$. By the definition of forcing in \widehat{M} , $\Gamma \widehat{\Vdash} \Box Y$.

If $\Gamma \not\vdash \Box Y$, then $\Delta \not\vdash Y$ for some Δ such that $\Gamma R \Delta$. Since \widehat{W} is closed under R , $\Delta \in \widehat{W}$. By the induction hypothesis, $\Delta \not\vdash Y$. Hence, $\Gamma \not\widehat{\Vdash} \Box Y$. \square

By Lemma 8, F is false at Γ_0 . This concludes the proof of Theorem 3. \square

6 Arithmetical semantics for S4LP and S4LPN

Arithmetical semantics for S4LP and S4LPN is given by interpreting $\Box F$ via the strong provability operator

F is true and provable in Peano Arithmetic PA,

together with interpreting $t:F$ as before:

t is a proof of F in Peano Arithmetic PA.

Using the strong provability operator to obtain S4-compliant logics has been a well established tradition in provability logic (cf. [1, 12, 13, 21, 24, 25, 28, 29]).

Theorem 4. (Arithmetical soundness of epistemic logic with justification) *Let \mathcal{CS} be a finite constant specification. If $\text{S4LPN}_{\mathcal{CS}} \vdash F$, then F is true under any arithmetical interpretation which translates $\Box F$ as strong provability in PA and $t:F$ as “ t is a proof of F in PA.”*

Proof. By induction on F . The validity of LP axioms and rules was shown in [3, 4]. The validity of S4 axioms and rules under the strong provability interpretation was shown in many sources, cf. [13].

The validity of the connection axiom $t:F \rightarrow \Box F$ is a combination of the validity of the explicit reflection $t:F \rightarrow F$, which is an LP axiom, already checked, and a first order tautology $\text{Prf}(t, \varphi) \rightarrow \exists x \text{Prf}(x, \varphi)$, where $\text{Prf}(x, y)$ is an arithmetical formula for x is a proof of y .

Finally, the negative introspection axiom $\neg t:F \rightarrow \Box \neg t:F$ is a special case of σ -completeness of the arithmetic, cf. [13]. \square

An arithmetically complete system GrzLPN of strong provability with proofs can be axiomatized by adding to S4LP the modal axiom by Grzegorzczuk $\Box(\Box(F \rightarrow \Box F) \rightarrow F) \rightarrow F$. Models for GrzLPN are F-models with reflexive partially ordered frames. This can be established by a combination of the methods from [9, 28, 29].

References

- [1] S. Artemov. Kolmogorov logic of problems and a provability interpretation of intuitionistic logic. *Theoretical Aspects of Reasoning about Knowledge - III Proceedings*, Morgan Kaufman Pbl., pp.257-272, 1990.
- [2] S. Artemov. Logic of Proofs. *Annals of Pure and Applied Logic*, v. 67, pp. 29–59, 1994.
- [3] S. Artemov. *Operational Modal Logic*. Tech. Rep. MSI 95-29, Cornell University, December 1995.

- [4] S. Artemov. Explicit provability and constructive semantics. The *Bulletin for Symbolic Logic*, v.7, No. 1, pp. 1–36, 2001.
- [5] S. Artemov. Kolmogorov’s and Gödel’s approach to intuitionistic logic: current developments. *Russian Mathematical Surveys*, v. 59, issue 2, pp.203-229, March-April 2004.
- [6] S. Artemov. *Evidence-based common knowledge*. Technical Report TR-2004018, CUNY Ph.D. Program in Computer Science, 2004.
<http://www.cs.gc.cuny.edu/tr/techreport.php?id=140>.
- [7] S. Artemov and L. Beklemishev. Provability logic. In D. Gabbay and F. Guenther, editors, *Handbook of Philosophical Logic, 2nd ed.*, volume 13, pages 229–403. Kluwer, Dordrecht, 2004.
- [8] S. Artemov, E. Kazakov and D. Shapiro. *On logic of knowledge with justifications*. Technical Report CFIS 99-12, Cornell University, 1999.
<http://www.cs.gc.cuny.edu/~sartemov/publications/S5LP.ps>.
- [9] 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.
<http://www.cs.gc.cuny.edu/tr/techreport.php?id=106>.
- [10] V. Brezhnev. *On explicit counterparts of modal logics*. Technical Report CFIS 2000-05, Cornell University, 2000.
- [11] 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/>.
- [12] G. Boolos, *The Unprovability of Consistency: An Essay in Modal Logic*, Cambridge University Press, 1979.
- [13] G. Boolos, *The Logic of Provability*, Cambridge University Press, 1993.
- [14] D. de Jongh and G. Japaridze. The Logic of Provability. Ed. S. Buss, *Handbook of Proof Theory*. Studies in Logic and the Foundations of Mathematics, Vol.137., Elsevier, pp. 475-546, 1998.
- [15] R. Fagin, J. Halpern, Y. Moses, M. Vardi, *Reasoning About Knowledge*, MIT Press, 1995.
- [16] M. C. Fitting. A semantics for the logic of proofs. Technical Report TR-2003012, CUNY Ph.D. Program in Computer Science, 2003.
<http://www.cs.gc.cuny.edu/tr/techreport.php?id=86>.
- [17] M. C. Fitting. *Semantics and Tableaus for LPS4*. Technical Report TR-2004016, CUNY Ph.D. Program in Computer Science, 2004.
<http://www.cs.gc.cuny.edu/tr/techreport.php?id=118>.

- [18] M. C. Fitting. The logic of proofs, semantically. *Annals of Pure and Applied Logic*, 132(1):1–25, 2005.
- [19] K. Gödel. Eine Interpretation des intuitionistischen Aussagenkalküls. *Ergebnisse Math. Colloq.*, Bd. 4 (1933), S. 39–40.
- [20] K. Gödel. Vortrag bei Zilsel (1938). In S. Feferman, ed. *Kurt Gödel Collected Works. Volume III*, Oxford University Press, 1995.
- [21] R. Goldblatt, “Arithmetical necessity, provability and intuitionistic logic”, *Theoria*, 44, pp. 38–46, 1978.
- [22] S. Kleene. *Introduction to Metamathematics*, Van Norstrand, 1952.
- [23] A. Kolmogoroff. Zur Deutung der intuitionistischen logik. *Mathematische Zeitschrift*, 35:58–65, 1932. In German. English translation in V.M. Tikhomirov, editor, *Selected works of A.N. Kolmogorov. Volume I: Mathematics and Mechanics*, pages 151–158. Kluwer, Dordrecht 1991.
- [24] A.V. Kuznetsov and A.Yu. Muravitsky, “Magari algebras,” *Fourteenth All-Union Algebra Conf., Abstract part 2: Rings, Algebraic Structures*, pp. 105–106, 1977 (in Russian)
- [25] A.V. Kuznetsov and A.Yu. Muravitsky, “On superintuitionistic logics as fragments of proof logic,” *Studia Logica*, v. XLV, pp. 76–99, 1986
- [26] A. Mkrtychev. Models for the Logic of Proofs. Lecture Notes in Computer Science, v. 1234, *Logical Foundations of Computer Science’ 97, Yaroslavl’*, pp. 266–275, 1997.
- [27] Y. Moses. Resource-bounded knowledge. In *Theoretical Aspects of Reasoning about Knowledge*, ed. M. Vardi, Morgan Kaufmann, pp. 261–276, 1988.
- [28] E. Nogina. *Logic of proofs with the strong provability operator*. ILLC Pre-publication Series ML-94-10, Institute for Logic, Language and Computation, University of Amsterdam, 1994.
- [29] E. Nogina. Grzegorzcyk logic with arithmetical proof operators. *Fundamental and Applied Mathematics*, v. 2, No.2, pp.483–499, (Russian), 1996
<http://mech.math.msu.su/fpm/eng/96/962/96206.htm>
- [30] R. Parikh. Knowledge and the problem of logical omniscience. *ISMIS-87 (International Symposium on Methodology for Intellectual Systems)*. Ed. Z. Ras and M. Zemankova, North Holland, pp.432–439, 1987.
- [31] R. Parikh. Logical omniscience. In *Logic and Computational Complexity*, Ed. Leivant, Springer Lecture Notes in Computer Science, No.960, pp. 22–29, 1995.

- [32] T. Sidon, “Provability Logic with Operations on Proofs”, Lecture Notes in Computer Science, v. 1234, *Logical Foundations of Computer Science’ 97, Yaroslavl’*, pp. 342-353, 1997.
- [33] J. van Benthem. Reflections on epistemic logic. *Logique & Analyse*, v. 133-134, pp. 5-14, 1991.
- [34] T. Yavorskaya (Sidon). Logic of proofs and provability. *Annals of Pure and Applied Logic*, 113(1-3):345–372, 2002.