

A logic of interactive proofs

David Lehnerr Zoran Ognjanović* Thomas Studer†

Abstract

We introduce the probabilistic two-agent justification logic IPJ , a logic in which we can reason about agents that perform interactive proofs. In order to study the growth rate of the probabilities in IPJ , we present a new method of parametrizing IPJ over certain negligible functions. Further, our approach leads to a new notion of zero-knowledge proofs.

Keywords: interactive proof system, zero-knowledge proof, epistemic logic, justification logic, probabilistic logic

1 Introduction

An interactive proof system [7, 11] is a protocol between two agents, the prover and the verifier. The aim of the protocol is that the prover can prove its knowledge of a secret to the verifier. To achieve this, the prover must answer a challenge provided by the verifier. Usually, the protocols are such that the verifier only knows with high probability that the prover knows the secret, that is the probability is a negligible function in the length of the challenge.

Several formalizations of the notion *proof of knowledge* are compared and analyzed in [8]. The aim of the present paper is to provide an epistemic logic model for interactive proofs of knowledge.

Our logic of interactive proofs and justifications IPJ_1 will be a combination of modal logic, justification logic, and probabilistic logic. The logic includes two agents, P (the prover) and V (the verifier). The modal part of IPJ_1

*Supported by the Science Fund of the Republic of Serbia project AI4TrustBC.

†Supported by the Swiss National Science Foundation grant 200020.184625.

consists of two S4 modalities \Box_P and \Box_V . As usual, \Box_a means *agent a knows that*. Justification logic adds explicit reasons for the agents' knowledge [5, 17]. We have formulas of the form $t{:}_a\alpha$, which stand for *agent a knows α for reason t* . The reason represented by the term t , can be a formal proof as in the first justification logic, the Logic of Proofs [2, 16], the execution of an interactive proof protocol, the result of an agent's reasoning, or any other justification of knowledge like, e.g., direct observation. For IPJ_1 , we will use a two-agent version of the logic of proofs together with the justification yields belief principle $t{:}_a\alpha \rightarrow \Box_a\alpha$. The third ingredient of IPJ_1 are probability operators of the form $\mathcal{P}_{\geq r}$ and $\mathcal{P}_{\approx r}$ meaning *with probability greater than or equal to r* and *with probability approximately r* , respectively. For the probabilistic part, we use the approach of [18, 19], which has been adapted to justification logic in [13, 14]. In order to deal with approximate probabilities, we need probability measures that can take non-standard values. Logics of this kind have been investigated in [20, 21].

Goldwasser et al. [11] introduced interactive proof systems as follows. Let \mathcal{L} be a language and P and V a pair of interacting (probabilistic) Turing machines, where P has unrestricted computational power and V is polynomial time. $\langle P, V \rangle$ is an interactive proof system for \mathcal{L} if the following conditions hold:

1. **Completeness:** For all $k \in \mathbb{N}$, there exists an $m \in \mathbb{N}$ such that for all inputs $x \in \mathcal{L}$ with $|x| > m$, the probability of $\langle P, V \rangle$ accepting x is at least $1 - |x|^{-k}$.
2. **Soundness:** For all $k \in \mathbb{N}$, there exists an $m \in \mathbb{N}$ such that for all inputs $x \notin \mathcal{L}$ with $|x| > m$ and any interactive Turing machine P' , the probability of $\langle P', V \rangle$ accepting x is at most $|x|^{-k}$.

Less formally, the agent P tries to prove its knowledge about a proposition α to the agent V . They may do that by following a challenge-response scheme. That is, V sends a challenge to P who then tries to answer it using his knowledge about α . On success, V 's confidence in P knowing α is increased. Moreover, the harder the challenge, the stronger is V 's belief. However, P may be dishonest and hence V may be convinced (with a low probability) that a wrong statement is true.

In order to model this in IPJ_1 , we introduce terms of the form f_t^n that represents V 's view of the run of the protocol where P has evidence t and n is a measure for the complexity of the run (this may refer to the complexity of

the challenge in a challenge response scheme). The outcome of a run will be formalized as $\mathcal{P}_{\geq r}(f_t^n :_V \Box_P \alpha)$ meaning that with probability greater than or equal to r , the run of the protocol with complexity n provides a justification for V that P knows α . Note that we are abstracting away the concrete protocol. Moreover, the subscript t in f_t^n does not imply that V has access to t ; it only states that P 's role in the protocol depends on t . We say that a formula α is interactively provable if the following two conditions hold:

1. **Completeness:** Assume $t :_P \alpha$. For all $k \in \mathbb{N}$, there exists a degree of complexity $m \in \mathbb{N}$ such that, for $n > m$ the probability of f_t^n justifying $\Box_P \alpha$ from V 's view is at least $1 - n^{-k}$.
2. **Soundness:** Assume $\neg t :_P \alpha$. For all $k \in \mathbb{N}$, there exists a degree of complexity $m \in \mathbb{N}$ such that, for $n > m$ the probability of f_t^n justifying $\Box_P \alpha$ from V 's view is at most n^{-k} .

Since IPJ_1 is a propositional logic, we need a way to express the soundness and completeness condition without quantifiers. For integers m, k , we start with sets of formulas $\mathsf{I}_{m,k}$ and define the set of interactively provable formulas

$$\mathsf{I} := \bigcap_k \bigcup_m \mathsf{I}_{m,k}.$$

If a formula α belongs to $\mathsf{I}_{m,k}$, then the following two conditions must hold for $n > m$:

1. $t :_P \alpha \rightarrow \mathcal{P}_{\geq 1 - \frac{1}{n^k}}(f_t^n :_V \Box_P \alpha)$
2. $\neg(t :_P \alpha) \rightarrow \mathcal{P}_{\leq \frac{1}{n^k}}(f_t^n :_V \Box_P \alpha)$

Therefore, if $\alpha \in \mathsf{I}$ and $t :_P \alpha$ then, for every k , there exists an m such that $\alpha \in \mathsf{I}_{m,k}$ and thus $\mathcal{P}_{\geq 1 - \frac{1}{n^k}}(f_t^n :_V \Box_P \alpha)$. Observe that this closely resembles the previously stated completeness property of interactive proof systems. The soundness property is obtained analogously.

Furthermore, we allow the probability operators to take non-standard values and consider protocols with transfinite complexity ω to capture the notion of a limit. Hence we can express statements of the form

if $t :_P \alpha$, then the probability of $f_t^\omega :_V \Box_P \alpha$ is almost 1.

Using the operator $\mathcal{P}_{\approx r}$, we add two more conditions for interactively provable formulas:

3. $t:P\alpha \rightarrow \mathcal{P}_{\approx 1}(f_t^\omega:V\Box_P\alpha)$ if $\alpha \in \mathbb{I}$;
4. $\neg(t:P\alpha) \rightarrow \mathcal{P}_{\approx 0}(f_t^\omega:V\Box_P\alpha)$ if $\alpha \in \mathbb{I}$.

We also include a principle saying that the justifications f_t^n are monotone in the complexity n :

5. $f_t^m:a\alpha \rightarrow f_t^n:a\alpha$ if $m < n$.

Justification logics with interacting agents are not new. Yavorskaya [25] introduced the evidence verification operator $!_P^V$ that can be used by V to verify P 's evidence, i.e. her system includes the axiom $t:P\alpha \rightarrow !_P^V t:Vt:P\alpha$. This resembles the definition of the complexity class NP as interactive proof system, see, e.g., [1]. There, the verifier is a deterministic Turing machine. The prover generates a proof certificate t for α (where the complexity of t is polynomial in α), i.e. we have $t:P\alpha$. Now P sends this certificate t to V and V checks it (which can be done in polynomial time). A successful check results in $!_P^V t$ being a justification for V that P knows the proof certificate t for α , i.e. $!_P^V t:Vt:P\alpha$.

2 Syntax

Let \mathbb{N} be the set of natural numbers and $\mathbb{N}^+ := \mathbb{N} \setminus \{0\}$. We define

$$\mathbf{Comp} := \mathbb{N} \cup \{\omega\}$$

where $\omega > n$ for each $n \in \mathbb{N}$.

We start with a countable set of justification variables and justification constants. Further we have a symbol f^n for each $n \in \mathbf{Comp}$. The set of *terms* \mathbf{Tm} is given by the following grammar

$$t ::= c \mid x \mid t \cdot t \mid t + t \mid !t \mid f^n t$$

where c is a justification constant and x is a justification variable. In the following, we usually write f_t^n for $f^n t$.

Our language is based on two agents, the prover P and the verifier V . We write a for an arbitrary agent, i.e. either P or V . Further, we use a countable set of atomic propositions \mathbf{Prop} . The set of *epistemic formulas* \mathbf{eFml} is given by the following grammar:

$$\alpha ::= p \mid \neg\alpha \mid \alpha \wedge \alpha \mid \Box_a\alpha \mid t:a\alpha$$

where p is an atomic proposition, t is a term and a is an agent.

For our formal approach, we consider probabilities that range over the unit interval of a non-archimedean recursive field that contains all rational numbers. We proceed as in [21] by choosing the unit interval of the Hardy field $\mathbb{Q}[\epsilon]$. The set $\mathbb{Q}[\epsilon]$ consists of all rational functions of a fixed non-zero infinitesimal $\epsilon \in \mathbb{R}^*$, where \mathbb{R}^* is a non-standard extension of \mathbb{R} (see [22]) for further details). Its positive elements have the form:

$$\epsilon^k \frac{\sum_{i=0}^n a_i \epsilon^i}{\sum_{i=0}^m b_i \epsilon^i},$$

where $a_i, b_i \in \mathbb{Q}$ for all $i \geq 0$ and $a_0 \cdot b_0 \neq 0$. We use S to denote the unit interval of $\mathbb{Q}[\epsilon]$.

The set of *formulas* Fml is given by the following grammar:

$$A ::= \alpha \mid \mathcal{P}_{\geq s} \alpha \mid \mathcal{P}_{\approx r} \alpha \mid \neg A \mid A \wedge A$$

where α is an epistemic formula, $s \in S$, and $r \in \mathbb{Q} \cap [0, 1]$.

Since any epistemic formula is a formula, we sometimes use latin letters to denote epistemic formulas, e.g. in $t:A \rightarrow \mathcal{P}_{\approx 1} B$, the letters A and B stand for epistemic formulas.

The remaining propositional connectives are defined as usual. Further we use the following syntactical abbreviations:

$$\begin{aligned} \mathcal{P}_{< s} \alpha &\text{ denotes } \neg \mathcal{P}_{\geq s} \alpha & \mathcal{P}_{\leq s} \alpha &\text{ denotes } \mathcal{P}_{\geq 1-s} \neg \alpha \\ \mathcal{P}_{> s} \alpha &\text{ denotes } \neg \mathcal{P}_{\leq s} \alpha & \mathcal{P}_{=s} \alpha &\text{ denotes } \mathcal{P}_{\leq s} \alpha \wedge \mathcal{P}_{\geq s} \alpha \end{aligned}$$

Our Logic of Interactive Proofs IPJ_l depends on a parameter l . We will introduce that parameter later when it will be relevant. We start with presenting the axioms of IPJ_l , which are divided into three groups: epistemic axioms, probabilistic axioms, interaction axioms.

Epistemic axioms

For both modal operators \Box_P and \Box_V we have the axioms for the modal logic S4.

- (p) all propositional tautologies
- (k) $\Box_a(A \rightarrow B) \rightarrow (\Box_a A \rightarrow \Box_a B)$
- (t) $\Box_a A \rightarrow A$
- (4) $\Box_a A \rightarrow \Box_a \Box_a A$

For both agents, we have the axioms for the Logic of Proofs [2] and the connection axiom (jyb). This yields the system S4LP from [6].

- (j) $s{:}_a(A \rightarrow B) \rightarrow (t{:}_a A \rightarrow_a s \cdot t{:}_a B)$
- (j+) $(s{:}_a A \vee t{:}_a A) \rightarrow (s + t){:}_a A$
- (jt) $t{:}_a A \rightarrow A$
- (j4) $t{:}_a A \rightarrow !t{:}_a t{:}_a A$
- (jyb) $t{:}_a A \rightarrow \Box_a A$

Probabilistic axioms

The probabilistic axioms correspond to the axiomatization of approximate conditional probabilities used in [20, 21] adapted to the unconditional case.

- (p1) $P_{\geq 0}A$
- (p2) $P_{\leq s}A \rightarrow P_{< t}A$, where $s < t$
- (p3) $P_{< s}A \rightarrow P_{\leq s}A$
- (p4) $P_{\geq 1}(A \leftrightarrow B) \rightarrow (P_{=s}A \rightarrow P_{=s}B)$
- (p5) $P_{\leq s}A \leftrightarrow P_{\geq 1-s}\neg A$
- (p6) $(P_{=s}A \wedge P_{=t}B \wedge P_{\geq 1}\neg(A \wedge B)) \rightarrow P_{=\min(1,s+t)}(A \vee B)$
- (pa1) $P_{\approx r}A \rightarrow P_{\geq r_1}A$, for every rational $r_1 \in [0, r]$
- (pa2) $P_{\approx r}A \rightarrow P_{\leq r_1}A$, for every rational $r_1 \in (r, 1]$

Interaction axioms

So far, we have axioms for an epistemic justification logic with approximate probabilities. Let us now add axioms for terms of the form f_t^n that model interactive proof protocols. These axioms depend on the parameter I in IPJ_I , which we introduce next.

An *interaction specification* I is a function $\mathsf{I} : \mathbb{N} \times \mathbb{N} \rightarrow \mathcal{P}(\text{eFml})$, i.e. to each $m, k \in \mathbb{N}$ we assign a set of epistemic formulas $\mathsf{I}(m, k)$. In the following, we write $\mathsf{I}_{m,k}$ for $\mathsf{I}(m, k)$. Further, we overload the notation and use I also to denote the set

$$\mathsf{I} := \bigcap_k \bigcup_m \mathsf{I}_{m,k}.$$

The interaction axioms are:

- (m) $f_t^m \cdot_a \alpha \rightarrow f_t^n \cdot_a \alpha$ for all $m, n \in \mathbf{Comp}$ such that $m < n$
- (c) $t :_P \alpha \rightarrow P_{\geq 1 - \frac{1}{n^k}}(f_t^n \cdot_V \Box_P \alpha)$ if $n > m$ and $\alpha \in \mathbf{I}_{m,k}$
- (s) $\neg(t :_P \alpha) \rightarrow P_{\leq \frac{1}{n^k}}(f_t^n \cdot_V \Box_P \alpha)$ if $n > m$ and $\alpha \in \mathbf{I}_{m,k}$
- (c ω) $t :_P \alpha \rightarrow P_{\approx 1}(f_t^\omega \cdot_V \Box_P \alpha)$ if $\alpha \in \mathbf{I}$
- (s ω) $\neg(t :_P \alpha) \rightarrow P_{\approx 0}(f_t^\omega \cdot_V \Box_P \alpha)$ if $\alpha \in \mathbf{I}$

Inference rules

The rules of \mathbf{IPJ}_1 are the following. We have modus ponens:

$$\frac{A \quad A \rightarrow B}{B}$$

\mathbf{IPJ}_1 also includes the modal necessitation rule as well as the axiom necessitation rule from justification logic:

$$\frac{A}{\Box A} \quad \frac{A \text{ is an axiom of } \mathbf{IPJ}_1}{c_1 \cdot_{a_1} c_2 \cdot_{a_2} \cdots c_n \cdot_{a_n} A}$$

for arbitrary constants c_i and agents a_i . Of course, it would be possible to parameterize \mathbf{IPJ}_1 additionally by a constant specification as it is often done in justification logic. This would not affect our treatment of interactive proofs.

We have the following rules for the probabilistic part:

1. From A infer $P_{\geq 1}A$
2. From $B \rightarrow P_{\neq s}A$ for all $s \in S$ infer $B \rightarrow \perp$
3. From $B \rightarrow P_{\geq r - \frac{1}{n}}A$ and $B \rightarrow P_{\leq r + \frac{1}{n}}A$ for all integers n , infer

$$B \rightarrow P_{\approx r}A$$

Of course in the last rule, only premises $B \rightarrow P_{\geq r - \frac{1}{n}}A$ are considered for which $r - \frac{1}{n} > 0$ holds and $B \rightarrow P_{\leq r + \frac{1}{n}}A$ is only considered if $r + \frac{1}{n} < 1$.

3 Semantics

For this section, we assume that we are given an arbitrary interaction specification I . Many notions in this chapter will depend on that parameter. For any set X we use $\mathcal{P}(X)$ to denote the power set of X . We will use a Fitting-style semantics [10] for justification logic, but modular models [4, 15] would work as well.

Definition 1 (Evidence relation). An *evidence relation* is a mapping

$$\mathcal{E} : \text{Tm} \rightarrow \mathcal{P}(\text{eFml})$$

from terms to sets of epistemic formulas such that for all $s, t \in \text{Tm}$, $\alpha \in \text{eFml}$, constants c_i , and agents a_i :

1. $\mathcal{E}(s) \cup \mathcal{E}(t) \subseteq \mathcal{E}(s + t)$;
2. $\mathcal{E}(s) \cdot \mathcal{E}(t) \subseteq \mathcal{E}(s \cdot t)$;
3. $t:\mathcal{E}(t) \subseteq \mathcal{E}(!t)$;
4. $c_2:a_2 \cdots c_n:a_n A \in \mathcal{E}(c_1)$ if α is an axiom;
5. $\alpha \in \mathcal{E}(f_t^n)$, if $\alpha \in \mathcal{E}(f_t^m)$ for $n > m$.

Definition 2 (Epistemic model). An *epistemic model* for IPJ_1 is a tuple $M = \langle W, R, \mathcal{E}, V \rangle$ where:

1. W is a non-empty set of objects called worlds.
2. R maps each agent a to a reflexive and transitive accessibility relation R_a on W .
3. \mathcal{E} maps each world w and each agent a to an evidence relation \mathcal{E}_w^a .
4. V is a valuation mapping each world to a set of atomic propositions.

Definition 3 (Truth within a world). Let $M = \langle W, R, \mathcal{E}, V \rangle$ be an epistemic model for IPJ_1 and let w be a world in W . For an epistemic formula $\alpha \in \text{eFml}$, we define $M, w \Vdash \alpha$ inductively by:

1. $M, w \Vdash \beta$ iff $\beta \in V(w)$ for $\beta \in \text{Prop}$
2. $M, w \Vdash \neg\beta$ iff $M, w \not\Vdash \beta$
3. $M, w \Vdash \beta \wedge \gamma$ iff $M, w \Vdash \beta$ and $M, w \Vdash \gamma$

4. $M, w \Vdash \Box_a \beta$ iff $M, u \Vdash \beta$ for all $u \in W$ with $R_a w u$
5. $M, w \Vdash t:{}_a \beta$ iff $\beta \in \mathcal{E}_w^a(t)$ and $M, u \Vdash \beta$ for all $u \in W$ with $R_a w u$

Definition 4 (Algebra). Let U be a non-empty set and let H be a non-empty subset of $\mathcal{P}(U)$. H will be called an algebra over U if the following hold:

- $U \in H$
- $X, Y \in H \rightarrow X \cup Y \in H$
- $X \in H \rightarrow U \setminus X \in H$

Definition 5 (Finitely additive measure). Let H be an algebra over U and $\mu : H \rightarrow S$, where S is the unit interval of the hardy field $\mathbb{Q}[\epsilon]$. We call μ a *finitely additive measure* if the following hold:

1. $\mu(U) = 1$
2. $X \cap Y = \emptyset \implies \mu(X \cup Y) = \mu(X) + \mu(Y)$ for all $X, Y \in H$.

Definition 6 (Probability space). A *probability space* is a triple $\langle U, H, \mu \rangle$ where:

1. U is a non-empty set
2. H is an algebra over U
3. $\mu : H \rightarrow S$ is a finitely additive measure

Definition 7 (Quasimodel). A quasimodel for IPJ_1 is a tuple

$$M = \langle W, R, \mathcal{E}, V, U, H, \mu, w_0 \rangle$$

such that

1. $\langle W, R, \mathcal{E}, V \rangle$ is an epistemic model for IPJ_1
2. $U \subseteq W$
3. $\langle U, H, \mu \rangle$ is a probability space
4. $w_0 \in U$

Let $M = \langle W, R, \mathcal{E}, V, U, H, \mu, w_0 \rangle$ be a quasimodel, $w \in W$, and $\alpha \in \mathbf{eFml}$. Since M contains an epistemic model, we write $M, w \Vdash \alpha$ for

$$\langle W, R, \mathcal{E}, V \rangle, w \Vdash \alpha.$$

Definition 8 (Events). Let $M = \langle W, R, \mathcal{E}, V, U, H, \mu, w_0 \rangle$ be a quasimodel. For an epistemic formula $\alpha \in \mathbf{eFml}$, we define the event that α occurs as

$$[\alpha]_M := \{u \in U \mid M, u \Vdash \alpha\}$$

We use $[\alpha]_M^C$ for the complement event $U \setminus [\alpha]_M$.

When the quasimodel M is clear from the context, we often drop the subscript M in $[\alpha]_M$.

Definition 9 (Independent events). Let M be a quasimodel. We say that two events $S, T \in H$ are independent in M if

$$\mu(S \cap T) = \mu(S) \cdot \mu(T).$$

Definition 10 (Probability almost r). Let $\langle U, H, \mu \rangle$ be a probability space. For $r \in \mathbb{Q} \cap [0, 1]$, we say that $X \in H$ has probability almost r ($\mu(X) \approx r$) if for all $n \in \mathbb{N}^+$ $\mu(X) \in [r - \frac{1}{n}, r + \frac{1}{n}]$.

Definition 11 (Truth in a quasimodel). Let

$$M = \langle W, R, \mathcal{E}, V, U, H, \mu, w_0 \rangle$$

be quasimodel for \mathbf{IPJ}_1 . We define $M \models A$ inductively by:

1. $M \models A$ iff $M, w_0 \Vdash A$ for $A \in \mathbf{eFml}$; otherwise
2. $M \models \neg B$ iff $M \not\models B$
3. $M \models B \wedge C$ iff $M \models B$ and $M \models C$
4. $M \models \mathcal{P}_{\geq s} \alpha$ iff $\mu([\alpha]) \geq s$
5. $M \models \mathcal{P}_{\approx r} \alpha$ iff $\mu([\alpha]) \approx r$

Definition 12 (Measurable model). A quasimodel

$$M = \langle W, R, \mathcal{E}, V, U, H, \mu, w_0 \rangle$$

is called *measurable* if $[\alpha] \in H$ for all $\alpha \in \mathbf{eFml}$.

Definition 13 (Model). A *model* for IPJ_1 is a measurable quasimodel M for IPJ_1 that satisfies:

1. $M \models t :_P \alpha \rightarrow \mathcal{P}_{\geq 1 - \frac{1}{n^k}}(f_t^n :_V \Box^P \alpha)$ if $n > m$ and $\alpha \in \mathbf{I}_{m,k}$;
2. $M \models \neg(t :_P \alpha) \rightarrow \mathcal{P}_{\leq \frac{1}{n^k}}(f_t^n :_V \Box^P \alpha)$ if $n > m$ and $\alpha \in \mathbf{I}_{m,k}$.

We say that a formula A is IPJ_1 -valid if $M \models A$ for all models M for IPJ_1 .

4 Properties and Results

We start with two auxiliary lemmas.

Lemma 14. *Let β, γ be epistemic formulas. IPJ_1 proves*

1. $\mathcal{P}_{=s}\gamma \rightarrow \mathcal{P}_{\leq s}(\gamma \wedge \beta)$.
2. $\mathcal{P}_{\leq s}\gamma \wedge \mathcal{P}_{< r}\beta \rightarrow \mathcal{P}_{< r+s}(\gamma \vee \beta)$ where $r + s \leq 1$.

Proof. For the first claim, suppose $\mathcal{P}_{=s}\gamma$. Thus we get $\mathcal{P}_{=1-s}\neg\gamma$. Further let t be such that $\mathcal{P}_{=t}(\neg\beta \wedge \gamma)$. Using axiom (p6) we infer

$$\mathcal{P}_{=(1-s)+t}(\neg\gamma \vee (\neg\beta \wedge \gamma)).$$

Since $(1-s) + t = 1 - (s-t)$, this is equivalent to

$$\mathcal{P}_{=s-t}(\gamma \wedge \neg(\neg\beta \wedge \gamma)).$$

By axiom (p4) we find

$$\mathcal{P}_{=s-t}(\gamma \wedge \beta).$$

We conclude $\mathcal{P}_{\leq s}(\gamma \wedge \beta)$.

To show the second claim, suppose $\mathcal{P}_{\leq s}\gamma$. By the first claim we get

$$\mathcal{P}_{\leq s}(\gamma \wedge \neg\beta).$$

From $\mathcal{P}_{< r}\beta$ we obtain using axiom (p6) that $\mathcal{P}_{< r+s}((\gamma \wedge \neg\beta) \vee \beta)$. Using axiom (p4) we conclude $\mathcal{P}_{< r+s}(\gamma \vee \beta)$. \square

We can read the operator $\mathcal{P}_{\approx 1}$ as *it is almost certain that*. This operator provably behaves like a normal modality.

Lemma 15. *Let α, β be epistemic formulas.*

1. IPJ_1 proves $\mathcal{P}_{\approx 1}(\alpha \rightarrow \beta) \rightarrow (\mathcal{P}_{\approx 1}\alpha \rightarrow \mathcal{P}_{\approx 1}\beta)$.
2. The rule $\frac{\alpha}{\mathcal{P}_{\approx 1}\alpha}$ is derivable in IPJ_1 .

Proof. We first establish that IPJ_1 proves

$$\mathcal{P}_{\approx 1}(\gamma \vee \beta) \wedge \mathcal{P}_{\approx 0}\gamma \rightarrow \mathcal{P}_{\approx 1}\beta. \quad (1)$$

From $\mathcal{P}_{\approx 1}(\gamma \vee \beta)$ we get

$$\forall r < 1 \text{ we have } \mathcal{P}_{\geq r}(\gamma \vee \beta). \quad (2)$$

From $\mathcal{P}_{\approx 0}\gamma$ we get

$$\forall s > 0 \text{ we have } \mathcal{P}_{\leq s}\gamma. \quad (3)$$

From (2) and (3) we obtain $\mathcal{P}_{\approx 1}\beta$. Suppose towards a contradiction that there exists $r < 1$ with $\neg\mathcal{P}_{\geq r}\beta$. By the definition of $\mathcal{P}_{< r}$ this is $\mathcal{P}_{< r}\beta$. Together with (3) this yields by the second claim of the previous lemma that

$$\mathcal{P}_{< r+s}(\gamma \vee \beta) \quad \forall s > 0 \text{ with } r + s < 1.$$

For $s' = \frac{1-r}{2}$ we have $r + s' = \frac{1+r}{2} < 1$. Thus there exists $q < 1$ with $\mathcal{P}_{< q}(\gamma \vee \beta)$, which contradicts (2). Hence (1) is established. Let γ be $\neg\alpha$ and observe that $\mathcal{P}_{\approx 1}\alpha \rightarrow \mathcal{P}_{\approx 0}\neg\alpha$ is provable in IPJ_1 . Now the first claim of this lemma immediately follows from (1).

It remains to show that the rule of $\mathcal{P}_{\approx 1}$ necessitation is derivable. Suppose that α is derivable. Thus $\mathcal{P}_{\geq 1}\alpha$ is derivable. Using axioms (p2) and (p3) we obtain $\mathcal{P}_{\geq 1 - \frac{1}{n}}\alpha$ for all integers n . Thus we infer $\mathcal{P}_{\approx 1}\alpha$. \square

An immediate consequence of these lemmas is the following. If t justifies the prover's knowledge of α , then, with almost certainty, the interactive proof protocol based on t will be successful in providing the verifier with a justification for α .

Corollary 16. *For $\alpha \in \mathbb{I}$, IPJ_1 proves $t :_P \alpha \rightarrow \mathcal{P}_{\approx 1}(c \cdot f_t^\omega :_V \alpha)$ for a arbitrary constant c .*

The deductive system IPJ_1 is sound with respect to IPJ_1 -models.

Theorem 17 (Soundness). *Let l be an arbitrary interaction specification. For any formula F we have that*

$$\vdash F \text{ implies } F \text{ is IPJ}_1\text{-valid.}$$

Proof. As usual by induction on the length of the derivation. The interesting case is when F is an instance of $(\mathsf{c}\omega)$. But first note that axioms (m) and (c) are IPJ₁-valid because of Definition 1 and Definition 13, respectively.

Now let F be an instance of $(\mathsf{c}\omega)$. Then F is of the form

$$t:{}_P\alpha \rightarrow P_{\approx 1}(f_t^\omega:{}_V\Box_P\alpha)$$

for some $\alpha \in \mathsf{l}$. Let $M = \langle W, R, \mathcal{E}, V, U, H, \mu, w_0 \rangle$ be an arbitrary model for IPJ₁ and assume $M \models t:{}_P\alpha$. We need to show

$$\mu([f_t^\omega:{}_V\Box_P\alpha]) \in \left[1 - \frac{1}{n}, 1\right] \text{ for all } n \in \mathbb{N}^+. \quad (4)$$

We fix an arbitrary $n \in \mathbb{N}^+$. Because of $\alpha \in \mathsf{l}$, we know that there exists an m such that $\alpha \in \mathsf{l}_{m,1}$. By soundness of axiom (c) we find that for each $n' > m$

$$\mu([f_t^{n'}:{}_V\Box_P\alpha]) \geq 1 - \frac{1}{n'}.$$

Let $n'' \in \mathbb{N}$ be such that $n'' > m$ and $n'' \geq n$. We find

$$\mu([f_t^{n''}:{}_V\Box_P\alpha]) \geq 1 - \frac{1}{n''} \geq 1 - \frac{1}{n}. \quad (5)$$

By soundness of axiom (m) we get that for each $w \in W$

$$M, w \Vdash f_t^{n''}:{}_V\Box_P\alpha \text{ implies } M, w \Vdash f_t^\omega:{}_V\Box_P\alpha.$$

Therefore, and by finite additivity of μ , we obtain

$$\mu([f_t^\omega:{}_V\Box_P\alpha]) \geq \mu([f_t^{n''}:{}_V\Box_P\alpha]). \quad (6)$$

Taking (5) and (6) together yields (4). \square

In practice, one often considers interactive proofs systems that are round-based, see [1].

Definition 18 (Round-based interactive proof system). An interactive protocol $\langle P, V \rangle$ is called *round-based* if the following two conditions hold:

1. **Completeness:** Let $x \in \mathcal{L}$. There exists a polynomial $p(x)$ such that the probability that $\langle P, V \rangle$ halts in an accepting state after $p(x)$ many messages is at least $\frac{2}{3}$.
2. **Soundness:** Let $x \notin \mathcal{L}$ and let $p(x)$ be any polynomial. For any interactive Turing machine P' , the probability that $\langle P', V \rangle$ halts in an accepting state after $p(x)$ many messages is at most $\frac{1}{3}$.

This definition achieves negligible (resp. overwhelming) probabilities by repeating the protocol several times and deciding based on a majority vote. Although this definition is simple to model in IPJ_1 , it is not suitable for a limit analysis because our measure is not σ -additive. Note that to properly formalize σ -additivity one needs countable conjunctions and disjunctions [12], which we do not want to include here. However, for finitely many rounds, we can describe how the probability increases throughout the rounds (given that they are pairwise independent).

Lemma 19. *Let M be an IPJ_1 -model for an arbitrary interaction specification 1. Consider justification terms s_1, \dots, s_n and an epistemic formula α such that*

1. $M \models s_i :_V \alpha$ for each s_i ;
2. $[s_i :_V \alpha]$ and $[s_j :_V \alpha]$ are independent events for all $i \neq j$.

We find that $M \models \bigwedge_{i=1, \dots, n} \mathcal{P}_{\geq 1-r}(s_i :_V \alpha) \rightarrow \mathcal{P}_{\geq 1-r^n} \alpha$.

Proof. Whenever $s_i :_V \alpha$ is true at a world w , α is true at w by soundness of axiom (jt). Hence, by monotonicity of μ we find

$$\mu([\alpha]) \geq \mu\left(\bigcup_{i=1}^n [s_i :_V \alpha]\right) = 1 - \mu\left(\bigcap_{i=1}^n [s_i :_V \alpha]^C\right) \stackrel{\text{indep.}}{\geq} 1 - \prod_{i=1}^n r = 1 - r^n$$

□

An interactive proof protocol for a language \mathcal{L} has the zero-knowledge property if, from a successful execution, the verifier only learns that x belongs to \mathcal{L} but nothing else. Formally, a protocol is perfectly zero-knowledge if there exists a probabilistic Turing machine T that generates proof transcripts¹ that are indistinguishable from original ones. If the verifier can

¹In the setting of interactive Turing machines, a proof transcript is everything that V sees on the public tapes during the protocol.

obtain additional information with negligible probability, then the protocol is said to be statistically zero-knowledge.

However, we cannot directly implement this definition because it would require to model the Turing machine T as an agent and we would need to reason about something like indistinguishable terms. Simplified, a protocol is zero-knowledge if the verifier cannot compute the prover's secret. In our setting the prover's secret is represented by the term t . Hence, $f_t^n :_V t :_P \alpha$ means that the prover's secret has been revealed to the verifier. In fact, $f_t^n :_V t :_P \alpha$ being unlikely is a direct consequence of the protocol being statistically zero-knowledge because the probability of the verifier knowing the prover's secret is bound by its ability to distinguish between proof transcripts. This gives rise to the following definition of zero-knowledge in IPJ_1 .

Definition 20 (Evidentially zero-knowledge). A protocol is *evidentially zero-knowledge* if for all inputs x belonging to \mathcal{L} , the probability of the verifier knowing the prover's evidence for x belonging to \mathcal{L} is negligible.

To address evidentially zero-knowledge protocols, we add the following two axioms to IPJ_1 :

1. $t :_P \alpha \rightarrow \mathcal{P}_{\leq \frac{1}{n^k}}(f_t^n :_V t :_P \alpha)$ if $n > m$ and $\alpha \in \mathsf{I}_{m,k}$;
2. $t :_P \alpha \rightarrow \mathcal{P}_{\approx 0}(f_t^\omega :_V t :_P A)$ if $\alpha \in \mathsf{I}$.

Models for IPJ_1 are adjusted by requiring the condition:

$$M \models t :_P \alpha \rightarrow \mathcal{P}_{\leq \frac{1}{n^k}}(f_t^n :_V t :_P \alpha) \text{ if } n > m \text{ and } \alpha \in \mathsf{I}_{m,k}.$$

It is easy to show that this extension is sound with respect to its models. The proof of soundness for the second axiom is similar to the soundness proof of $(c\omega)$.

5 Conclusion

We presented the probabilistic two-agent justification logic IPJ_1 , in which we can reason about agents that perform interactive proofs. The foundation of this work is based on probabilistic justification logic combined with interacting evidence systems. We further proposed a new technique that asserts a countable axiomatization and makes it possible to reason about the

growth rate of a probability measure. Intuitively, the set $I = \bigcap_k \bigcup_m I_{m,k}$ can be thought of as the set of all formulas that are known to be interactively provable. For a formula $\alpha \in I_{m,k}$ and a term t with $t:P\alpha$,

$$\mathcal{P}_{\geq 1 - \frac{1}{n^k}}(f_t^n :_V \Box_P \alpha)$$

holds for all $n > m$. Hence, if $\alpha \in I$, then the following first order sentence is true

$$\forall k \exists m \forall (n > m) \mu([f_t^n :_V \Box_P \alpha]) \geq 1 - \frac{1}{n^k},$$

which is the definition of an overwhelming function.

Our approach of modelling limits with the help of specification sets is quite versatile as the following example shows.

Example 21. Consider a sequence of the form:

$$\mathcal{P}_{=L+0.5}(f_t^1 :_V \alpha) \quad \mathcal{P}_{=L+0.25}(f_t^2 :_V \alpha) \quad \mathcal{P}_{=L+0.125}(f_t^3 :_V \alpha) \quad \dots$$

The sentence we want to model is:

$$(\forall \epsilon > 0)(\exists m \geq 0)(\forall n > m)(\mathcal{P}_{\leq L+\epsilon}(f_t^n :_V \alpha) \wedge \mathcal{P}_{\geq L-\epsilon}(f_t^n :_V \alpha))$$

Again, for $\epsilon, L \in \mathbb{Q}$ and $m \in \mathbb{N}$, we define sets $\text{Conv}_{\epsilon,m}^L$ and let

$$\text{Conv}^L := \bigcap_{\epsilon \in \mathbb{Q}} \bigcup_{m \in \mathbb{N}} \text{Conv}_{\epsilon,m}^L.$$

With the following formulas, we can express that a sequence of probabilities converges:

1. $\mathcal{P}_{\leq L+\epsilon}(f_t^n :_V \alpha) \wedge \mathcal{P}_{\geq L-\epsilon}(f_t^n :_V \alpha)$ if $n > m$ and $\alpha \in \text{Conv}_{\epsilon,m}^L$;
2. $\mathcal{P}_{\approx L}(f_t^\omega :_V \alpha)$ if $\alpha \in \text{Conv}^L$.

Additionally, we showed that our model can address a round-based definition of interactive proofs, however only for finitely many rounds since our measure is not σ -additive. Further, we also investigated zero-knowledge proofs. As it turns out, IPJ_1 cannot model the original definition because we cannot compare justification terms in IPJ_1 . However, we introduced the notion of evidentially zero knowledge, which fits nicely in our framework.

Moreover, we established soundness of IPJ_1 . Our axiomatization is a combination of systems that are known to be complete and we conjecture that IPJ_1 is complete, too.

From a more general perspective, this paper complements the list of motivations for justification logic. There are the "classical" applications of justification logic in epistemology and proof theory [3, 5, 17]. Recently, justification logic also turned out to be useful to analyze certain deontic situations [9] as well as a paradox in quantum physics [23], both having to do with certain forms of consistency requirements. The presented logical analysis of zero knowledge proofs is a novel example that shows the importance of the distinction between explicit (where the justification is shown) and implicit (where the justification is hidden) knowledge. The essence of a zero knowledge proof of a proposition α is that the verifier knows that the prover knows α , but the verifier does not know the prover's justification for α . Thus the verifier does not know why the prover knows α (this hints at possible connections with the logic of knowing why [24]). That is, the verifier has explicit knowledge of the implicit knowledge of the prover.

References

- [1] S. Arora and B. Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009.
- [2] S. Artemov. Explicit provability and constructive semantics. *Bulletin of Symbolic Logic*, 7(1):1–36, Mar. 2001.
- [3] S. Artemov. The logic of justification. *The Review of Symbolic Logic*, 1(4):477–513, Dec. 2008.
- [4] S. Artemov. The ontology of justifications in the logical setting. *Studia Logica*, 100(1–2):17–30, Apr. 2012.
- [5] S. Artemov and M. Fitting. *Justification Logic: Reasoning with Reasons*. Cambridge University Press, 2019.
- [6] S. Artemov and E. Nogina. Introducing justification into epistemic logic. *Journal of Logic and Computation*, 15(6):1059–1073, Dec. 2005.

- [7] L. Babai. Trading group theory for randomness. In *Proceedings of the Seventeenth Annual ACM Symposium on Theory of Computing*, STOC '85, pages 421–429. Association for Computing Machinery, 1985.
- [8] M. Bellare and O. Goldreich. On defining proofs of knowledge. In E. F. Brickell, editor, *Advances in Cryptology — CRYPTO' 92*, pages 390–420, Berlin, Heidelberg, 1993. Springer Berlin Heidelberg.
- [9] F. Faroldi, M. Ghari, E. Lehmann, and T. Studer. Impossible and conflicting obligations in justification logic. In A. Marra, F. Liu, P. Portner, and F. Van De Putte, editors, *Proceedings of DEON 2020*, 2020.
- [10] M. Fitting. The logic of proofs, semantically. *Annals of Pure and Applied Logic*, 132(1):1–25, Feb. 2005.
- [11] S. Goldwasser, S. Micali, and C. Rackoff. The knowledge complexity of interactive proof-systems. In *Proceedings of the Seventeenth Annual ACM Symposium on Theory of Computing*, STOC '85, pages 291–304. Association for Computing Machinery, 1985.
- [12] N. Ikodinović, Z. Ognjanović, A. Perović, and M. Rašković. Completeness theorems for σ -additive probabilistic semantics. *Ann. Pure Appl. Log.*, 171(4), 2020.
- [13] I. Kokkinis, P. Maksimović, Z. Ognjanović, and T. Studer. First steps towards probabilistic justification logic. *Logic Journal of the IGPL*, 23(4):662–687, 2015.
- [14] I. Kokkinis, Z. Ognjanović, and T. Studer. Probabilistic justification logic. *Journal of Logic and Computation*, 30(1):257–280, 2020.
- [15] R. Kuznets and T. Studer. Justifications, ontology, and conservativity. In T. Bolander, T. Braüner, S. Ghilardi, and L. Moss, editors, *Advances in Modal Logic, Volume 9*, pages 437–458. College Publications, 2012.
- [16] R. Kuznets and T. Studer. Weak arithmetical interpretations for the logic of proofs. *Logic Journal of the IGPL*, 24(3):424–440, 2016.
- [17] R. Kuznets and T. Studer. *Logics of Proofs and Justifications*. College Publications, 2019.

- [18] Z. Ognjanović and M. Rašković. Some first order probability logics. *Theoretical Computer Science*, 247:191–212, 2000.
- [19] Z. Ognjanović, M. Rašković, and Z. Marković. *Probability Logics - Probability-Based Formalization of Uncertain Reasoning*. Springer, 2016.
- [20] Z. Ognjanović, N. Savić, and T. Studer. Justification logic with approximate conditional probabilities. In *Logic, Rationality, and Interaction - 6th International Workshop, LORI 2017, Sapporo, Japan, September 11-14, 2017, Proceedings*, pages 681–686, 2017.
- [21] M. Rašković, Z. Marković, and Z. Ognjanović. A logic with approximate conditional probabilities that can model default reasoning. *International Journal of Approximate Reasoning*, 49(1):52–66, 2008.
- [22] A. Robinson. *Non-standard Analysis*. Princeton University Press, 1996.
- [23] T. Studer. A conflict tolerant logic of explicit evidence. *Logical Investigations*, 27(1):124–144, 2021.
- [24] C. Xu, Y. Wang, and T. Studer. A logic of knowing why. *Synthese*, 198:1259–1285, 2021.
- [25] T. Yavorskaya (Sidon). Interacting explicit evidence systems. *Theory of Computing Systems*, 43(2):272–293, Aug. 2008.