Tracking Evidence

Sergei Artemov*

Graduate Center CUNY
365 Fifth Avenue, New York City, NY 10016
sartemov@gc.cuny.edu

April 25, 2010

Abstract

In this case study we describe an approach to a general logical framework for tracking evidence within epistemic contexts. We consider as basic an example which features two justifications for a true statement, one which is correct and one which is not. We formalize this example in a system of Justification Logic with two knowers: the object agent and the observer, and we show that whereas the object agent does not logically distinguish between factive and non-factive justifications, such distinctions can be attained at the observer level by analyzing the structure of evidence terms. Basic logic properties of the corresponding two-agent Justification Logic system have been established, which include Kripke-Fitting completeness.

We also argue that a similar evidence-tracking approach can be applied to analyzing paraconsistent systems.

1 Introduction

In this paper, commencing from seminal works [21, 14], the following analysis of basic epistemic notions was adopted: for a given agent,

$$F \textit{ is known} \quad \sim \quad F \textit{ holds in all epistemically possible situations.} \tag{1}$$

The notion of justification, an essential component of epistemic studies, was introduced into the mathematical models of knowledge within the framework of Justification Logic in [1, 2, 3, 5, 6, 8, 13, 16, 18, 19, 22] and other papers; a comprehensive account of this approach is given in [4]. At the foundational level, Justification Logic furnishes a new, evidence-based semantics for the logic of knowledge, according to which

$$F \textit{ is known} \quad \sim \quad F \textit{ has an adequate justification.} \tag{2}$$

*This work was supported by NSF grant 0830450.

Within Justification Logic, we can reason about justifications, simple and compound, and track different pieces of evidence pertaining to the same fact.

In this paper we develop a sufficiently general mechanism of evidence tracking which is crucial for distinguishing between factive and nonfactive justifications. Some preliminary observations leading to this mechanism have been discussed in [4].

1.1 Basics of Justification Logic

Evidence (justification) terms are built from justification variables $x, y, z, \ldots$ and evidence constants $a, b, c, \ldots$ by means of the operations **application** '$\cdot$,' **sum** '$+$,' and **evidence verifier** '!.' The list of operations is flexible: more elaborate justification logic systems also use additional operations on justifications such as negative verifier '?.' On the other hand, it makes sense to consider subsets of operations such as $\{\cdot, +\}$ or even $\{\cdot\}$ (cf. [4]). However, these features do not alter the main results of this paper and, for the sake of convenience, we choose to work with the set of operations $\{\cdot, +, !\}$.

Formulas of Justification Logic are built from logical atomic propositions by means of the usual classical logical connectives $\wedge, \vee, \neg, \ldots$ with an additional formation rule: if t is an evidence term and F a formula, then $t{:}F$ is a formula. Using p to denote any sentence letter and t for an evidence term, we define the formulas by the grammar

$$F = p \mid F \wedge F \mid F \vee F \mid F \rightarrow F \mid \neg F \mid t{:}F \ .$$

The basic introspective justification logic considered in this paper is called $\mathsf{J4}_0$. It contains the following postulates:

1. *Classical propositional axioms and rule Modus Ponens,*
2. *Application Axiom* $s{:}(F \rightarrow G) \rightarrow (t{:}F \rightarrow (s \cdot t){:}G)$,
3. *Monotonicity Axiom* $s{:}F \rightarrow (s + t){:}F$, $s{:}F \rightarrow (t + s){:}F$.
4. *Introspection Axiom* $t{:}F \rightarrow !t{:}(t{:}F)$.

Constants denote justifications of assumptions. To postulate that an axiom A is justified, one has to assume

$$c : A$$

for some evidence constant c. Furthermore, if, in addition, we want to postulate that this new principle $c : A$ is also justified, we can use Introspection and conclude $!c : (c : A)$, etc.

A **Constant Specification** CS for a given logic $\mathcal{L}$ is a set of formulas of form $c : A$ where A's are axioms of $\mathcal{L}$ and c's are evidence constants. We distinguish the following types of constant specifications:

- *axiomatically appropriate*: for each axiom A there is a constant c such that $c : A \in CS$;
- *total* (called TCS): for any axiom A and constant c, $c : A \in CS$.

For a given Constant Specification CS,

$$\mathsf{J4}_{CS} = \mathsf{J4}_0 + CS, \qquad \mathsf{J4} = \mathsf{J4}_0 + TCS.$$

An alternative description of $\mathsf{J4}$ is given by

$$\mathsf{J4} = \mathsf{J4}_0 + \mathrm{R4}$$

where R4 is the *Axiom Internalization Rule*:

For any axiom A and constant c, infer $c : A$.

Finite CS's constitute a representative class of constant specifications: any derivation in $\mathsf{J4}$ may be regarded as a derivation in $\mathsf{J4}_{CS}$ for some finite constant specification CS.

The Deduction Theorem holds in $\mathsf{J4}_{CS}$ for each constant specification CS. The following Internalization property is characteristic for Justification Logic systems.

Theorem 1 [cf. [4]] *For an axiomatically appropriate constant specification CS, $\mathsf{J4}_{CS}$ enjoys Internalization:*

If $\vdash F$, then $\vdash p{:}F$ for some justification term p.

1.2 Epistemic Semantics

A Kripke-Fitting model [13] $\mathcal{M} = (W, R, \mathcal{E}, \Vdash)$ is a Kripke model $(W, R, \Vdash)$ with transitive accessibility relation R (for $\mathsf{J4}$-style systems), augmented by an **admissible evidence function** $\mathcal{E}$ which for any evidence term t and formula F, specifies the set of possible worlds where t is considered admissible evidence for F, $\mathcal{E}(t, F) \subseteq W$. The admissible evidence function $\mathcal{E}$ must satisfy the closure conditions with respect to operations $\cdot, +, !$ as follows:

- *Application*: $\mathcal{E}(s, F \to G) \cap \mathcal{E}(t, F) \subseteq \mathcal{E}(s \cdot t, G)$,
- *Sum*: $\mathcal{E}(s, F) \cup \mathcal{E}(t, F) \subseteq \mathcal{E}(s+t, F)$,
- *Verifier*: $\mathcal{E}(t, F) \subseteq \mathcal{E}(!t, t{:}F)$.

In addition, $\mathcal{E}$ should be monotone with respect to R, i.e.,

$$u \in \mathcal{E}(t, F) \text{ and } uRv \text{ yield } v \in \mathcal{E}(t, F).$$

We say that $\mathcal{E}(t, F)$ *holds at a given world* u if $u \in \mathcal{E}(t, F)$.

Given $\mathcal{M} = (W, R, \mathcal{E}, \Vdash)$, the forcing relation $\Vdash$ on all formulas is defined as follows: for $u \in W$,

1. $\Vdash$ respects Boolean connectives at each world;
2. $u \Vdash t{:}F$ iff $v \Vdash F$ for every $v \in W$ with uRv (the usual Kripke condition) and $u \in \mathcal{E}(t, F)$.

According to this definition, the admissible evidence function $\mathcal{E}$ may be regarded as a Fagin-Halpern-style awareness function [12], but equipped with the structure of justifications.

A model $\mathcal{M} = (W, R, \mathcal{E}, \Vdash)$ *respects a Constant Specification CS* if $\mathcal{E}(c, A) = W$ for all formulas $c{:}A$ from CS.

Theorem 2 [cf. [4]] *For any Constant Specification CS,* $\mathsf{J4}_{CS}$ *is sound and complete for the corresponding class of Kripke-Fitting models respecting CS.*

The information about Kripke structure in Kripke-Fitting models can be completely encoded by the admissible evidence function; this feature is captured by *Mkrtychev models*, which are Kripke-Fitting models with a single world. Naturally, the condition of monotonicity of the evidence function $\mathcal{E}$ with respect to the accessibility relation R becomes void in Mkrtychev models.

Theorem 3 *For any Constant Specification CS,* $\mathsf{J4}_{CS}$ *is sound and complete for the class of Mkrtychev models respecting CS.*

Mkrtychev models play an important theoretical role in establishing decidability and complexity bounds in Justification Logic [4, 8, 15, 16, 17, 18, 19]. Kripke-Fitting models take into account both epistemic Kripke structure and evidence structure and can be useful as natural models of epistemic scenarios.

Corollary 1 [cf. [4]] *For any constant specification CS,* $\mathsf{J4}_{CS}$ *is consistent and has a model.*

1.3 Correspondence between modal and justification logics

The natural modal epistemic counterpart of the evidence assertion $t{:}F$ is $\Box F$ read as

$$\textit{for some } x,\ x{:}F.$$

This observation leads to the notion of *forgetful projection* which replaces each occurrence of $t{:}F$ by $\Box F$ and hence converts a Justification Logic sentence S to a corresponding Modal Logic sentence S^o. Obviously, different Justification Logic sentences may have the same forgetful projection, hence S^o loses certain information that was contained in S. However, it is easily observed that the forgetful projection always maps valid formulas of Justification Logic (e.g., axioms of $\mathsf{J4}$) to valid formulas of a corresponding Epistemic Logic (which in our case is $\mathsf{K4}$). The converse also holds: any valid formula of Epistemic Logic is a forgetful projection of some valid formula of Justification Logic. This follows from Correspondence Theorem 4. We assume that forgetful projection is naturally extended from sentences to logics.

Theorem 4 [Correspondence Theorem, cf. [4]] $\mathsf{J4}^o = \mathsf{K4}$.

This correspondence holds for other pairs of Justification and Modal systems, cf. [4]. Within the core of the Correspondence Theorem is the Realizaton Theorem.

Theorem 5 [Realization Theorem] *There is an algorithm which, for each modal formula F derivable in* $\mathsf{K4}$*, assigns evidence terms to each occurrence of modality in F in such a way that the resulting formula F^r is derivable in* $\mathsf{J4}$*. Moreover, the realization assigns evidence variables to the negative occurrences of modality in F, thus respecting the existential reading of epistemic modality.*

The Correspondence Theorem shows that modal logic $\mathsf{K4}$ has an exact Justification Logic counterpart $\mathsf{J4}$. Note that the Realization Theorem is not at all trivial. Known realization algorithms which recover evidence terms in modal theorems use cut-free derivations in the corresponding modal logics [1, 2, 7, 8].

2 Russell's Example: Induced Factivity

In this paper we offer a Justification Logic technique of handling different justifications for the same fact, e.g., when some of the justifications are factive and some are not. We will formalize and analyze Russell's well-known example from [20].

> If a man believes that the late Prime Minister's last name began with a 'B,' he believes what is true, since the late Prime Minister was Sir Henry Campbell Bannerman[1]. But if he believes that Mr. Balfour was the late Prime Minister[2], he will still believe that the late Prime Minister's last name began with a 'B,' yet this belief, though true, would not be thought to constitute knowledge.

Here we have to deal with two justifications for a true statement, one which is correct and one which is not. Let B be a sentence (propositional atom), w be a designated evidence variable for the wrong reason for B and r a designated evidence variable for the right (hence factive) reason for B. Then, Russell's example prompts the following set of assumptions[3]:

$$\mathcal{R} = \{w{:}B,\ r{:}B,\ r{:}B \rightarrow B\}.$$

Somewhat counter to our intuition, we can logically deduce factivity of w from $\mathcal{R}$:

1. $r{:}B$ - *an assumption*;
2. $r{:}B \rightarrow B$ - *an assumption*;
3. B - *from 1 and 2, by Modus Ponens*;
4. $B \rightarrow (w{:}B \rightarrow B)$ - *a propositional axiom*;
5. $w{:}B \rightarrow B$ - *from 3 and 4, by Modus Ponens*.

The question is, how can we distinguish the 'real' factivity of $r{:}B$ from the 'induced factivity' of $w{:}B$ when the agent can deduce both sentences $r{:}B \rightarrow B$ and $w{:}B \rightarrow B$? The intuitive answer lies in the fact that the derivation $w{:}B \rightarrow B$ is based on the factivity of r for B. Some sort of evidence-tracking mechanism is needed here to formalize this argument.

3 Two-agent setting: observer and object agent

Let us call 'a man' from Russell's example the *object agent*. Tracking object agent reasoning does not appear to be sufficient since the object agent can easily derive the factivity of w for B from the bare assumption $s{:}B$ for any s. Indeed,

[1] Which was true back in 1912. There is a linguistical problem with this example. The correct spelling of this person's last name is Campbell-Bannerman; strictly speaking, this name begins with a 'C.'

[2] Which was false in 1912.

[3] Here we ignore a possible objection that the justifications 'the late Prime Minister was Sir Henry Campbell Bannerman' and 'Mr. Balfour was the late Prime Minister' are mutually exclusive since there could be only one Prime Minister at a time. If the reader is not comfortable with this, we suggest a slight modification of Russell's example in which 'Prime Minister' is replaced by 'member of the Cabinet.' The compatibility concern then disappears since justifications 'X was the member of the late Cabinet' and 'Y was the member of the late Cabinet' with different X and Y are not necessarily incompatible.

1. $s{:}B$ - *an assumption*;
2 $B \rightarrow (w{:}B \rightarrow B)$ - *a propositional axiom*;
3. $c{:}[B \rightarrow (w{:}B \rightarrow B)]$ - *constant specification of 2*;
4. $s{:}B \rightarrow (c\cdot s){:}[w{:}B \rightarrow B]$ - *from 3, by application axiom and Modus Ponens*;
5. $(c\cdot s){:}[w{:}B \rightarrow B]$ *from 1 and 4, by Modus Ponens.*

It takes an outside observer to make such a distinction. More precisely, tracking the reasoning of the object agent, an outside observer can detect the induced character of the factivity of $w:B$. Note that we need Justification (vs. Modal) Logic at both levels: the object agent, since he faces different justifications of the same fact, and the observer, since we need to track the observer's evidence.

So, we consider a setting with the object agent and the observer possessing a justification system obeying $\mathsf{J4}$ and hence not necessarily factive. The fact that the latter is the observer is reflected by the condition that all assumptions of the object agent and assumptions $\mathcal{R}$ are known to the observer.

Here is the formal definition of system $\mathsf{J4(J4)}$. The language contains two disjoint sets of evidence terms built from variables $x, y, z, \ldots$ and constants $a, b, c, \ldots$ for the observer; variables $u, v, w, \ldots$ and constants $k, l, m, \ldots$ for the object agent. We will not be using different symbols for similar operations 'application' and 'sum' for the observer and the object agent, and hope this will not lead to ambiguity. However, for better readability, we will be using different notation for evidence assertions:

$$[\![s]\!]F \quad \sim \quad \textit{s is a justification of F for the observer},$$

$$t{:}F \quad \sim \quad \textit{t is a justification of F for the object agent.}$$

Using p to denote any sentence letter, s for an evidence term of the observer, and t for an evidence term of the object agent, we define the formulas of $\mathsf{J4(J4)}$ by the grammar

$$F = p \mid F \wedge F \mid F \vee F \mid F \rightarrow F \mid \neg F \mid t{:}F \mid [\![s]\!]F \ .$$

The list of postulates of $\mathsf{J4(J4)}$ contains the following principles:

1. **Classical propositional axioms and rule Modus Ponens**

2. **Axioms of $\mathsf{J4}$ (including Total Constant Specification) for the Object Agent:**

   [A1] *Application* $t_1{:}(F \rightarrow G) \rightarrow (t_2{:}F \rightarrow (t_! \cdot t_2){:}G)$,

   [A2] *Monotonicity* $t_1{:}F \rightarrow (t_1 + t_2){:}F,\ t_2{:}F \rightarrow (t_1 + t_2){:}F$,

   [A3] *Verification* $t{:}F \rightarrow\ !t{:}t{:}F$,

   [A4] *Total Constant Specification for* 1, A1–A3:

   $$TCS_A = \{k{:}F \mid\ k \text{ is an object agent evidence constant and } F \text{ is from 1, A1–A3}\},$$

3. **Similar axioms of $\mathsf{J4}$ (including Total Constant Specification) for the Observer**

   [O1] *Application* $[\![s_1]\!](F \to G) \to ([\![s_2]\!]F \to [\![s_1 \cdot s_2]\!]G)$,

   [O2] *Monotonicity* $[\![s_1]\!]F \to [\![s_1 + s_2]\!]F,\ [\![s_2]\!]F \to [\![s_1 + s_2]\!]F$,

   [O3] *Verification* $[\![s]\!]F \to [\![!s]\!][\![s]\!]F$,

   [O4] *Total Constant Specification for* 1, O1–O3:

   $$TCS_O = \{[\![a]\!]F \mid\ a \text{ is an observer evidence constant and } F \text{ is from 1, O1–O3}\},$$

4. **Total Constant Specification for the observer of the object agent postulates**:

   $$TCS_{OA} = \{[\![b]\!]F \mid\ b \text{ is an observer evidence constant and } F \text{ is from 1, A1–A4}\}.$$

System $\mathsf{J4}(\mathsf{J4})$ provides a general setup with the object agent and the observer, both of reasoning type $\mathsf{J4}$, but does not yet reflect the specific structure of Russell's Prime Minister example.

4 Some Model Theory

A Kripke-Fitting model for $\mathsf{J4}(\mathsf{J4})$ is

$$\mathcal{M} = (W, R_A, R_O, \mathcal{E}_A, \mathcal{E}_O, \Vdash)$$

such that

$(W, R_A, \mathcal{E}_A, \Vdash)$ is a $\mathsf{J4}$-model which respects TCS_A;

$(W, R_O, \mathcal{E}_O, \Vdash)$ is a $\mathsf{J4}$-model which respects TCS_O and TCS_{OA}.

Soundness of $\mathsf{J4}(\mathsf{J4})$ with respect to these models is straightforward and follows from the soundness of $\mathsf{J4}$.

Theorem 6 $\mathsf{J4}(\mathsf{J4})$ *is complete with respect to the class of* $\mathsf{J4}(\mathsf{J4})$*-models.*

Proof. By the standard maximal consistent set construction. Let W be the set of all maximal consistent sets of $\mathsf{J4}(\mathsf{J4})$-formulas;

$$\Gamma R_A \Delta \quad \textit{iff} \quad \{F \mid t{:}F \in \Gamma \textit{ for some } t\} \subseteq \Delta;$$

$$\Gamma R_O \Delta \quad \textit{iff} \quad \{F \mid [\![s]\!]F \in \Gamma \textit{ for some } s\} \subseteq \Delta;$$

$$\Gamma \in \mathcal{E}_A(t, F) \quad \textit{iff} \quad t{:}F \in \Gamma;$$

$$\Gamma \in \mathcal{E}_O(s, F) \quad \textit{iff} \quad [\![s]\!]F \in \Gamma;$$

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

First, we notice that R_A and R_O transitive. Second, we check the closure conditions as well as the monotonicity for $\mathcal{E}_A$ and $\mathcal{E}_O$. These are all rather standard checkups, performed in the same way as the completeness proof for $\mathsf{J4}$ (cf. [4]). Finally, we have to check that evidence functions $\mathcal{E}_A$and $\mathcal{E}_O$ respect the corresponding constant specifications of $\mathsf{J4}(\mathsf{J4})$. This is secured by the definition of the evidence functions, since

$$TCS_A \cup TCS_O \cup TCS_{OA} \subseteq \Gamma$$

for every maximal consistent set Γ.

Lemma 1 [Truth Lemma] *For each formula F and world $\Gamma \in W$,*

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

Proof. The proof is also rather standard and proceeds by induction on F. The base case holds by the definition of the forcing relation $\Vdash$; Boolean connectives are straightforward. Let F be $t{:}G$. If $t{:}G \in \Gamma$, then $\Gamma \in \mathcal{E}_A(t, G)$; moreover, by the definition of R_A, $G \in \Delta$ for each Δ such that $\Gamma R_A \Delta$. By the Induction Hypothesis, $\Delta \Vdash G$, therefore, $\Gamma \Vdash t{:}G$. If $t{:}G \notin \Gamma$, then $\Gamma \notin \mathcal{E}_A(t, G)$ and $\Gamma \nVdash t{:}G$.

The Induction step in case $F = [\![s]\!]G$ is considered in a similar way. □

Corollary 2 *TCS_A, TCS_O, and TCS_{OA} hold at each node.*

Indeed, $TCS_A \cup TCS_O \cup TCS_{OA} \subseteq \Gamma$ since Γ contains all postulates of $\mathsf{J4}(\mathsf{J4})$. By the Truth Lemma, $\Gamma \Vdash TCS_A \cup TCS_O \cup TCS_{OA}$.

To complete the proof of Theorem 6, consider F which is not derivable in $\mathsf{J4}(\mathsf{J4})$. The set $\{\neg F\}$ is therefore consistent. By the standard Henkin construction, $\{\neg F\}$ can be extended to a maximal consistent set Γ. Since $F \notin \Gamma$, by the Truth Lemma, $\Gamma \nVdash F$. □

5 Distinguishing induced factivity

Russell's Prime Minister example can be formalized over $\mathsf{J4}(\mathsf{J4})$ by the set of assumptions $\mathcal{R}$ and $\mathcal{IR}$: the latter stands for 'Internalized Russell'

$$\mathcal{IR} = \{[\![x]\!]r{:}B,\ [\![y]\!](r{:}B \rightarrow B),\ [\![z]\!]w{:}B\}.$$

Here B, r, and w are as in $\mathcal{R}$, and x, y, z are designated proof variables for the observer.

First, we check that the observer knows the factivity of w for B, e.g., that

$$\mathsf{J4}(\mathsf{J4}) + \mathcal{R} + \mathcal{IR} \ \vdash [\![s]\!](w{:}B \rightarrow B)$$

for some proof term s. Here is the derivation, which is merely the internalization of the corresponding derivation from Section 2:

1. $[\![x]\!]r{:}B$ - *an assumption*;
2. $[\![y]\!](r{:}B \rightarrow B)$ - *an assumption*;
3. $[\![y{\cdot}x]\!]B$ - *from 1 and 2, by application*;
4. $[\![a]\!][B \rightarrow (w{:}B \rightarrow B)]$ - *by TCS_O for a propositional axiom*;
5. $[\![a{\cdot}(y{\cdot}x)]\!](w{:}B \rightarrow B)$ - *from 3 and 4, by application.*

Finally, let us establish that the observer cannot conclude $w{:}B \rightarrow B$ other than by using the factivity of r. In our formal setting, this amounts to proving the following theorem.

Theorem 7 *If*

$$\mathsf{J4}(\mathsf{J4}) + \mathcal{R} + \mathcal{IR} \ \vdash [\![\tilde{s}]\!](w{:}B \rightarrow B),$$

then term $\tilde{s}$ contains both proof variables x and y.

Proof. Following [15], we axiomatize the *reflected fragment* of $\mathsf{J4(J4)} + \mathcal{R} + \mathcal{IR}$ consisting of all formulas $[\![s]\!]F$ derivable in $\mathsf{J4(J4)} + \mathcal{R} + \mathcal{IR}$.

The principal tool here is the so-called $*$-calculus (cf. [19, 15]).

Calculus

$$*[\mathsf{J4(J4)} + \mathcal{R} + \mathcal{IR}]$$

has **axioms** $TCS_O \cup TCS_{OA} \cup \mathcal{IR}$ and rules of inference

Application: given $[\![s_1]\!](F \to G)$ and $[\![s_2]\!]F$, derive $[\![s_1 \cdot s_2]\!]G$;

Sum: given $[\![s_1]\!]F$, derive $[\![s_1 + s_2]\!]F$ or $[\![s_2 + s_1]\!]F$;

Proof Checker: given $[\![s]\!]F$, derive $[\![!s]\!][\![s]\!]F$.

The following Lemma connects $\mathsf{J4(J4)} + \mathcal{R} + \mathcal{IR}$ and its reflected fragment $*[\mathsf{J4(J4)} + \mathcal{R} + \mathcal{IR}]$.

Lemma 2 *For any formula* $[\![s']\!]F$,

$$\mathsf{J4(J4)} + \mathcal{R} + \mathcal{IR} \vdash [\![s']\!]F \quad \textit{iff} \quad * [\mathsf{J4(J4)} + \mathcal{R} + \mathcal{IR}] \vdash [\![s']\!]F.$$

Proof. It is obvious that if $*[\mathsf{J4(J4)} + \mathcal{R} + \mathcal{IR}] \vdash [\![s']\!]F$, then $\mathsf{J4(J4)} + \mathcal{R} + \mathcal{IR} \vdash [\![s']\!]F$. Indeed, all axioms of $*[\mathsf{J4(J4)} + \mathcal{R} + \mathcal{IR}]$ are provable in $\mathsf{J4(J4)} + \mathcal{R} + \mathcal{IR}$; the rules of the former correspond to axioms of the latter.

In order to establish the converse, let us suppose that $*[\mathsf{J4(J4)} + \mathcal{R} + \mathcal{IR}] \nvdash [\![s']\!]F$. We build a singleton $\mathsf{J4(J4)}$-model $\mathcal{M} = (W, R_A, R_O, \mathcal{E}_A, \mathcal{E}_O, \Vdash)$ in which $\mathcal{R} \cup \mathcal{IR}$ holds but $[\![s']\!]F$ does not: this will be sufficient to conclude that $\mathsf{J4(J4)} + \mathcal{R} + \mathcal{IR} \nvdash [\![s']\!]F$.

- $W = \{1\}$;
- $R_A = \emptyset$, $R_O = \{(1, 1)\}$;
- $\mathcal{E}_A(t, G) = W$ for each t, G;
- $\mathcal{E}_O(s, G)$ holds at 1 iff $*[\mathsf{J4(J4)} + \mathcal{R} + \mathcal{IR}] \vdash [\![s]\!]G$;
- $1 \Vdash p$ for all propositional variables, including B.

Note that R_A and R_O are transitive. Let us check the closure properties of the evidence functions. $\mathcal{E}_A$ is universal and hence closed. $\mathcal{E}_O$ is closed under *application*, *sum*, and *verifier* since the calculus $*[\mathsf{J4(J4)} + \mathcal{R} + \mathcal{IR}]$ is.

Monotonicity of $\mathcal{E}_A$ and $\mathcal{E}_O$ vacuously hold since W is a singleton.

Furthermore, $TCS_{A,O,OA}$ hold in $\mathcal{M}$. To check this, we first note that since $R_A = \emptyset$, a formula $t{:}G$ holds at 1 if and only if $\mathcal{E}_A(t, G)$. Therefore, all formulas $t{:}G$ hold at 1, in particular, $1 \Vdash TCS_A$. Hence all axioms A1–A4 of $\mathsf{J4(J4)}$ hold at 1. This yields that $1 \Vdash TCS_{OA}$. Indeed, for each $[\![c]\!]A \in TCS_{OA}$, $1 \Vdash A$ (just established) and $\mathcal{E}_O(c, A)$, since $[\![c]\!]A$ is an axiom of $*[\mathsf{J4(J4)} + \mathcal{R} + \mathcal{IR}]$. By the same reasons, $1 \Vdash \mathcal{R}$.

Since $*[\mathsf{J4(J4)} + \mathcal{R} + \mathcal{IR}] \vdash TCS_O$, $\mathcal{E}_O(c, A)$ holds for all $[\![c]\!]A \in TCS_O$. In addition, each such A is an axiom O1–O3, hence $1 \Vdash A$. Therefore, $1 \Vdash TCS_O$. By similar reasons, $1 \Vdash \mathcal{IR}$.

We have just established that $\mathcal{M}$ is a model for $\mathsf{J4(J4)} + \mathcal{R} + \mathcal{IR}$.

We claim that $\mathcal{M} \not\Vdash [\![s']\!]F$ which follows immediately from the assumption that

$$*[\mathsf{J4}(\mathsf{J4}) + \mathcal{R} + \mathcal{IR}] \not\vdash [\![s']\!]F$$

since then $\mathcal{E}_O(s', F)$ does not hold at 1. Therefore,

$$\mathsf{J4}(\mathsf{J4}) + \mathcal{R} + \mathcal{IR} \not\vdash [\![s']\!]F \ .$$

This concludes the proof of Lemma 2. □

Lemma 3 [Subterm property of $*$-derivations] *In a tree-form derivation of a formula* $[\![s]\!]F$ *in* $*[\mathsf{J4}(\mathsf{J4}) + \mathcal{R} + \mathcal{IR}]$, *if* $[\![s']\!]G$ *is derived at some node, then* s' *is a subterm of* s.

Proof. Obvious, from the fact that all rules of $*[\mathsf{J4}(\mathsf{J4})+\mathcal{R}+\mathcal{IR}]$ have such a subterm property. □

Lemma 4 *If* $*[\mathsf{J4}(\mathsf{J4}) + \mathcal{R} + \mathcal{IR}] \vdash [\![\tilde{s}]\!](w{:}B \to B)$, *then term* $\tilde{s}$ *contains* x.

Proof. Suppose the opposite, i.e., that $\tilde{s}$ does not contain x. Then, by the subterm property, the proof of $[\![\tilde{s}]\!](w{:}B \to B)$ in $*[\mathsf{J4}(\mathsf{J4}) + \mathcal{R} + \mathcal{IR}]$ does not use axiom $[\![x]\!]r{:}B$. Moreover, since $*[\mathsf{J4}(\mathsf{J4}) + \mathcal{R} + \mathcal{IR}]$ does not really depend on $\mathcal{R}$, $[\![\tilde{s}]\!](w{:}B \to B)$ is derivable without $\mathcal{R}$ and $[\![x]\!]r{:}B$. Since such a proof can be replicated in $\mathsf{J4}(\mathsf{J4}) + \mathcal{IR}$ without $[\![x]\!]r{:}B$, it should be the case that

$$\mathsf{J4}(\mathsf{J4}) + [\![y]\!](r{:}B \to B) + [\![z]\!]w{:}B \vdash [\![\tilde{s}]\!](w{:}B \to B).$$

To get a contradiction, it now suffices to build a $\mathsf{J4}(\mathsf{J4})$-model $\mathcal{M} = (W, R_A, R_O, \mathcal{E}_A, \mathcal{E}_O, \Vdash)$ in which $[\![y]\!](r{:}B \to B)$ and $[\![z]\!]w{:}B$ hold, but $[\![\tilde{s}]\!](w{:}B \to B)$ does not. Here is the model:

- $W = \{1\}$;
- $R_A = \emptyset$, $R_O = \{(1, 1)\}$;
- $\mathcal{E}_A(r, B) = \emptyset$ and $\mathcal{E}_A(t, F) = W$ for all other pairs t, F;
- $\mathcal{E}_O(s, G) = W$ for all s, G;
- $1 \not\Vdash p$ for all propositional variables, including B.

First, we check that $\mathcal{M}$ is a $\mathsf{J4}(\mathsf{J4})$-model. Closure and monotonicity conditions on $R_A, R_O, \mathcal{E}_A, \mathcal{E}_O$ are obviously met. We claim that $TCS_{A,O,OA}$ hold in $\mathcal{M}$. Since $R_A = \emptyset$, a formula $t{:}F$ holds at 1 if and only if $\mathcal{E}_A(t, F)$ holds at 1. Therefore, $1 \not\Vdash r{:}B$ and $1 \Vdash t{:}F$ for all other pairs t, and F. In particular,

$$1 \Vdash TCS_A.$$

Since $R_O = \{(1, 1)\}$ and $\mathcal{E}_O(s, G) = W$, for any observer evidence term s, $1 \Vdash [\![s]\!]G$ if and only if $1 \Vdash G$. All observer axioms hold at 1 and hence

$$1 \Vdash TCS_O.$$

As we have shown, $1 \Vdash TCS_A$ and hence

$$1 \Vdash TCS_{OA}.$$

Furthermore, since $1 \Vdash r{:}B \to B$,

$$1 \Vdash [\![y]\!](r{:}B \to B),$$

and since $1 \Vdash w{:}B$,

$$1 \Vdash [\![z]\!]w{:}B.$$

Finally, since $1 \nVdash w{:}B \to B$,

$$1 \nVdash [\![\tilde{s}]\!](w{:}B \to B).$$

□

Lemma 5 *If* $*[\mathsf{J4}(\mathsf{J4}) + \mathcal{R} + \mathcal{IR}] \vdash [\![\tilde{s}]\!](w{:}B \to B)$*, then term* $\tilde{s}$ *contains* y*.*

Proof. Suppose the opposite, i.e., that $\tilde{s}$ does not contain y. Then, by the subterm property, the derivation of $[\![\tilde{s}]\!](w{:}B \to B)$ in $*[\mathsf{J4}(\mathsf{J4}) + \mathcal{R} + \mathcal{IR}]$ does not use axiom $[\![y]\!](r{:}B \to B)$. From this, we can find a derivation of $[\![\tilde{s}]\!](w{:}B \to B)$ in

$$\mathsf{J4}(\mathsf{J4}) + [\![x]\!]r{:}B + [\![z]\!]w{:}B.$$

To obtain a contradiction, it suffices to present a $\mathsf{J4}(\mathsf{J4})$-model $\mathcal{M} = (W, R_A, R_O, \mathcal{E}_A, \mathcal{E}_O, \Vdash)$ in which $[\![x]\!]r{:}B$ and $[\![z]\!]w{:}B$ hold, but $[\![\tilde{s}]\!](w{:}B \to B)$ does not hold. Here is this model:

- $W = \{1\}$;
- $R_A = \emptyset$, $R_O = \{(1,1)\}$;
- $\mathcal{E}_A(t, F) = W$ for all t, F;
- $\mathcal{E}_O(s, G) = W$ for all s, G;
- $1 \nVdash p$ for all propositional variables, including B.

Conditions on $R_A, R_O, \mathcal{E}_A, \mathcal{E}_O$ are obviously met. Let us check constant specifications of $\mathsf{J4}(\mathsf{J4})$. Since $R_A = \emptyset$, and $\mathcal{E}_A(t, F)$ holds at 1 for all t, F, $t{:}F$ holds at for all t, F. In particular,

$$1 \Vdash TCS_A.$$

For the same reasons, $1 \Vdash r{:}B$ and $1 \Vdash w{:}B$.

Furthermore, $1 \Vdash [\![s]\!]F$ if and only if $1 \Vdash F$, because $\mathcal{E}_O(s, F) = \{1\}$ and $R_O = \{(1,1)\}$. Therefore,

$$1 \Vdash TCS_{OA}.$$

Since all axioms O1–O3 are true at 1,

$$1 \Vdash TCS_O.$$

Finally, since $1 \Vdash r{:}B$ and $1 \Vdash w{:}B$,

$$1 \Vdash [\![x]\!]r{:}B \text{ and } 1 \Vdash [\![z]\!]w{:}B.$$

It remains to establish that $1 \nVdash [\![\tilde{s}]\!](w{:}B \to B)$, for which it suffices to check that $1 \nVdash w{:}B \to B$, which is the case since $1 \Vdash w{:}B$ and $1 \nVdash B$. □

Theorem 7 now follows from Lemmas 2, 4, and 5. □

5.1 Observer's factivity

Another natural candidate for the observer logic is the Logic of Proofs LP (cf. [2, 4, 13]) which is J4 augmented by the *Factivity Axiom*

$$[\![s]\!]F \to F,$$

with the corresponding extension of constant specifications to include constants corresponding to this axiom. Kripke-Fitting models for LP are J4-models with a reflexive accessibility relation.

An assumption that the observer (the reader, for example) is LP-compliant is quite reasonable since, according to [2], the Logic of Proofs LP is a universal logic of mathematical reasoning for a wide range of natural formal systems (knowers)[4]. So we could therefore define a two-agent system LP(J4) and proceed with the same evidence-tracking analysis. The main result: an analogue of Theorem 7 and its proof hold for LP(J4). In particular, all models built in the proof of Theorem 7 are intentionally made reflexive with respect to the observer's accessibility relation so they fit for LP as the observer's logic.

6 Conclusions

The formalization of Russell's example given in this paper can obviously be extended to other situations with multiple justifications of the same facts. The principal technique consisting of

- introducing the observer and a two-layer reasoning system;
- working in the reflected fragment of the observer's reasoning;
- formalizing dependencies of assumptions via variable occurrences in proof terms;
- reasoning in Fitting/Mkrtychev models for formally establishing independence,

is of a general character and can be useful for evidence-tracking in a general setting.

On the other hand, the whole power of J4(J4) is not needed for Russell's example, e.g., there is no use of '+' operations here. However, we have intentionally considered J4(J4) in its entirety to introduce a basic introspective system of evidence-tracking.

[4]As was shown in [9], the same LP serves as the logic of proofs for polynomially bounded agents as well.

Verification principles for both the object agent and the observer have been used only to simplify formulations of Constant Specifications. The same evidence tracking can be done within the framework of the basic justification logic J for both the object agent and the observer. Moreover, Theorem 7 and its proof hold for a wide range of systems, e.g., J, J4, J45 for the object agent and, independently, J, JT, J4, LP, J45, JT45, etc. (cf. [4] for the definitions) for the observer.

It appears that a similar evidence-tracking approach can be applied to analyzing paraconsistent systems. For example, the set of formulas A[5]

$$A = \{p_1,\ \ p_1 \to p_2,\ \ p_2 \to p_3,\ \ \ldots,\ \ p_{n-1} \to p_n,\ \ \neg p_n\}$$

is obviously inconsistent. However, any derivation from A which does not use **all** $n+1$ assumptions of A is contradiction-free. This argument can be naturally formalized in Justification Logic.

We wish to think that this approach to evidence tracking could be also useful in distributed knowledge systems (cf. [10, 11, 12]).

7 Acknowledgements

The author is very grateful to Mel Fitting, Vladimir Krupski, Roman Kuznets, and Elena Nogina, whose advice helped with this paper. Many thanks to Karen Kletter for editing this text.

References

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

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

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

[4] S. Artemov. The Logic of Justification. *The Review of Symbolic Logic*, 1(4):477-513, 2008.

[5] S. Artemov and R. Kuznets. Logical omniscience as a computational complexity problem. In *Aviad Heifetz, editor, Theoretical Aspects of Rationality and Knowledge, Proceedings of the Twelfth Conference (TARK 2009), Stanford University, California, July 6–8, 2009. ACM.* , pages 14–23, 2009.

[6] S. Artemov and E. Nogina. Introducing justification into epistemic logic. *Journal of Logic and Computation*, 15(6):1059–1073, 2005.

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

[8] V. Brezhnev and R. Kuznets. Making knowledge explicit: How hard it is. *Theoretical Computer Science*, 357(1-3):23–34, 2006.

[5] Here $p_1, p_2, \ldots, p_n$ are propositional letters.

[9] E. Goris. Logic of proofs for bounded arithmetic. In D. Grigoriev, J. Harrison, and E.A. Hirsch, editors, *Computer Science - Theory and Applications. CSR 2006*, volume 3967 of *Lecture Notes in Computer Science*, pages 191–201. Springer, 2006.

[10] Y. Gurevich and I. Neeman. DKAL: Distributed-Knowledge Authorization Language. *21st IEEE Computer Security Foundations Symposium (CSF 2008)*, pp. 149-162, 2008.

[11] Y. Gurevich and I. Neeman. The Infon Logic. *Bulletin of European Association for Theoretical Computer Science*, 98:150–178, 2009.

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

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

[14] J. Hintikka. *Knowledge and Belief.* Cornell University Press, Ithaca, 1962.

[15] N.V. Krupski. On the complexity of the reflected logic of proofs. *Theoretical Computer Science*, 357(1):136–142, 2006.

[16] R. Kuznets. On the complexity of explicit modal logics. In *Computer Science Logic 2000*, volume 1862 of *Lecture Notes in Computer Science*, pages 371–383. Springer-Verlag, 2000.

[17] R. Kuznets. *Complexity Issues in Justification Logic.* Ph.D. thesis, CUNY Graduate Center, 2008. http://sites.google.com/site/kuznets/PhD.pdf

[18] R. Milnikel. Derivability in certain subsystems of the Logic of Proofs is Π_2^p-complete. *Annals of Pure and Applied Logic*, 145(3):223–239, 2007.

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

[20] B. Russell. *The Problems of Philosophy.* London: Williams and Norgate; New York: Henry Holt and Company, 1912.

[21] G.H. von Wright. *An essay in modal logic.* North-Holland, Amsterdam, 1951.

[22] T. Yavorskaya (Sidon). Multi-agent Explicit Knowledge. In D. Grigoriev, J. Harrison, and E.A. Hirsch, editors, *Computer Science - Theory and Applications. CSR 2006*, volume 3967 of *Lecture Notes in Computer Science*, pages 369–380. Springer, 2006.