2008

TR-2008010: The Logic of Justification

Sergei Artemov

The Logic of Justification

Sergei Artemov[*]

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

September 8, 2008

Abstract

We describe a general logical framework, Justification Logic, for reasoning about epistemic justification. Justification Logic is based on classical propositional logic augmented by justification assertions $t{:}F$ that read *t is a justification for F*. Justification Logic absorbs basic principles originating from both mainstream epistemology and the mathematical theory of proofs. It contributes to the studies of the well-known *Justified True Belief vs. Knowledge* problem. We state a general Correspondence Theorem showing that behind each epistemic modal logic, there is a robust system of justifications. This renders a new, evidence-based foundation for epistemic logic.

As a case study, we offer a resolution of the Goldman-Kripke 'Red Barn' paradox and analyze Russell's 'prime minister example' in Justification Logic. Furthermore, we formalize the well-known Gettier example and reveal hidden assumptions and redundancies in Gettier's reasoning.

1 Introduction

The celebrated account of Knowledge as Justified True Belief commonly attributed to Plato (cf. [29; 34]) was widely accepted until 1963 when a paper by Edmund Gettier [29] opened the door to a broad philosophical discussion of the subject (cf. [19; 32; 45; 53; 62] and many others).

Meanwhile, commencing from seminal works [37; 67], the notions of Knowledge and Belief have acquired formalization by means of modal logic with atoms $\mathbf{K}F$ (*F is known*) and $\mathbf{B}F$ (*F is believed*). Within this approach, the following analysis was adopted: for a given agent,

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

[*]This work has been supported by NSF grant 0830450, CUNY Collaborative Incentive Research Grant CIRG1424, and PSC CUNY Research Grant PSCREG-39-721

The resulting *Epistemic Logic* has been remarkably successful in terms of developing a rich mathematical theory and applications (cf. [23; 48], and other sources). However, the notion of justification, which has been an essential component of epistemic studies, was conspicuously absent in the mathematical models of knowledge within the epistemic logic framework. This deficiency is displayed most prominently, in the *Logical Omniscience* defect of the modal logic of knowledge (cf. [21; 22; 38; 51; 56]). In the provability domain, the absence of an adequate description of the logic of justifications (here mathematical proofs) remained an impediment to both formalizing the Brouwer-Heyting-Kolmogorov semantics of proofs and providing a long-anticipated exact provability semantics for Gödel's provability logic $\mathsf{S4}$ and intuitionistic logic ([3; 4; 6; 66]). This lack of a justification component has, perhaps, contributed to a certain gap between epistemic logic and mainstream epistemology ([34; 35]). We would like to think that Justification Logic is a step towards filling this void.

The contribution of this paper to epistemology can be briefly summarized as follows.

> *We describe basic logical principles for justifications and relate them to both mainstream and formal epistemology. The result is a long-anticipated mathematical notion of justification, making epistemic logic more expressive. We now have the capacity to reason about justifications, simple and compound. We can compare different pieces of evidence pertaining to the same fact. We can measure the complexity of justifications, which leads to a coherent theory of logical omniscience. Justification Logic provides a novel, evidence-based mechanism of truth-tracking which seems to be a key ingredient of the analysis of knowledge. Finally, Justification Logic furnishes a new, evidence-based foundation for the logic of knowledge, according to which*
>
> $$F \textit{ is known} \qquad \sim \qquad F \textit{ has an adequate justification.} \tag{2}$$

There are several natural interpretations of Justification Logic. Justification assertions of the format $t{:}F$ read generically as

$$t \textit{ is a justification of } F. \tag{3}$$

There is also a more strict 'justificationist' reading in which $t{:}F$ is understood as

$$t \textit{ is accepted by agent as a justification of } F. \tag{4}$$

The language and tools of Justification Logic accommodate both readings of $t:F$. Moreover, Justification Logic is general enough to incorporate other semantics that are not necessarily terminologically related to justifications or proofs. For example, $t{:}F$ can be read as

$$t \textit{ is a sufficient resource for } F. \tag{5}$$

Tudor Protopopescu suggests that $t:F$ could also be assigned an externalist, non-justificationist reading, something like

$$F \textit{ satisfies conditions } t. \tag{6}$$

In this setting, t would be something like a set of causes or counterfactuals. Such a reading would still maintain the distinction between partial and factive justifications, since t may not be all that is required for *belief that* F to count as *knowledge that* F.

Within Justification Logic, we do not directly analyze what it means for t to justify F beyond the format $t{:}F$, but rather attempt to characterize this relation axiomatically. This is similar to the way Boolean logic treats its connectives, say, disjunction: it does not analyze the formula $p \vee q$ but rather assumes certain logical axioms and truth tables about this formula.

There are several design decisions made for this installment of Justification Logic.

1. We decide to limit our attention, at this stage, to *propositional* and *quantifier-free* systems of Justification Logic, and leave quantified systems for further study.

2. We build our systems on the simplest base: *classical Boolean logic*, though we are completely aware that there are much more elaborate logical models, e.g., intuitionistic and substructural logics, conditionals, relevance logics, and logics of counterfactual reasoning, just to name a few. There are several good reasons for choosing the Boolean logic base here. At this stage, we are concerned first with *justifications*, which provide a sufficiently serious challenge on even the simplest Boolean base. Once this case is sorted out in a satisfactory way, we can move on to incorporating justifications into other logics. Second, the paradigmatic examples which we will consider (e.g., Goldman-Kripke and Gettier), can be handled with Boolean Justification Logic. Third, the core of Epistemic Logic consists of modal systems with a classical Boolean base (K, T, K4, S4, K45, KD45, S5, etc.). We provide each of them with a corresponding Justification Logic companion based on Boolean logic.

3. Within the Justification Logic framework, we treat both partial and factive justifications. This helps to capture the essence of discussion on these matters in epistemology, where justifications are not generally assumed to be factive.

4. In this paper, we consider the case of one agent only, although several multi-agent Justification Logic systems have already been developed ([5; 12; 68]).

Formal logical methods do not directly solve philosophical problems, but rather provide a tool for analyzing assumptions and to ensure that we draw correct conclusions. Our hope is that Justification Logic will do just that.

2 Preliminary Analysis of Principles Involved

In this section, we will survey the Logic of Proofs, Gettier's examples [29], and examine some classical post-Gettier sources to determine what logical principles in the given Justification Logic format (propositional Boolean logic with justification assertions $t:F$) may be extracted. As is usual with converting informally stated principles into formal ones, a certain amount of good will is required. This does not at all mean that the considerations adduced in [19; 32; 45; 53; 62] may be readily formulated in the Boolean Justification Logic. The aforementioned papers are written in natural language, which is richer than any formal one; a more sophisticated formal language could probably provide a better account here, which we leave to future studies.

2.1 The Logic of Proofs

The Logic of Proofs LP was suggested by Gödel in [31] and developed in full in [2; 4]. LP gives a complete axiomatization of the notion of mathematical proof with natural operations 'application,' 'sum,' and 'proof checker.' We discuss these operations below in a more general epistemic setting.

In LP, justifications are represented by *proof polynomials*, which are terms built from *proof variables* $x, y, z, \ldots$ and *proof constants* $a, b, c, \ldots$ by means of two binary operations: *application* '$\cdot$' and *sum* (*union, choice*) '$+$,' and one unary operation *proof checker* '!'. The formulas of LP are those of propositional classical logic augmented by the formation rule: *if t is a proof polynomial and F a formula, then $t{:}F$ is again a formula.*

The Logic of Proofs LP contains the postulates of classical propositional logic and the rule of *Modus Ponens* along with

$$s{:}(F \to G) \to (t{:}F \to (s \cdot t){:}G) \qquad (Application)$$
$$s{:}F \to (s+t){:}F, \quad t{:}F \to (s+t){:}F \qquad (Sum)$$
$$t{:}F \to !t{:}(t{:}F) \qquad (Proof\ Checker)$$
$$t{:}F \to F \qquad (Reflection).$$

Proof constants in LP represent 'atomic' proofs of axioms which are not analyzed any further. In addition to the usual logical properties, such as being closed under substitution and respecting the Deduction Theorem, LP enjoys the Internalization property:

If $\vdash F$, then there is a proof polynomial p such that $\vdash p{:}F$.

2.2 Gettier Examples

Gettier in [29] described two situations, Case I and Case II, that were supposed to provide examples of justified true beliefs which should not be considered knowledge. In this paper we will focus on formalizing Case I, which proved to be more challenging. Case II can be easily formalized in a similar fashion.

Here is a shortened exposition of Case I from [29].

Suppose that Smith and Jones have applied for a certain job. And suppose that Smith has strong evidence for the following conjunctive proposition:

(d) Jones is the man who will get the job, and Jones has ten coins in his pocket.

Proposition (d) entails:

(e) The man who will get the job has ten coins in his pocket.

Let us suppose that Smith sees the entailment from (d) to (e), and accepts (e) on the grounds of (d), for which he has strong evidence. In this case, Smith is clearly justified in believing that (e) is true. But imagine, further, that unknown to Smith, he himself, not Jones, will get the job. And, also, unknown to Smith, he himself has ten coins in his pocket. Then, all of the following are true:

1) (e) is true,

2) Smith believes that (e) is true, and

3) Smith is justified in believing that (e) is true.

But it is equally clear that Smith does not know that (e) is true....

Gettier uses a version of the epistemic closure principle, closure of justification under logical consequence:

> *... if Smith is justified in believing P, ... and Smith deduces Q from P ..., then Smith is justified in believing Q.*

Here is its natural formalization:

Smith is justified in believing P can be formalized as "for some t, $t{:}P$";

Smith deduces Q *from* P — "there is a deduction of $P\to Q$ (available to Smith)";

Smith is justified in believing Q — "$t{:}Q$ for some t."

Such a rule holds for the Logic of Proofs, as well as for all other Justification Logic systems considered in this paper. It is a combination of the Internalization Rule:

$$\textit{if} \vdash F, \textit{ then} \vdash s{:}F \textit{ for some } s \tag{7}$$

and the Application Axiom:

$$s{:}(P\to Q)\to(t{:}P\to(s\cdot t){:}Q). \tag{8}$$

Indeed, suppose $t{:}P$ and there is a deduction of $P\to Q$. By the Internalization Rule, $s{:}(P\to Q)$ for some s. From the Application Axiom, by *Modus Ponens* twice, we get $(s\cdot t){:}Q$.

2.3 Goldman's Reliabilism

Goldman in [32] offered the 'fourth condition' to be added to the Justified True Belief definition of knowledge. According to [32],

> *a subject's belief is justified only if the truth of a belief has caused the subject to have that belief (in the appropriate way), and for a justified true belief to count as knowledge, the subject must also be able to correctly reconstruct (mentally) that causal chain.*

Goldman's principle makes it clear that a justified belief (in our language, a situation t *justifies* F for some t) for an agent occurs **only if** F is true, which provides the Factivity Axiom for 'knowledge-producing' justifications

$$t{:}F\to F \qquad\qquad \textit{(Factivity Axiom).} \tag{9}$$

The Factivity Axiom is assumed for *factive justifications* (systems JT, LP, $\mathsf{JT45}$ below) but not for general justification systems J, $\mathsf{J4}$, $\mathsf{J45}$, $\mathsf{JD45}$.

With a certain amount of good will, we can assume that *the 'causal chain' leading from the truth of* F *to a justified belief that* F manifests itself in the Principle of Internalization which holds for many Justification Logic systems:

$$\textit{If } F \textit{ is valid, then one could construct a justification } p \textit{ such that } p{:}F \textit{ is valid.} \tag{10}$$

Internalization is usually represented in an equivalent form (in the presence of the Completeness Theorem) as a meta-rule (7). The algorithm which builds a justified belief $p\,{:}\,F$ from a strong evidence (proof) of the validity of F seems to be an instance of Goldman's 'causal chain.'

2.4 Lehrer and Paxson's Indefeasibility Condition

Lehrer and Paxson in [45] offered the following 'indefeasibility condition':

> *there is no further truth which, had the subject known it, would have defeated [subject's] present justification for the belief.*

The 'further truth' here could refer to a possible update of the subject's database, or some possible-worlds situation, etc.: these readings lie outside the scope of our language of Boolean Justification Logic. A natural reading of 'further truth' in our setting could be 'other postulate or assumption of the system,' which means a simple consistency property which vacuously holds for all Justification Logic systems considered here. Another plausible reading of 'further truth' could be 'further evidence,' and we assume this particular reading here. Since there is no temporal or update component in our language yet, 'any further evidence' could be understood for now as 'any other justification,' or just 'any justification.'

Furthermore, Lehrer and Paxson's condition seems to involve a negation of an existential quantifier over justifications 'there is no further truth ... ,' or

there is no justification....

However, within the classical logic tradition, we can read this as a universal quantifier over justifications followed by a negation

for any further evidence, it is not the case....

Denoting 'present justification for the belief' as the assertion $s{:}F$, we reformulate Lehrer–Paxson's condition as

given $s{:}F$, for any evidence t, it is not the case that t would have defeated $s{:}F$.

The next step is to formalize 't does not defeat $s{:}F$.' This informal statement seems to suggest an implication

> *if $s{:}F$ holds, then the joint evidence of s and t, which we denote here as $s+t$, is also an evidence for F, i.e., $(s+t){:}F$ holds.*

Here is the resulting formal version of Lehrer–Paxson's condition: for any proposition F and any justifications s and t, the following holds

$$s{:}F \rightarrow (s+t){:}F \qquad \textit{(Monotonicity Axiom).} \tag{11}$$

2.5 Further Assumptions

In order to build a formal account of justification, we will make some basic structural assumptions: *justifications are abstract objects which have structure, operations on justifications are potentially executable, agents do not lose or forget justifications, agents apply the laws of classical logic and accept their conclusions, etc.*

In the following, we consider both: **justifications**, which do not necessarily yield the truth of a belief, and **factive justifications**, which yield the truth of the belief.

3 Basic Principles and Systems

3.1 Application

The *Application* operation takes justifications s and t and produces a justification $s \cdot t$ such that if $s{:}(F \to G)$ and $t{:}F$, then $(s \cdot t){:}G$. Symbolically,

$$s{:}(F \to G) \to (t{:}F \to (s \cdot t){:}G). \tag{12}$$

This is a basic property of justifications assumed in combinatory logic and λ-calculi (cf. [64]), BHK-semantics ([65]), Kleene realizability ([39]), the Logic of Proofs LP ([4]), etc. Application Principle (12) is related to the epistemological closure principle (cf., for example, [20; 46]) that one knows everything that one knows to be implied by what one knows. However, (12) does not rely on this closure principle, since (12) deals with a broader spectrum of justifications, not necessarily linked to knowledge.

Note that the epistemological closure principle which could be formalized using the knowledge modality $\mathbf{K}$ as

$$\mathbf{K}(F \to G) \to (\mathbf{K}F \to \mathbf{K}G), \tag{13}$$

smuggles the *logical omniscience* defect into modal epistemic logic. The latter does not have the capacity to measure how hard it is to attain knowledge [21; 22; 38; 51; 56]. Justification Logic provides natural means of escaping logical omniscience by keeping track of the size of justification terms [10].

3.2 Monotonicity of Justification

The *Monotonicity* property of justification has been expressed by the operation *sum* '+,' which can be read from (11). If $s{:}F$, then whichever evidence t occurs, the combined evidence $s + t$ remains a justification for F. Operation '+' takes justifications s and t and produces $s + t$, which is a justification for everything justified by s or by t.

$$s{:}F \to (s+t){:}F \quad \text{and} \quad s{:}F \to (t+s){:}F.$$

A similar operation '+' is present in the Logic of Proofs LP, where the sum '$s+t$' can be interpreted as a concatenation of proofs s and t.

Correspondence Theorem 7 uses Monotonicity to connect Justification Logic with epistemic modal logic. However, it is an intriguing challenge to develop a theory of non-monotonic justifications which prompt belief revision. Some Justification Logic systems without Monotonicity have been studied in [13; 41; 42].

3.3 Basic Justification Logic J_0

Justification terms (polynomials) are built from justification variables $x, y, z, \ldots$ and justification constants $a, b, c, \ldots$ (with indices $i = 1, 2, 3 \ldots$ which we will be omitting whenever it is safe) by means of the operations **application** '$\cdot$' and **sum** '+'[1]. Constants denote atomic justifications which the system no longer analyzes; variables denote unspecified justifications.

Basic Logic of Justifications J_0:

A1. *Classical propositional axioms and rule Modus Ponens,*

A2. *Application Axiom* $s{:}(F \to G) \to (t{:}F \to (s \cdot t){:}G)$,

A3. *Monotonicity Axiom* $s{:}F \to (s+t){:}F,\ s{:}F \to (t+s){:}F$,

J_0 is the logic of general (not necessarily factive) justifications for an absolutely skeptical agent for whom no formula is provably justified, i.e., J_0 does not derive $t{:}F$ for any t and F. Such an agent is, however, capable of making *relative justification conclusions* of the form

$$\textit{if } x{:}A,\ y{:}B,\ \ldots,\ z{:}C \textit{ hold, then } t{:}F.$$

J_0 is able, with this capacity, to adequately emulate other Justification Logic systems in its language.

3.4 Logical Awareness and Constant Specifications

The *Logical Awareness principle* states that logical axioms are justified *ex officio*: an agent accepts logical axioms (including the ones concerning justifications) as justified. As stated here, Logical Awareness is too restrictive and Justification Logic offers a flexible mechanism of Constant Specifications to represent all shades of logical awareness.

Justification Logic distinguishes between an assumption and a justified assumption. Constants are used to denote justifications of assumptions in situations when we don't analyze these justifications any further. Suppose we want to postulate that an axiom A is justified for a given agent. The way to say it in Justification Logic is to postulate

$$e_1{:}A$$

for some evidence constant e_1 with index 1. Furthermore, if we want to postulate that this new principle $e_1{:}A$ is also justified, we can postulate

$$e_2{:}(e_1{:}A)$$

for the similar constant e_2 with index 2, etc. Keeping track of indices is not necessary, but it is easy and helps in decision procedures (cf. [44]). The set of all assumptions of this kind for a given logic is called a *Constant Specification*. Here is a formal definition.

A **Constant Specification** CS for a given logic $\mathcal{L}$ is a set of formulas

$$e_n : e_{n-1} : \ldots : e_1 : A \quad (n \geq 1),$$

where A is an axiom of $\mathcal{L}$, and $e_1, e_2, \ldots, e_n$ are similar constants with indices $1, 2, \ldots, n$. We also assume that CS contains all intermediate specifications, i.e., whenever $e_n : e_{n-1} : \ldots : e_1 : A$ is in CS, then $e_{n-1} : \ldots : e_1 : A$ is in CS too. In this paper, we will distinguish the following types of constant specifications:

[1]More elaborate models considered below in this paper also use additional operations on justifications, e.g., verifier '!' and negative verifier '?'.

- *empty*: $CS = \emptyset$. This corresponds to an absolutely skeptical agent (cf. a comment after axioms of J_0).

- *finite*: CS is a finite set of formulas. This is a representative case, since any specific derivation in Justification Logic concerns only finite sets of constants and constant specifications.

- *axiomatically appropriate*: for each axiom A there is a constant e_1 such that $e_1{:}A$ is in CS, and if
$$e_n{:}e_{n-1}{:}\ldots{:}e_1{:}A \in CS,$$
then
$$e_{n+1}{:}e_n{:}e_{n-1}{:}\ldots{:}e_1{:}A \in CS.$$
Axiomatically appropriate CS's are necessary for ensuring the Internalization property.

- *total*: for each axiom A and **any** constants $e_1, e_2, \ldots, e_n$,
$$e_n{:}e_{n-1}{:}\ldots{:}e_1{:}A \in CS.$$
We are reserving the name TCS for the total constant specification (for a given logic). Naturally, the total constant specification is axiomatically appropriate.

Logic of Justifications with given Constant Specification

$$\mathsf{J}_{CS} = \mathsf{J}_0 + CS.$$

Logic of Justifications

$$\mathsf{J} = \mathsf{J}_0 + \mathrm{R4},$$

where R4 is the **Axiom Internalization Rule**:

For each axiom A and any constants $e_1, e_2, \ldots, e_n$, infer $e_n{:}e_{n-1}{:}\ldots{:}e_1{:}A$.

Note that J_0 is $\mathsf{J}_\emptyset$, and J coincides with J_{TCS}. The latter reflects the idea of the unrestricted Logical Awareness for J. A similar principle appeared in the Logic of Proofs LP; it has also been anticipated in Goldman's [32]. Note that any **specific** derivation in J may be regarded as a derivation in J_{CS} for a corresponding **finite** constant specification CS, hence finite CS's constitute an important representative class of constant specifications.

Logical Awareness expressed by axiomatically appropriate constant specifications is an explicit incarnation of the Necessitation Rule in modal epistemic logic:

$$\vdash F \quad \Rightarrow \quad \vdash \mathbf{K}F \tag{14}$$

applied to axioms.

Let us consider some basic examples of derivations in J. In Examples 1 and 2, only constants of level 1 have been used; in such situations we skip indices completely.

Example 1 This example shows how to build a justification of a conjunction from justifications of the conjuncts. In the traditional modal language, this principle is formalized as

$$\Box A \wedge \Box B \rightarrow \Box (A \wedge B).$$

In J we express this idea in a more precise justification language.

1. $A \rightarrow (B \rightarrow (A \wedge B))$, *a propositional axiom*;
2. $c{:}[A \rightarrow (B \rightarrow (A \wedge B))]$, *from 1, by R4*;
3. $x{:}A \rightarrow (c \cdot x){:}(B \rightarrow (A \wedge B))$, *from 2, by A2 and Modus Ponens*;
4. $x{:}A \rightarrow (y{:}B \rightarrow ((c \cdot x) \cdot y){:}(A \wedge B))$, *from 3, by A2 and some propositional reasoning*;
5. $x{:}A \wedge y{:}B \rightarrow ((c \cdot x) \cdot y){:}(A \wedge B)$, *from 5, by propositional reasoning.*

The derived formula 5 contains constant c, which was introduced in line 2, and the complete reading of the result of this derivation is

$$x{:}A \wedge y{:}B \rightarrow ((c \cdot x) \cdot y){:}(A \wedge B),\ \textit{given}\ c{:}[A \rightarrow (B \rightarrow (A \wedge B))].$$

Example 2 This example shows how to build a justification of a disjunction from justifications of either of the disjuncts. In the usual modal language this is represented by

$$\Box A \vee \Box B \rightarrow \Box (A \vee B).$$

Let us see how this would look in J.

1. $A \rightarrow (A \vee B)$, *by A1*;
2. $a{:}[A \rightarrow (A \vee B)]$, *from 1, by R4*;
3. $x{:}A \rightarrow (a \cdot x){:}(A \vee B)$, *from 2, by A2 and Modus Ponens*;
4. $B \rightarrow (A \vee B)$, *by A1*;
5. $b{:}[B \rightarrow (A \vee B)]$, *from 4, by R4*;
6. $y{:}B \rightarrow (b \cdot y){:}(A \vee B)$ *from 5, by A2 and Modus Ponens*;
7. $(a \cdot x){:}(A \vee B) \rightarrow (a \cdot x + b \cdot y){:}(A \vee B)$, *by A3*;
8. $(b \cdot y){:}(A \vee B) \rightarrow (a \cdot x + b \cdot y){:}(A \vee B)$, *by A3*;
9. $(x{:}A \vee y{:}B) \rightarrow (a \cdot x + b \cdot y){:}(A \vee B)$ *from 3, 6, 7, 8, by propositional reasoning.*

The complete reading of the result of this derivation is

$$(x{:}A \vee y{:}B) \rightarrow (a \cdot x + b \cdot y){:}(A \vee B),\ \textit{given}\ a{:}[A \rightarrow (A \vee B)]\ \textit{and}\ \ b{:}[B \rightarrow (A \vee B)].$$

Explicit mention of Constant Specifications of Justification Logic systems is normally used when semantic issues are concerned: e.g., arithmetical, symbolic, and epistemic semantics. To define the truth value of a formula under a given interpretation, one should be given a specification of constants involved.

For each constant specification CS, J_{CS} enjoys the Deduction Theorem, because J_0 contains propositional axioms and *Modus Ponens* as the only rule of inference.

Theorem 1 *For each axiomatically appropriate constant specification CS, J_{CS} enjoys Internalization:*

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

Proof. Induction on derivation length. Suppose $\vdash F$. If F is an axiom, then, since CS is axiomatically appropriate, there is a constant e suchb that $e{:}F$ is in CS, hence an axiom of J_{CS}. If F is in CS, then, since CS is axiomatically appropriate, $e{:}F$ is in CS for some constant e. If F is obtained by *Modus Ponens* from $X \to F$ and X, then, by the Induction Hypothesis, $\vdash s{:}(X \to F)$ and $\vdash t{:}X$ for some s, t. By the Application Axiom, $\vdash (s{\cdot}t){:}F$. Note that Internalization can require a growth of constant specification sets; if $\vdash F$ with a Constant Specification CS, then the proof of $p{:}F$ may need some Constant Specification CS' which is different from CS. $\square$

4 Red Barn Example and Tracking Justifications

We begin illustrating new capabilities of Justification Logic with a paradigmatic Red Barn Example which Kripke developed in 1980 in objection to Nozick's account of knowledge (cf. article *The Epistemic Closure Principle* in Stanford Encyclopedia of Philosophy [46], from which we borrow the formulation, with some editing for brevity).

> Suppose I am driving through a neighborhood in which, unbeknownst to me, papier-mâché barns are scattered, and I see that the object in front of me is a barn. Because I have barn-before-me percepts, I believe that the object in front of me is a barn. Our intuitions suggest that I fail to know barn. But now suppose that the neighborhood has no fake red barns, and I also notice that the object in front of me is red, so I know a red barn is there. This juxtaposition, being a red barn, which I know, entails there being a barn, which I do not, "is an embarrassment"[2].

We proceed in the spirit of the Red Barn Example and consider it a general test for theories that explain knowledge. What we want is a way to represent what is going on here which maintains epistemic closure,

one knows everything that one knows to be implied by what one knows, (15)

but also preserves the problems the example was intended to illustrate.

We present plausible formal analysis of the Red Barn Example in epistemic modal logic (subsections 4.1 and 4.2) and in Justification Logic (subsections 4.3 and 4.4). We will see that epistemic modal logic is capable only of telling us that there is a problem, whereas Justification Logic helps to analyse what has gone wrong. We see that closure holds as it is supposed to, and we see that if we keep track of justifications we can analyse why we had a problem.

4.1 Red Barn in modal logic of belief

In our first formalization, the logical derivation will be made in epistemic modal logic with 'my belief' modality $\Box$. We then interpret some of the occurrences of $\Box$ as 'knowledge' according to the problem's description. We will not try to capture the whole scenario formally; to make our point, it suffices to formalize and verify its "entailment" part. Let

[2] Dretske [20].

- B be 'the object in front of me is a barn,'
- R be 'the object in front of me is red,'
- $\Box$ be 'my belief' modality.

The formulation considers observations 'I see a barn' and 'I see a red barn,' and claims logical dependencies between them. The following is a natural formalization of these assumptions in the epistemic modal logic of belief:

1. $\Box B$, 'I believe that the object in front of me is a barn';
2. $\Box(B \wedge R)$, 'I believe that the object in front of me is a red barn.'

At the metalevel, we assume that 2 is knowledge, whereas 1 is not knowledge by the problem's description. So, we could add factivity of 2, $\Box(B \wedge R) \rightarrow (B \wedge R)$, to the formal description, but this would not matter for our conclusions. We note that indeed 1 logically follows from 2 in the modal logic of belief K:

3. $(B \wedge R) \rightarrow B$, *logical axiom*;
4. $\Box[(B \wedge R) \rightarrow B]$*, from 3, by Necessitation. As a logical truth, this is a case of knowledge too*;
5. $\Box(B \wedge R) \rightarrow \Box B$*, from 4, by modal logic.*

Within this formalization, it appears that Closure Principle (15) is violated: $\Box(B \wedge R)$ is knowledge by the problem's description, $\Box[(B \wedge R) \rightarrow B]$ is knowledge as a simple logical axiom, whereas $\Box B$ is not knowledge.

4.2 Red Barn in modal logic of knowledge

Now we will use epistemic modal logic with 'my knowledge' modality $\mathbf{K}$. Here is a straightforward formalization of Red Barn Example assumptions:

1. $\neg \mathbf{K} B$, 'I do not know that the object in front of me is a barn';
2. $\mathbf{K}(B \wedge R)$, 'I know that the object in front of me is a red barn.'

It is easy to see that these assumptions are inconsistent in the modal logic of knowledge. Indeed,

3. $\mathbf{K}(B \wedge R) \rightarrow (\mathbf{K} B \wedge \mathbf{K} R)$*, by normal modal logic*;
4. $\mathbf{K} B \wedge \mathbf{K} R$*, from 2 and 3, by Modus Ponens*;
5. $\mathbf{K} B$*, from 4, by propositional logic.*

Lines 1 and 5 formally contradict each other.

Modal logic of knowledge does not seem to apply here.

4.3 Red Barn in Justification Logic of belief

Justification Logic seems to provide a more fine-grained analysis of the Red Barn Example. We naturally refine assumptions by introducing individual justifications u for belief that B, and v for belief that $B \wedge R$. The set of assumptions in the Justification Logic is

1. $u{:}B$, 'u is the reason to believe that the object in front of me is a barn';

2. $v{:}(B \wedge R)$, 'v is the reason to believe that the object in front of me is a red barn.' *On the metalevel, the description states that this is a case of knowledge, not merely a belief.*

Again, we can add the factivity condition for 2, $v{:}(B \wedge R) \to (B \wedge R)$, but this does not change the analysis here. Let us try to reconstruct the reasoning of the agent in J:

3. $(B \wedge R) \to B$, *logical axiom*;
4. $a{:}[(B \wedge R) \to B]$, *from 3, by Axiom Internalization. This is also knowledge, as before*;
5. $v{:}(B \wedge R) \to (a \cdot v){:}B$, *from 4, by Application and Modus Ponens*;
6. $(a \cdot v){:}B$, *from 2 and 5, by Modus Ponens.*

Closure holds! Instead of deriving 1 from 2 as in Section 4.1, we have obtained a correct conclusion that $(a \cdot v){:}B$, i.e., 'I know B for reason $a \cdot v$,' which seems to be **different** from u: the latter is the result of a perceptual observation, whereas the former is the result of logical reasoning. In particular, we cannot conclude that 2, $v{:}(B \wedge R)$, entails 1, $u{:}B$; moreover, with some basic model theory of J in Section 5, we can show that 2 **does not entail** 1. Hence, after observing a red façade, I indeed know B, but this knowledge does not come from 1, which remains a case of belief rather than of knowledge.

4.4 Red Barn in Justification Logic of knowledge

Within this formalization, $t{:}F$ is interpreted as

'I know F for reason t.'

As in Section 4.2, we assume

1. $\neg u{:}B$, 'u is not a sufficient reason to know that the object is a barn';
2. $v{:}(B \wedge R)$, 'v is a sufficient reason to know that the object is a red barn.'

This is a perfectly consistent set of assumptions in the logic of factive justifications

J + *Factivity Principle* $(t{:}F \to F)$.

As in 4.3, we can derive $(a \cdot v){:}B$ where $a{:}[(B \wedge R) \to B]$, but this does not lead to a contradiction. Claims $\neg u{:}B$ and $(a \cdot v){:}B$ naturally co-exist. They refer to different justifications u and $a \cdot v$ of the same fact B; one of them insufficient and the other quite sufficient for my knowledge that B.

It appears that in 4.3 and 4.4, Justification Logic represents the structure of the argument made by Kripke in his Red Barn Example, and which was not captured by traditional epistemic modal tools. The Justification Logic formalization represents what seems to be happening in such a case; we can maintain closure of knowledge under logical entailment, even though 'barn' is not perceptually known.

The formal analyses provided in 4.3 and 4.4 is similar in spirit to a "conclusive reason" style of analysis ([19]) which is formulated in terms of evidence tracking rather than belief tracking.

5 Basic Epistemic Semantics

The standard epistemic semantics for J has been provided by the proper adaptation of Kripke-Fitting models [25] and Mkrtychev models [50].

A Kripke-Fitting J**-model** $\mathcal{M} = (W, R, \mathcal{E}, \Vdash)$ is a Kripke model $(W, R, \Vdash)$ enriched with an **admissible evidence function** $\mathcal{E}$ such that $\mathcal{E}(t, F) \subseteq W$ for any justification t and formula F. Informally, $\mathcal{E}(t, F)$ specifies the set of possible worlds where t is considered admissible evidence for F. The intended use of $\mathcal{E}$ is in the truth definition for justification assertions:

$u \Vdash t{:}F$ if and only if

1. *F holds for all possible situations, i.e., $v \Vdash F$ for all v such that uRv;*

2. *t is an admissible evidence for F at u, i.e., $u \in \mathcal{E}(t, F)$.*

An admissible evidence function $\mathcal{E}$ must satisfy the closure conditions with respect to operations '$\cdot$' and '$+$':

- *Application*: $\mathcal{E}(s, F \to G) \cap \mathcal{E}(t, F) \subseteq \mathcal{E}(s \cdot t, G)$. This condition states that whenever s is an admissible evidence for $F \to G$ and t is an admissible evidence for F, their 'product,' $s \cdot t$, is an admissible evidence for G.

- *Sum*: $\mathcal{E}(s, F) \cup \mathcal{E}(t, F) \subseteq \mathcal{E}(s+t, F)$. This condition guarantees that $s+t$ is an admissible evidence for F whenever either s is admissible for F or t is admissible for F.

These are natural conditions to place on $\mathcal{E}$ because they are necessary for making basic axioms of Application and Monotonicity valid.

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

Given a model $\mathcal{M} = (W, R, \mathcal{E}, \Vdash)$, the forcing relation $\Vdash$ is extended from sentence variables to all formulas as follows: 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 $u \nVdash F$, etc.);

2. $u \Vdash t{:}F$ iff $u \in \mathcal{E}(t, F)$ and $v \Vdash F$ for every $v \in W$ with uRv.

Note that an admissible evidence function $\mathcal{E}$ may be regarded as a Fagin-Halpern awareness function [23] equipped with the structure of justifications.

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

Theorem 2 *For any Constant Specification CS, J_{CS} is sound and complete for the class of all Kripke-Fitting models respecting CS.*

Proof.

1. Fix a Constant Specification CS and consider J_{CS}.

Soundness is straightforward. Induction on derivations in J_{CS}. Let us check the axioms.

Application. Suppose $u \Vdash s{:}(F \to G)$ and $u \Vdash t{:}F$. Then, by the definition of forcing, $u \in \mathcal{E}(s, F \to G)$ and $u \in \mathcal{E}(t, F)$, hence, by the closure condition for $\mathcal{E}$, $u \in \mathcal{E}(s{\cdot}t, G)$. Moreover, for each v such that uRv, $v \Vdash F \to G$ and $v \Vdash F$, hence $v \Vdash G$. Thus $u \Vdash (s{\cdot}t){:}G$ and $u \Vdash s{:}(F \to G) \to (t{:}F \to (s{\cdot}t){:}G)$.

Sum. Suppose $u \Vdash t{:}F$. Then $u \in \mathcal{E}(t, F)$, hence, by the closure condition for $\mathcal{E}$, $u \in \mathcal{E}(s{+}t, F)$. In addition, $v \Vdash F$ for each v such that uRv, hence $u \Vdash (s{+}t){:}F$. Thus $u \Vdash t{:}F \to (s{+}t){:}F$.

Axioms from CS hold at each world, since the models respect CS. The Induction Step corresponds to the use of *Modus Ponens*, which is clearly a sound rule here.

To establish completeness, we use standard canonical model construction. The canonical model $\mathcal{M} = (W, R, \mathcal{E}, \Vdash)$ for J_{CS} is defined as follows:

- W is the set of all maximal consistent sets in J_{CS}. Following an established tradition, we denote elements of W as Γ, Δ, etc.;
- $\Gamma R \Delta$ iff $\Gamma^\sharp \subseteq \Delta$, where $\Gamma^\sharp = \{F \mid t{:}F \in \Gamma \text{ for some } t\}$;
- $\mathcal{E}(s, F) = \{\Gamma \in W \mid s{:}F \in \Gamma\}$;
- $\Gamma \Vdash p$ iff $p \in \Gamma$.

The Truth Lemma claims that for all F's,

$$\Gamma \Vdash F \quad \textit{if and only if} \quad F \in \Gamma.$$

This is established by standard induction on the complexity of F. The atomic cases are covered by the definition of '$\Vdash$.' The Boolean induction steps are standard. Consider the case when F is $t{:}G$ for some t and G.

If $t{:}G \in \Gamma$, then $G \in \Delta$ for all Δ such that $\Gamma R \Delta$ by the definition of R. By the Induction Hypothesis, $\Delta \Vdash G$. In addition, $\Gamma \in \mathcal{E}(t, G)$ by the definition of $\mathcal{E}$. Hence $\Gamma \Vdash t{:}G$, i.e., $\Gamma \Vdash F$.

If $t{:}G \notin \Gamma$, then $\Gamma \notin \mathcal{E}(t, G)$, i.e., $\Gamma \not\Vdash t{:}G$ and $\Gamma \not\Vdash F$.

Furthermore, $\mathcal{M}$ respects CS at each node. Indeed, by the construction of $\mathcal{M}$, $CS \subseteq \Gamma$ for each $\Gamma \in W$. By the Truth Lemma, $\Gamma \Vdash c{:}A$ for each $c{:}A \in CS$.

The conclusion of the proof of Theorem 2 is standard. Let F be not derivable in J_{CS}. Then the set $\{\neg F\}$ is consistent. Using the standard saturation construction ([23; 48]), extend $\{\neg F\}$ to a maximal consistent set Γ. By consistency, $F \notin \Gamma$. By the Truth Lemma, $\Gamma \not\Vdash F$. □

There are several features of the canonical model which could be included into the formulation of the Completeness Theorem to make it stronger.

Strong Evidence. We can show that the canonical model considered in this proof satisfies the Strong Evidence property

$$\Gamma \in \mathcal{E}(t, F) \quad \textit{implies} \quad \Gamma \Vdash t{:}F.$$

Indeed, let $\Gamma \in \mathcal{E}(t, F)$. By the definition of $\mathcal{E}$, $t{:}F \in \Gamma$, hence $F \in \Gamma^\sharp$ and $F \in \Delta$ for each Δ such that $\Gamma R \Delta$. By the Truth Lemma, $\Delta \Vdash F$, hence $\Gamma \Vdash t{:}F$. In a model with the Strong Evidence property there are no void or irrelevant justifications; if t is an admissible evidence for F, then t is a 'real evidence' for F, i.e., F holds at all possible worlds.

Fully Explanatory property for axiomatically appropriate Constant Specifications:

If $\Delta \Vdash F$ for all Δ such that $\Gamma R \Delta$, then $\Gamma \Vdash t{:}F$ for some t.

Note that for axiomatically appropriate constant specifications CS, the Internalization property holds: if G is provable in J_{CS}, then $t{:}G$ is also provable there for some term t. Here is the proof of the Fully Explanatory property for canonical models[3]. Suppose $\Gamma \nVdash t{:}F$ for any justification term t. Then the set $\Gamma^\sharp \cup \{\neg F\}$ is consistent. Indeed, otherwise for some $t_1{:}X_1, t_2{:}X_2, \ldots, t_n{:}X_n \in \Gamma$, $X_1 \rightarrow (X_2 \rightarrow \ldots \rightarrow (X_n \rightarrow F)\ldots)$ is provable. By Internalization, there is a justification s such that $s{:}(X_1 \rightarrow (X_2 \rightarrow \ldots \rightarrow (X_n \rightarrow F)\ldots))$ is also provable. By Application, $t_1{:}X_1 \rightarrow (t_2{:}X_2 \rightarrow \ldots \rightarrow (t_n{:}X_n \rightarrow (s \cdot t_1 \cdot t_2 \cdot \ldots \cdot t_n){:}F)\ldots)$ is provable, hence $\Gamma \vdash t{:}F$ for $t = s \cdot t_1 \cdot t_2 \cdot \ldots \cdot t_n$. Therefore, $\Gamma \Vdash t{:}F$ — a contradiction. Let Δ be a maximal consistent set extending $\Gamma^\sharp \cup \{\neg F\}$. By the definition of R, $\Gamma R \Delta$, by the Truth Lemma, $\Delta \nVdash F$, which contradicts the assumptions.

Mkrtychev semantics is a predecessor of Kripke-Fitting semantics ([50]). *Mrktychev models* are Kripke-Fitting models with a single world, and the proof of Theorem 2 can be easily modified to establish completeness of J_{CS} with respect to Mkrtychev models.

Theorem 3 *For any Constant Specification CS, J_{CS} is sound and complete for the class of Mrktychev models respecting CS.*

Proof. Soundness follows immediately from Theorem 2. For completeness, define the canonical model as in Theorem 2 except for R, which should be taken empty. This assumption makes the condition '$\Delta \Vdash F$ for all Δ such that $\Gamma R \Delta$' vacuously true, and the forcing condition for justification assertions $\Gamma \Vdash t{:}F$ becomes equivalent to $\Gamma \in \mathcal{E}(t, F)$, i.e., $t{:}F \in \Gamma$. This simplification immediately verifies the Truth Lemma.

The conclusion of the proof of Theorem 3 is standard. Let F be not derivable in J_{CS}. Then the set $\{\neg F\}$ is consistent. Using the standard saturation construction, extend it to a maximal consistent set Γ containing $\neg F$. By consistency, $F \notin \Gamma$. By the Truth Lemma, $\Gamma \nVdash F$. The Mkrtychev model consisting of this particular Γ is the desired counter-model for F. The rest of the canonical model is irrelevant. $\square$

Note that Mkrtychev models built in Theorem 3 are not reflexive, and possess the Strong Evidence property. On the other hand, Mkrtychev models cannot be Fully Explanatory, since '$\Delta \Vdash F$ for all Δ such that $\Gamma R \Delta$' is vacuously true, but $\Gamma \Vdash t{:}F$ is not.

Theorem 3 shows that the information about Kripke structure in Kripke-Fitting models can be completely encoded by the admissible evidence function. Mkrtychev models play an important theoretical role in Justification Logic [7; 16; 40; 43; 49]. On the other hand, as we will see in Section 10, Kripke-Fitting models can be useful as counter-models with desired properties since they

[3] This proof for LP was offered by Fitting in [25].

take into account both epistemic Kripke structure and evidence structure. Speaking metaphorically, Kripke-Fitting models naturally reflect two reasons why a certain fact F can be unknown to an agent: F fails at some possible world or an agent does not have a sufficient evidence of F.

Another application area of Kripke-Fitting style models is Justification Logic with both epistemic modalities and justification assertions (cf. [5; 12]).

Corollary 1 [Model existence] *For any constant specification CS,* J_{CS} *is consistent and has a model.*

Proof. J_{CS} is consistent. Indeed, suppose J_{CS} proves $\perp$, and erase all justification terms (with ':'s) in each of its formulas. What remains is a chain of propositional formulas provable in classical logic (an easy induction on the length of the original proof) ending with $\perp$ – contradiction.

To build a model for J_{CS}, use the Completeness Theorem (Theorem 2). Since J_{CS} does not prove $\perp$, by Completeness, there is a J_{CS}-model (where $\perp$ is false, of course). $\square$

6 Factivity

Unlike Application and Monotonicity, *Factivity* of justifications is not required in basic Justification Logic systems, which makes the latter capable of representing both partial and factive justifications.

Factivity states that justifications of F are factive, i.e., sufficient for an agent to conclude that F is true. This yields the Factivity Axiom

$$t{:}F \rightarrow F, \tag{16}$$

which has a similar motivation to the Truth Axiom in epistemic modal logic

$$\mathbf{K}F \rightarrow F, \tag{17}$$

widely accepted as a basic property of knowledge (Plato, Wittgenstein, Hintikka, etc.).

The Factivity Axiom (16) first appeared in the Logic of Proofs LP as a principal feature of mathematical proofs. Indeed, in this setting (16) is valid: if there is a mathematical proof t of F, then F must be true.

We adopt the Factivity Axiom (16) for justifications that lead to knowledge. However, factivity alone does not warrant knowledge, which has been demonstrated by Gettier examples ([29]).

Logic of Factive Justifications:

$$\mathsf{JT}_0 = \mathsf{J}_0 + \mathrm{A4},$$

$$\mathsf{JT} = \mathsf{J} + \mathrm{A4},$$

with

A4. *Factivity Axiom* $t{:}F \rightarrow F$.

Systems JT_{CS} corresponding to Constant Specifications CS are defined as in Section 3.4.

JT-models are J-models with reflexive accessibility relations R. The reflexivity condition makes each possible world accessible from itself which exactly corresponds to the Factivity Axiom. The direct analogue of Theorem 1 hold for JT_{CS} as well.

Theorem 4 *For any Constant Specification CS, each of the logics JT_{CS} is sound and complete with respect to the class of JT-models respecting CS.*

Proof. We now proceed as in the proof of Theorem 2. The only addition to soundness is establishing that the Factivity Axiom holds in reflexive models. Let R be reflexive. Suppose $u \Vdash t{:}F$. Then $v \Vdash F$ for all v such that uRv. By reflexivity of R, uRu, hence $u \Vdash F$ as well.

For completeness, it suffices to check that R in the canonical model is reflexive. Indeed, if $s{:}F \in \Gamma$, then, by the properties of the maximal consistent sets, $F \in \Gamma$ as well, since JT derives $s{:}F \rightarrow F$ (with any CS). Hence $\Gamma^\sharp \subseteq \Gamma$ and $\Gamma R \Gamma$. ◻

Mkrtychev JT-models are singleton JT-models, i.e., JT-models with singleton W's.

Theorem 5 *For any Constant Specification CS, each of the logics JT_{CS} is sound and complete with respect to the class of Mkrtychev JT-models respecting CS.*

Proof. Soundness follows from Theorem 4. For completeness, we follow the footprints of Theorem 2, Theorem 3, but define the accessibility relation R as

$$\Gamma R \Delta \quad \textit{iff} \quad \Gamma = \Delta.$$

◻

6.1 Russell's Example: Induced Factivity

Here is Russell's well-known example from [60] of an epistemic scenario which can be meaningfully analyzed in Justification Logic.

> 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[4]. But if he believes that Mr. Balfour was the late Prime Minister, 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.

As in the Red Barn Example (Section 4), we have to handle a wrong reason for a true justified fact. Again, the tools at Justification Logic seem to be useful and adequate here.

Let B stand for

the late Prime Minister's last name began with a 'B.'

[4]Which was common knowledge back in 1912.

Furthermore, let w be a wrong reason for B and r the right (hence factive) reason for B. Then, Russell's example yields the following assumptions:

$$\{w{:}B,\ r{:}B,\ r{:}B \to B\}. \tag{18}$$

In the original setting (18), we do not claim that w is a factive justification for B; moreover, such factivity is not completely consistent with our intuition. Paradoxically, however, in the basic Justification Logic J, we can logically deduce factivity of w from (18):

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

However, this derivation utilizes the fact that r is a factive justification for B to conclude $w{:}B \to B$, which constitutes the case of 'induced factivity' of $w{:}B$. The question is, how can we distinguish the 'real' factivity of $r{:}B$ from an 'induced factivity' of $w{:}B$? Again, some sort of truth-tracking is needed here, and Justification Logic seems to do the job. The natural approach would be to consider the set of assumptions (18) **without** $r{:}B$, i.e.,

$$\{w{:}B,\ r{:}B \to B\}, \tag{19}$$

and establish that factivity of w, i.e., $w : B \to B$ is not derivable from (19). Here is a J-model $\mathcal{M} = (W, R, \mathcal{E}, \Vdash)$ in which (19) holds but $w{:}B \to B$ does not.

$W = \{\mathbf{0}\}$, $R = \emptyset$, $\mathbf{0} \nVdash B$, and $\mathcal{E}(t, F)$ holds for all pairs (t, F) except (r, B). It is easy to see that the closure conditions *Application* and *Sum* on $\mathcal{E}$ are fulfilled. At $\mathbf{0}$, $w{:}B$ holds, i.e.,

$$\mathbf{0} \Vdash w{:}B,$$

since w is an admissible evidence for B at $\mathbf{0}$ and there are no possible worlds accessible from $\mathbf{0}$. Furthermore,

$$\mathbf{0} \nVdash r{:}B,$$

since, according to $\mathcal{E}$, r is not an admissible evidence for B at $\mathbf{0}$. Hence

$$\mathbf{0} \Vdash r{:}B \to B.$$

On the other hand,

$$\mathbf{0} \nVdash w{:}B \to B$$

since B does not hold at $\mathbf{0}$.

7 Additional Principles and Systems

In this section, we discuss other principles and operations which may or may not be added to the core Justification Logic systems.

7.1 Positive Introspection

One of the common principles of knowledge is identifying *knowing* and *knowing that one knows.* In the formal modal setting, this corresponds to

$$\mathbf{K}F \to \mathbf{KK}F.$$

This principle has an adequate explicit counterpart: the fact that the agent accepts t as a sufficient evidence of F serves as a sufficient evidence that $t{:}F$. Often, such meta-evidence has a physical form, e.g., a referee report certifying that a proof of a paper is correct, a computer verification output given a formal proof t of F as an input, a formal proof that t is a proof of F, etc. *Positive Introspection* assumes that given t, the agent produces a justification $!t$ of $t{:}F$ such that

$$t{:}F \to\, !t{:}(t{:}F).$$

Positive Introspection in this operational form first appeared in the Logic of Proofs LP [2; 4].

We define

$$\mathsf{J4} = \mathsf{J} + \mathrm{A5}$$

and

$$\mathsf{LP} \ = \mathsf{JT} + \mathrm{A5},[5]$$

with

A5. *Positive Introspection Axiom* $t{:}F \to\, !t{:}(t{:}F)$.

We also define $\mathsf{J4}_0$, $\mathsf{J4}_{CS}$, LP_0, and LP_{CS} in the natural way (cf. Section 3.4). The direct analogue of Theorem 1 holds for $\mathsf{J4}_{CS}$ and LP_{CS} as well.

Note that in the presence of the Positive Introspection Axiom, one could limit the scope of the Axiom Internalization Rule R4 to internalizing axioms which are not yet of the form $e{:}A$. This is how it has been done in LP: the Axiom Internalization can then be emulated by using $!!e{:}(!e{:}(e{:}A))$ instead of $e_3{:}(e_2{:}(e_1{:}A))$, etc. The notion of Constant Specification could also be simplified accordingly.

Such modifications are minor and they do not affect the main theorems and applications of Justification Logic.

7.2 Negative Introspection

Pacuit and Rubtsova considered in [54; 55; 57; 58] the *Negative Introspection* operation '?' which verifies that a given justification assertion is false. A possible motivation for considering such an operation could be that the positive introspection operation '!' may well be regarded as capable of providing conclusive verification judgments about the validity of justification assertions $t{:}F$. So, when t is not a justification for F, such a '!' should conclude that $\neg t{:}F$. This is normally the case

[5] In our notation, LP can be assigned the name $\mathsf{JT4}$. However, in virtue of a fundamental role played by LP for Justification Logic, we suggest keeping the name LP for this system.

for computer proof verifiers, proof checkers in formal theories, etc. This motivation is, however, nuanced: the examples of proof verifiers and proof checkers work with both t and F as inputs, whereas the Pacuit-Rubtsova format $?t$ suggests that the only input for '?' is a justification t, and the result $?t$ is supposed to justify propositions $\neg t{:}F$ uniformly for all F's for which $t{:}F$ does not hold. Such an operation '?' does not exist for formal mathematical proofs since $?t$ should be a single proof of infinitely many propositions $\neg t{:}F$, which is impossible[6]. For what it's worth, we include Negative Introspection in the list of additional justification principles, and leave the decision of whether to accept it or not to the user.

A6. *Negative Introspection Axiom* $\neg t{:}F \rightarrow ?t{:}(\neg t{:}F)$.

We define systems

$$\mathsf{J45}{=}\ \mathsf{J4} + \mathrm{A6},$$

$$\mathsf{JD45}{=}\ \mathsf{J45} + \neg t{:}\bot,$$

$$\mathsf{JT45} = \mathsf{J45} + \mathrm{A4},$$

and naturally extend these definitions to $\mathsf{J45}_{CS}$, $\mathsf{JD45}_{CS}$, and $\mathsf{JT45}_{CS}$.

The direct analogue of Theorem 1 holds for $\mathsf{J45}_{CS}$, $\mathsf{JD45}_{CS}$, and $\mathsf{JT45}_{CS}$.

7.3 More Epistemic Models

We now define epistemic models for other Justification Logic systems.

- $\mathsf{J4}$-models are J-models with *transitive* R and two additional conditions:

  Monotonicity with respect to R, i.e., $u \in \mathcal{E}(t, F)$ and uRv yield $v \in \mathcal{E}(t, F)$,

  Introspection closure: $\mathcal{E}(t, F) \subseteq \mathcal{E}(!t, t{:}F)$;

- LP-models are $\mathsf{J4}$-models with *reflexive* R (these are the original Kripke-Fitting models);

- $\mathsf{J45}$-models are $\mathsf{J4}$-models satisfying conditions:

  Negative Introspection closure: $[\mathcal{E}(t, F)]^c \subseteq \mathcal{E}(?t, \neg t{:}F)$ (Here $[X]^c$ denotes the complement of X.)

  Strong Evidence: $u \Vdash t{:}F$ for all $u \in \mathcal{E}(t, F)$ (i.e., only 'actual' evidence is admissible).

  Note that $\mathsf{J45}$-models satisfy the *Stability* property: uRv yields '$u \in \mathcal{E}(t, F)$ iff $v \in \mathcal{E}(t, F)$.' In other words, $\mathcal{E}$ is monotone with respect to R^{-1} as well. Indeed, the direction '$u \in \mathcal{E}(t, F)$ yields $v \in \mathcal{E}(t, F)$' is due to Monotonicity. Suppose $u \notin \mathcal{E}(t, F)$. By Negative Introspection closure, $u \in \mathcal{E}(?t, \neg t{:}F)$. By Strong Evidence, $u \Vdash ?t{:}(\neg t{:}F)$. By the definition of forcing, $v \Vdash \neg t{:}F$, i.e., $v \not\Vdash t{:}F$. By Strong Evidence, $v \notin \mathcal{E}(t, F)$.

[6] A proof-compliant way to represent negative introspection in Justification Logic was suggested in [9], but we will not consider it here.

Note also that the Euclidean property of the accessibility relation R is not required for $\mathsf{J45}$-models and is not needed to establish the soundness of $\mathsf{J45}$ with respect to $\mathsf{J45}$-models. However, the canonical model for $\mathsf{J45}$ is Euclidean, hence both soundness and completeness claims trivially survive an additional requirement that R is Euclidean.

- $\mathsf{JD45}$-models are $\mathsf{J45}$-models with the *Serial* condition on the accessibility relation R: for each u there is v such that uRv holds.

- $\mathsf{JT45}$-models are $\mathsf{J45}$-models with reflexive R. Again, the Euclidean property (or, equivalently, symmetry) of R is not needed for soundness. However, these properties hold for the canonical $\mathsf{JT45}$-model, hence they could be included into the formulation of the Completeness Theorem.

Theorem 6 *Each of the logics $\mathsf{J4}_{CS}$, LP_{CS}, $\mathsf{J45}_{CS}$, $\mathsf{JT45}_{CS}$ for any Constant Specification is sound and complete with respect to the corresponding class of epistemic models. $\mathsf{JD45}_{CS}$ is complete w.r.t. its epistemic models for axiomatically appropriate CS.*

Proof. We will follow the footprints of the proof of Theorem 2.

1. $\mathsf{J4}$. For soundness, it now suffices to check the validity of the Positive Introspection Axiom at each node of any $\mathsf{J4}$-model. Suppose $u \Vdash t{:}F$. Then $u \in \mathcal{E}(t, F)$ and $v \Vdash F$ for each v such that uRv. By the closure condition, $u \in \mathcal{E}(!t, t{:}F)$, and it remains to check that $v \Vdash t{:}F$. By monotonicity of $\mathcal{E}$, $v \in \mathcal{E}(t, F)$. Now, take any w such that vRw. By transitivity of R, uRw as well, hence $w \Vdash F$. Thus $v \Vdash t{:}F$, $u \Vdash !t{:}t{:}F$, and $u \Vdash t{:}F \rightarrow !t{:}t{:}F$.

Completeness is again established as in Theorem 2. It only remains to check that the accessibility relation R is transitive, the admissible evidence function $\mathcal{E}$ is monotone, and the additional closure condition on $\mathcal{E}$ holds.

Monotonicity. Suppose $\Gamma R \Delta$ and $\Gamma \in \mathcal{E}(t, F)$, i.e., $t{:}F \in \Gamma$. By maximality of Γ, $!t{:}t{:}F \in \Gamma$ as well, since $\mathsf{J4} \vdash t{:}F \rightarrow !t{:}t{:}F$. By definition, $t{:}F \in \Delta$, i.e., $\Delta \in \mathcal{E}(t, F)$.

Transitivity. Suppose $\Gamma R \Delta$, $\Delta R \Sigma$, and $t{:}F \in \Gamma$. Then, by monotonicity, $t{:}F \in \Delta$. By the definition of R, $F \in \Sigma$, hence $\Gamma R \Sigma$.

Closure. Suppose $\Gamma \in \mathcal{E}(t, F)$, i.e., $t{:}F \in \Gamma$. Then as above, $!t{:}t{:}F \in \Gamma$, hence $\Gamma \in \mathcal{E}(!t, t{:}F)$.

2. LP. This is the well-studied case of the Logic of Proofs, cf. [25].

3. $\mathsf{J45}$. Soundness. We have to check the Negative Introspection Axiom. Let $u \Vdash \neg t{:}F$, i.e., $u \not\Vdash t{:}F$. By the Strong Evidence condition, $u \notin \mathcal{E}(t, F)$. By Negative Introspection closure, $u \in \mathcal{E}(?t, \neg t{:}F)$. By Strong Evidence, $u \Vdash ?t{:}(\neg t{:}F)$.

Completeness. We follow the same canonical model construction as in J and $\mathsf{J4}$. The only addition is checking **Negative Introspection closure**. Let $\Gamma \notin \mathcal{E}(t, F)$. Then $t{:}F \notin \Gamma$. By maximality, $\neg t{:}F \in \Gamma$. By the Negative Introspection Axiom, $?t{:}(\neg t{:}F) \in \Gamma$, hence $\Gamma \in \mathcal{E}(?t, \neg t{:}F)$.

Here is an additional feature of the canonical model that can be included in the formulation of the Completeness Theorem to make it more specific.

R is Euclidean. Let $\Gamma R \Delta$ and $\Gamma R \Delta'$. It suffices to show that $\Delta^\sharp \subseteq \Delta'$. Let $F \in \Delta^\sharp$. Then for some t, $t{:}F \in \Delta$, i.e., $\Delta \in \mathcal{E}(t, F)$. By Stability, $\Gamma \in \mathcal{E}(t, F)$, hence $t{:}F \in \Gamma$ and $F \in \Gamma^\sharp$. By the definition of R, $F \in \Delta'$.

4. $\mathsf{JD45}$. The proof can be found in [44].

5. $\mathsf{JT45}$. For soundness, it suffices to check the Factivity Axiom, which easily follows from the reflexivity of R. For completeness, follow the footprints of 3 and note that R is reflexive. Indeed, $\Gamma^\sharp \subseteq \Gamma$ for reflexive theories.

The additional features of the canonical model are as follows: *R is an equivalence relation, the admissible evidence function does not distinguish equivalent worlds.* This follows easily from 5. $\Box$

Historical survey. The first Justification Logic system LP was introduced in 1995 in [2] (cf. also [4]). Such basic properties of Justification Logic as internalization, realization, arithmetical semantics [2; 4], symbolic models and complexity estimates ([16; 43; 49; 50]), and epistemic semantics and completeness [24; 25] were first established for LP.

A fair amount of work has already been done on Jusification Logics other than LP. Systems J, $\mathsf{J4}$, and JT were first considered in [15] under different names and in a slightly different setting[7]. $\mathsf{JT45}$ appeared independently in [54; 55] and [57; 58], and $\mathsf{JD45}$ in [54; 55]. $\mathsf{J45}$ has, perhaps, first been considered in this work. Systems combining epistemic modalities and justifications were studied in [5; 11; 12].

Mkrtychev semantics for J, JT, and $\mathsf{J4}$ with Completeness Theorem were found in [43]. Complexity bounds for LP and $\mathsf{J4}$ were found in [43; 49]. A comprehensive overview of all decidability and complexity results can be found in [44].

8 Forgetful Projection and the Correspondence Theorem

An intuitive connection between justification assertions and the justified belief modality $\Box$ involves the informal **existential quantifier**: $\Box F$ is read as

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

The language of Justification Logic does not have quantifiers over justifications, but instead has a sufficiently rich system of operations (polynomials) on justifications. We can use Skolem's idea of replacing quantifiers by functions and view Justification Logic systems as Skolemized logics of knowledge/belief. Naturally, to convert a Justification Logic sentence to the corresponding Epistemic Modal Logic sentence, one can use the **forgetful projection** '$\leadsto$' that replaces each occurrence of $t{:}F$ by $\Box F$.

Example: the sentence

$$x{:}P \rightarrow f(x){:}Q$$

can be regarded as a Skolem-style version of

$$\exists x(x{:}P) \rightarrow \exists y(y{:}Q),$$

which can be read as

$$\Box P \rightarrow \Box Q,$$

[7] [15] also considered variants of Justification Logic systems which, in our notations, would be called "JD" and "$\mathsf{JD4}$."

which is the forgetful projection of the original sentence $x{:}P \to f(x){:}Q$ (here, P, Q are assumed to be atomic sentences for simplicity's sake).

Examples (P, Q are atomic propositions):

$$t{:}P \to P \qquad \rightsquigarrow \qquad \Box P \to P,$$

$$t{:}P \to !t{:}(t{:}P) \qquad \rightsquigarrow \qquad \Box P \to \Box\Box P,$$

$$s{:}(P \to Q) \to (t{:}P \to (s\cdot t){:}Q) \qquad \rightsquigarrow \qquad \Box(P \to Q) \to (\Box P \to \Box Q).$$

Forgetful projection sometimes forgets too much, e.g., a logical triviality $x{:}P \to x{:}P$, a meaningful principle $x{:}P \to (x{+}y){:}P$, and a non-valid formula $x{:}P \to y{:}P$ have the same forgetful projection $\Box P \to \Box P$. However, '$\rightsquigarrow$' always maps valid formulas of Justification Logic to valid formulas of Epistemic Logic. 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 7. We assume that '$\rightsquigarrow$' is naturally extended from sentences to logics.

Theorem 7 [Consolidated Correspondence Theorem]

1. $\mathsf{J} \rightsquigarrow \mathsf{K}$
2. $\mathsf{JT} \rightsquigarrow \mathsf{T}$
3. $\mathsf{J4} \rightsquigarrow \mathsf{K4}$
4. $\mathsf{LP} \rightsquigarrow \mathsf{S4}$
5. $\mathsf{J45} \rightsquigarrow \mathsf{K45}$
6. $\mathsf{JD45} \rightsquigarrow \mathsf{KD45}$
7. $\mathsf{JT45} \rightsquigarrow \mathsf{S5}$

Proof. It is straightforward that the forgetful projection of each of the Justification Logic systems J, JT, J4, LP, J45, JD45, JT45 is derivable in the corresponding epistemic modal logics K, T, K4, S4, K45, KD45, S5, respectively.

The core of Theorem 7 is the Realization Theorem:

> *One can recover justification terms for all modal operators in valid principles of epistemic modal logics* K*,* T*,* K4*,* S4*,* K45*,* KD45*, and* S5 *such that the resulting formula is derivable in the corresponding Justification Logic system* J*,* JT*,* J4*,* LP*,* J45*,* JD45*, and* JT45*.*

The important feature of the Realization Theorem is that it recovers realizing functions according to the **existential reading of the modality**, i.e., negative occurrences of the modality are realized by (distinct) free variables, and the positive occurrences by justification polynomials, depending on these variables. For example, $\Box F \to \Box G$ will be realized by $x{:}F' \to f(x){:}G'$ where F', G' are realizations of F and G, respectively.

The Realization Theorem was first established for $\mathsf{S4}/\mathsf{LP}$ (case 4) in [2; 4], cases 1–3 are covered in [15]. The Realization Theorem for 7 is established in [58] using a very potent method from [25], and the proof for 5 and 6 is very similar to [25; 58] and can be safely omitted here. $\Box$

The Correspondence Theorem shows that the major epistemic modal logics K, $\mathsf{K4}$, $\mathsf{K45}$, $\mathsf{KD45}$ (for belief) and T, $\mathsf{S4}$, $\mathsf{S5}$ (for knowledge) have exact Justification Logic counterparts J, $\mathsf{J4}$, $\mathsf{J45}$, $\mathsf{JD45}$ (for partial justifications) and JT, LP, $\mathsf{JT45}$ (for factive justifications).

8.1 Foundational Consequences of the Correspondence Theorem

Is there anything new that we have learned from the Correspondence Theorem about epistemic modal logics?

First of all, this theorem provides a new semantics for major modal logics. In addition to the traditional Kripke-style 'universal' reading of $\Box F$ as

F holds in all possible situations,

there is now a rigorous 'existential' semantics for $\Box F$ that reads as

there is a witness (proof, justification) for F.

Perhaps the justification semantics plays a similar role in modal logic to that played by Kleene realizability in intuitionistic logic. In both cases, the intended semantics was **existential**: the Brouwer-Heyting-Kolmogorov interpretation of intuitionistic logic ([36; 65; 66]) and Gödel's provability reading of $\mathsf{S4}$ ([30; 31]). In both cases, a later possible-world semantics of **universal** character became a highly potent and dominant technical tool. However, in both cases, Kripke semantics did not solve the original semantical problems. It took Kleene realizability [39; 63] to reveal the computational semantics of intuitionistic logic and the Logic of Proofs [2; 4] to provide exact BHK semantics of proofs for intuitionistic and modal logic.

In the epistemic context, Justification Logic and the Correspondence Theorem add a new 'justification' component to modal logics of knowledge and belief. Again, this new component was in fact an old and central notion which has been widely discussed by mainstream epistemologists but has remained out of the scope of formal logical methods. The Correspondence Theorem tells us that justifications are compatible with Hintikka-style systems and hence can be regarded as a foundation for epistemic modal logic.

Another comparison suggests itself here: Skolem functions for first-order logic which provide a functional reading of quantifiers. It might seem that Skolem functions do not add much, since they do not suggest altering first-order logic. However, Skolem functions proved to be very useful for foundations (e.g., Henkin and Herbrand models, etc.), as well as for applications (Resolution, Logic Programming, etc.).

Note that the Realization Theorem is not at all trivial. For cases 1–4, realization algorithms are known that use cut-free derivations in the corresponding modal logics [2; 4; 15; 16]. For 5–7, the Realization Theorem has been established by Fitting's method or its proper modifications [25;

58]. In principle, these results also produce realization procedures which are based on exhaustive search.

It would be a mistake to draw the conclusion that **any** modal logic has a reasonable Justification Logic counterpart. For example, the logic of formal provability GL ([8; 14]) contains the *Löb Principle*

$$\Box(\Box F \rightarrow F) \rightarrow \Box F, \tag{20}$$

which does not seem to have an epistemically acceptable explicit version. Let us consider, for example, a case when F is the propositional constant $\perp$ for *false*. A Skolem-style reading of (20) suggests that there are justification terms s and t such that

$$x{:}(s{:}\perp \rightarrow \perp) \rightarrow t{:}\perp. \tag{21}$$

This is intuitively false for factive justification, though. Indeed, $s{:}\perp \rightarrow \perp$ is the Factivity Axiom. Apply Axiom Internalization R4 to obtain $c{:}[s{:}\perp \rightarrow \perp]$ for some constant c. This choice of c makes the antecedent of (21) intuitively true and the conclusion of (21) false[8]. In particular, (20) is not valid for proof interpretation (cf. [33] for a total account of which principles of GL are realizable).

9 Quantifier-Free First-Order Justification Logic

In this section, we extend J from the propositional language to the quantifier-free first-order language. To simplify formalities, we will regard here the first-order language without functional symbols, but with equality. Later, in Section 10, we will introduce definite descriptions in the form $\iota x F(x)$.

The language under consideration in this section is the first-order predicate language with individual variables and constants, predicate symbols of any arity and the equality symbol '=,' along with justification terms (including operations '$\cdot$' and '+') and the formula formation symbol ':' as in Section 3.3. Formulas are defined in the usual first-order way (without quantifiers) with an additional clause that if F is a formula and t is a justification polynomial, then $t{:}F$ is again a formula. The 'quantifier-free J' has all the axioms and rules of J, plus the equality axioms.

The formal system qfJ_0 has the following postulates:

A1. *Classical axioms of quantifier-free first-order logic with equality and Modus Ponens,*

A2. *Application Axiom* $s{:}(F \rightarrow G) \rightarrow (t{:}F \rightarrow (s \cdot t){:}G)$,

A3. *Monotonicity Axiom* $s{:}F \rightarrow (s+t){:}F$, $s{:}F \rightarrow (t+s){:}F$,

E1. $g = g$ *for any individual term* g *(reflexivity of equality)*;

E2. $f = g \rightarrow (P[f/x] \rightarrow P[g/x])$ *(substitutivity of equality)*, where f and g are individual terms, P is any atomic formula, $P[f/x]$ and $P[g/x]$ are the results of replacing all the occurrences of a variable x in P by f and g respectively; we will use notations $P(f)$, $P(g)$ for that.

[8]To be precise, we have to substitute c for x everywhere in s and t.

The system qfJ is qfJ_0 + R4, where

R4. *For each axiom A and any constants $e_1, e_2, \ldots, e_n$, infer $e_n\!:\!e_{n-1}\!:\!\ldots\!:\!e_1\!:\!A$.*

As in Section 3.4, we define Constant Specifications and systems qfJ_{CS}. In particular, $\mathsf{qfJ}_\emptyset$ is qfJ_0 and qfJ_{TCS} is qfJ.

The following proposition follows easily from the definitions.

Proposition 1 *Deduction Theorem holds for qfJ_{CS} for any constant specification CS. Internalization holds for qfJ_{CS} for an axiomatically appropriate constant specification CS.*

The following theorem provides a way to resolve the Frege puzzle ([28]) in an epistemic environment: equality of individual objects alone does not warrant substitutivity, but justified equality does.

Theorem 8 [Justified substitution] *For any individual terms f and g, justification variable u, and atomic formula $P(x)$, there is a justification term $s(u)$ such that qfJ proves*

$$u{:}(f = g) \rightarrow s(u){:}[P(f) \leftrightarrow P(g)].$$

The same holds for any qfJ_{CS} with an axiomatically appropriate constant specification CS.

Proof. Taking into account Example 1, it suffices to establish that for some $t(u)$,

$$u{:}(f = g) \rightarrow t(u){:}[P(f) \rightarrow P(g)].$$

From E2 it follows that qfJ proves

$$(f = g) \rightarrow [P(f) \rightarrow P(g)].$$

By R4, there is a justification constant c such that qfJ proves

$$c{:}\{(f = g) \rightarrow [P(f) \rightarrow P(g)]\}.$$

By A2, qfJ proves

$$c{:}\{(f = g) \rightarrow [P(f) \rightarrow P(g)]\} \rightarrow \{u{:}(f = g) \rightarrow (c{\cdot}u){:}[P(f) \rightarrow P(g)]\}.$$

By Modus Ponens, qfJ proves

$$u{:}(f = g) \rightarrow (c{\cdot}u){:}[P(f) \rightarrow P(g)].$$

It suffices now to pick $c{\cdot}u$ as $t(u)$. $\square$

An unjustified substitution can fail in qfJ. Namely, for any individual variables x and y, a predicate symbol P, and justification term s, the formula

$$(x = y) \rightarrow s{:}[P(x) \leftrightarrow P(y)] \tag{22}$$

is not valid. To establish this, one needs some model theory for qfJ.

We define qfJ-*models* as the usual first-order Kripke models[9] equipped with admissible evidence functions. A model is $(W, \{D_w\}, R, \mathcal{E}, \Vdash)$ such that the following properties hold.

[9] Equality is interpreted as identity in the model.

- W is an nonempty set of worlds.
- $\{D_w\}$ is the collection of nonempty domains D_w for each $w \in W$.
- R is the binary (accessibility) relation on W.
- $\mathcal{E}$ is the admissible evidence function which for each justification term t and formula F, returns the set of worlds $\mathcal{E}(t,F) \subseteq W$. Informally, these are the worlds where t is admissible evidence for F. We also assume that $\mathcal{E}$ satisfies the usual closure properties *Application* and *Sum* (Section 5).
- $\Vdash$ is the forcing (truth) relation such that

  $\Vdash$ assigns elements of D_w to individual variables and constants for each $w \in W$,

  for each n-ary predicate symbol P, and any $a_1, a_2, \ldots, a_n \in D_w$, it is specified whether $P(a_1, a_2, \ldots, a_n)$ holds in D_w,

  $\Vdash$ is extended to all the formulas by stipulating that

  $w \Vdash s = t$ iff '$\Vdash$' maps s and t to the same element of D_w,

  $w \Vdash P(t_1, t_2, \ldots, t_n)$ iff '$\Vdash$' maps t_i's to a_i's and $P(a_1, a_2, \ldots, a_n)$ holds in D_w,

  $w \Vdash F \wedge G$ iff $w \Vdash F$ and $w \Vdash G$,

  $w \Vdash \neg F$ iff $w \nVdash F$,

  $w \Vdash t{:}F$ iff $v \Vdash F$ for all v such that wRv, and $w \in \mathcal{E}(t,F)$.

The notion of a model respecting given constant specification is directly transfered from Section 5.

The following Theorem is established in the same manner as the soundness part of Theorem 2.

Theorem 9 *For any Constant Specification CS,* qfJ_{CS} *is sound with respect to the corresponding class of epistemic models.*

We are now ready to show that instances of unjustified substitution can fail in qfJ. To do this, it now suffices to build a qfJ-counter-model for (22) with the total constant specification. Obviously, the *maximal* $\mathcal{E}$ (i.e., $\mathcal{E}(t,F)$ contains each world for any t and F) respects any constant specification.

The Kripke-Fitting counter-model in Figure 1 exploits the traditional modal approach to refute a belief assertion by presenting a possible world where the object of this belief does not hold. In the picture, only true atomic formulas are shown next to possible worlds.

- $W = \{\mathbf{0}, \mathbf{1}\}$; $R = \{(\mathbf{0}, \mathbf{1})\}$; $D_0 = D_1 = \{a, b\}$;
- $\mathbf{1} \Vdash P(a)$ and $\mathbf{1} \nVdash P(b)$; the truth value of P at $\mathbf{0}$ does not matter;
- x and y are interpreted as a at $\mathbf{0}$; x is interpreted as a and y as b at $\mathbf{1}$;
- $\mathcal{E}$ is maximal at $\mathbf{0}$ and $\mathbf{1}$.

Obviously, $\mathbf{0} \Vdash x = y$. Since $\mathbf{1} \nVdash P(x) \leftrightarrow P(y)$, for any justification term s, $\mathbf{0} \nVdash s{:}[P(x) \leftrightarrow P(y)]$. Hence

$$\mathbf{0} \nVdash x = y \rightarrow s{:}[P(x) \leftrightarrow P(y)].$$

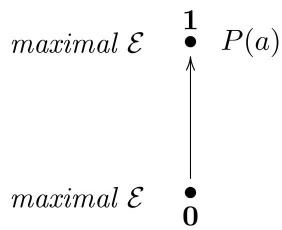

Figure 1: Fitting counter-model for unjustified substitution

10 Formalization of Gettier Examples

We consider Gettier's Case I in detail; Case II is much simpler logically and can be given similar treatment. We will present a complete formalization of Case I in qfJ with a definite description operation. Let

- $J(x)$ be the predicate *x gets the job*;
- $C(x)$ be the predicate *x has (ten) coins (in his pocket)*;
- `Jones` and `Smith` be individual constants denoting Jones and Smith, respectively[10];
- u be a justification variable.

10.1 Natural Model for Case I

Gettier's assumptions (d) and (e) contain a definite description

$$\textit{the man who will get the job.} \tag{23}$$

In this section, we will formalize Case I using a definite description ι-operation such that $\iota x P(x)$ is intended to denote

$$\textit{the } x \textit{ such that } P(x).$$

We interpret $\iota x P(x)$ in a given world of a qfJ-model as the element a such that $P(a)$ if there exists a unique a satisfying $P(a)$. Otherwise, $\iota x P(x)$ is undefined and any atomic formula where $\iota x P(x)$ actually occurs is taken to be false. Definite description terms are non-rigid designators: $\iota x P(x)$ may be given different interpretations in different worlds of the same qfJ-model (cf. [26]). The use of a definite description

Jones is the man who will get the job

as a justified belief by Smith hints that Smith has strong evidence for the fact that at most one person will get the job. This is implicit in Gettier's assumption.

We now present a Fitting model $\mathcal{M}$ which may be regarded as an exact epistemic formulation of Case I.

[10]Assuming that there are people seeking the job other than Jones and Smith does not change the analysis.

1. At the actual world $\mathbf{0}$, $J(\texttt{Smith})$, $C(\texttt{Smith})$, and $C(\texttt{Jones})$[11] hold and $J(\texttt{Jones})$ does not hold.

2. There is a possible belief world $\mathbf{1}$ for Smith at which $J(\texttt{Jones})$ and $\neg J(\texttt{Smith})$ hold. These conditions follow from proposition (d)

   Jones is the man who will get the job, and Jones has coins

   or, in logic form,
   $$(\texttt{Jones} = \iota x J(x)) \ \wedge \ C(\texttt{Jones})$$
   for which Smith has a strong evidence. In addition, Smith has no knowledge of 'Smith has coins' and there should be a possible world at which $C(\texttt{Smith})$ is false; we use $\mathbf{1}$ to represent this possibility.

3. World $\mathbf{1}$ is accessible from $\mathbf{0}$.

4. Smith has a *strong evidence of (d)*, which we will represent by introducing a justification variable u such that
   $$u{:}[(\texttt{Jones} = \iota x J(x)) \wedge C(\texttt{Jones})] \tag{24}$$
   holds at the actual world $\mathbf{0}$. We further assume that the admissible evidence function $\mathcal{E}$ respects the justification assertion (24), which yields
   $$\mathbf{0} \in \mathcal{E}(u, (\texttt{Jones} = \iota x J(x)) \wedge C(\texttt{Jones})).$$
   To keep things simple, we can assume that $\mathcal{E}$ is the maximal admissible evidence function, i.e., $\mathcal{E}(t, F) = \{\mathbf{0}, \mathbf{1}\}$ for each t, F.

These observations lead to the following model $\mathcal{M}$ on Figure 2.

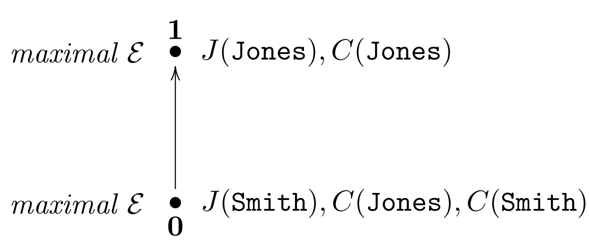

Figure 2: Natural Fitting model for Gettier Case I

[11] Strictly speaking, Case I explicitly states only that Smith has a strong evidence that $C(\texttt{Jones})$, which is not sufficient to conclude that $C(\texttt{Jones})$, since Smith's justifications are not necessarily factive. However, since the actual truth value of $C(\texttt{Jones})$ does not matter in Case I, we assume that in this instance, Smith's belief that $C(\texttt{Jones})$ was true.

- $W = \{\mathbf{0}, \mathbf{1}\}$; $R = \{(\mathbf{0}, \mathbf{1})\}$;
- $D_{\mathbf{0},\mathbf{1}} = \{\text{Jones}, \text{Smith}\}$, $\mathtt{Jones}$ is interpreted as 'Jones' and $\mathtt{Smith}$ as 'Smith';
- $\mathbf{0} \Vdash J(\mathtt{Smith}),\ C(\mathtt{Jones}),\ C(\mathtt{Smith}), \neg J(\mathtt{Jones})$;
- $\mathbf{1} \Vdash J(\mathtt{Jones}),\ C(\mathtt{Jones}),\ \neg J(\mathtt{Smith}), \neg C(\mathtt{Smith})$;
- $\iota x J(x)$ at $\mathbf{0}$ is interpreted as $\mathtt{Smith}$ and at $\mathbf{1}$ as $\mathtt{Jones}$;
- $\mathcal{E}$ is maximal at both $\mathbf{0}$ and $\mathbf{1}$.

It is interesting to compare this model with the axiomatic description of Case I. Here is the list of explicit assumptions:

$$J(\mathtt{Smith}),\quad C(\mathtt{Smith}),\quad C(\mathtt{Jones}),\quad \neg J(\mathtt{Jones}),\quad u{:}[(\mathtt{Jones} = \iota x J(x)) \wedge C(\mathtt{Jones})]. \tag{25}$$

It follows from the Soundness Theorem 9 that assumptions (25) provide a sound description of the actual world:

Proposition 2 $\mathsf{qfJ} + (25) \vdash F$ *entails* $\mathbf{0} \Vdash F$.

Example 3 The description of a model by (25) is not complete. For example, conditions (25) do not specifically indicate whether $t{:}C(\mathtt{Smith})$ holds at the actual world for some t, whereas it is clear from the model that $\mathbf{0} \nVdash t{:}C(\mathtt{Smith})$ for any t since $\mathbf{1} \nVdash C(\mathtt{Smith})$ and $\mathbf{1}$ is accessible from $\mathbf{0}$. Model $\mathcal{M}$ extends the set of assumptions (25) to a possible **complete** specification: every ground proposition F in the language of this example is either true or false at the 'actual' world $\mathbf{0}$ of the model.

10.2 Formalizing Gettier's Reasoning

Gettier's conclusion in Case I states that *Smith is justified in believing that 'The man who will get the job has ten coins in his pocket.'* In our formal language, this amounts to a statement that for some justification term t,

$$t{:}C(\iota x J(x)) \tag{26}$$

is derivable in qfJ from assumptions of Case I.

Theorem 10 *Gettier's conclusion $t{:}C(\iota x J(x))$ is derivable in qfJ from assumptions (25) of Case I. Furthermore, $t{:}C(\iota x J(x))$ holds at the 'actual world' $\mathbf{0}$ of the natural model $\mathcal{M}$ of Case I.*

Proof. In order to find t we may mimic Gettier's informal reasoning. First, we formally derive (e) (i.e., $C(\iota x J(x))$) from (d) (i.e., $\mathtt{Jones} = \iota x J(x) \wedge C(\mathtt{Jones})$) and then use the fact that (d) is justified (i.e., $u{:}[\mathtt{Jones} = \iota x J(x) \wedge C(\mathtt{Jones})]$). We will now show that this argument can be formalized in qfJ. Note that in qfJ, we may reason as follows:

1. $\mathtt{Jones} = \iota x J(x) \rightarrow [C(\mathtt{Jones}) \rightarrow C(\iota x J(x))]$, *an axiom of* qfJ*;*
2. $[\mathtt{Jones} = \iota x J(x) \wedge C(\mathtt{Jones})] \rightarrow C(\iota x J(x))$, *by propositional reasoning, from 1;*

3. $s{:}\{[\texttt{Jones} = \iota x J(x) \wedge C(\texttt{Jones})] \rightarrow C(\iota x J(x))\}$, *by Internalization, from 2;*
4. $u{:}[\texttt{Jones} = \iota x J(x) \wedge C(\texttt{Jones})] \rightarrow (s{\cdot}u){:}C(\iota x J(x))$, *by Axiom A2 and Modus Ponens, from 3;*
5. $u{:}[\texttt{Jones}{=}\iota x J(x) \wedge C(\texttt{Jones})]$, *an assumption from (25);*
6. $(s{\cdot}u){:}C(\iota x J(x))$, *by Modus Ponens, from 4 and 5.*

Now we can pick t to be $s{\cdot}u$. So,

$$\mathsf{qfJ} + (25) \vdash (s{\cdot}u){:}C(\iota x J(x))$$

and, by Proposition 2,

$$\mathbf{0} \Vdash (s{\cdot}u){:}C(\iota x J(x)).$$

$\square$

10.3 Eliminating Definite Descriptions, Russell-style

We can eliminate definite descriptions from Case I using, e.g., Russell's translation (cf. [27; 52; 59; 61]) of definite descriptions. According to Russell, $C(\iota x J(x))$ contains a hidden *uniqueness assumption* and reads as

$$\exists x[J(x) \wedge \forall y(J(y) \rightarrow y = x) \wedge C(x)], \tag{27}$$

and $\texttt{Jones} = \iota x J(x)$ as

$$J(\texttt{Jones}) \wedge \forall y(J(y) \rightarrow y = \texttt{Jones}). \tag{28}$$

In addition, in the universe of Case I consisting of two objects *Jones, Smith*, a universally quantified sentence $\forall y F(y)$ reads as

$$F(\texttt{Jones}) \wedge F(\texttt{Smith}),$$

and an existentially quantified statement $\exists x G(x)$ reads as

$$G(\texttt{Jones}) \vee G(\texttt{Smith}).$$

Taking into account all of these simplifying observations, we may assume that for Smith (and the reader), $\forall y(J(y) \rightarrow y = \texttt{Jones})$ reads as

$$[J(\texttt{Jones}) \rightarrow (\texttt{Jones} = \texttt{Jones})] \wedge [J(\texttt{Smith}) \rightarrow (\texttt{Smith} = \texttt{Jones})],$$

which is equivalent[12] to

$$\neg J(\texttt{Smith}).$$

Now, (28) is equivalent to

$$J(\texttt{Jones}) \wedge \neg J(\texttt{Smith}),$$

and the whole Gettier proposition (d) collapses to

$$J(\texttt{Jones}) \wedge \neg J(\texttt{Smith}) \wedge C(\texttt{Jones}). \tag{29}$$

[12] We assume that everybody is aware that $\texttt{Smith} \neq \texttt{Jones}$.

The assumption that (d) is justified for Smith can now be represented by

$$v{:}[J(\texttt{Jones}) \wedge \neg J(\texttt{Smith}) \wedge C(\texttt{Jones})], \tag{30}$$

for some justification variable v.

Smith's justified belief

$$\text{'the man who will get the job has coins,'} \tag{31}$$

according to Russell, should read as

$$\exists x[J(x) \wedge \forall y(J(y) \rightarrow y = x) \wedge C(x)]. \tag{32}$$

The same considerations as above show that

$$\forall y[J(y) \rightarrow (y = \texttt{Jones})]$$

is equivalent to

$$\neg J(\texttt{Smith}),$$

and

$$\forall y[J(y) \rightarrow (y = \texttt{Smith})]$$

is equivalent to

$$\neg J(\texttt{Jones}).$$

Since an existentially quantified formula $\exists x G(x)$ is logically equivalent to a disjunction $G(\texttt{Jones}) \vee G(\texttt{Smith})$, formula (32) is equivalent to

$$[J(\texttt{Jones}) \wedge \neg J(\texttt{Smith}) \wedge C(\texttt{Jones})] \vee [J(\texttt{Smith}) \wedge \neg J(\texttt{Jones}) \wedge C(\texttt{Smith})]. \tag{33}$$

Finally, the formalization of (31) in our language amounts to stating that for some justification term p,

$$p{:}\{[J(\texttt{Jones}) \wedge \neg J(\texttt{Smith}) \wedge C(\texttt{Jones})] \vee [J(\texttt{Smith}) \wedge \neg J(\texttt{Jones}) \wedge C(\texttt{Smith})]\}. \tag{34}$$

Theorem 11 *Gettier's claim (34) is derivable in* qfJ *from the assumption (30) of Case I, and holds in the 'actual world'* **0** *of the natural model* $\mathcal{M}$ *of Case I.*

Proof. After all the preliminary work and assumptions, there is not much left to do. We just note that (29) is a disjunct of (33). A derivation of (34) from (30) in qfJ reduces now to repeating steps of Example 2, which shows how to derive a justified disjunction from its justified disjunct. □

Comment 1 One can see clearly the essence of Gettier's example. In (33), one of two disjuncts is justified but false, whereas the other disjunct is unjustified but true. The resulting disjunction (33) is both justified and true, but not really known to Smith.

10.4 Hidden Uniqueness Assumption is Necessary

In this subsection, we study what happens if we deviate from Russell's reading of definite descriptions, in particular if we skip the uniqueness of the defined object. For example, let us read Gettier's proposition (d) as

$$\textit{Jones will get the job, and Jones has ten coins in his pocket,} \tag{35}$$

and proposition (e) as

$$\textit{A man who will get the job has ten coins in his pocket.} \tag{36}$$

Then a fair formalization of (35) would be

$$J(\texttt{Jones}) \wedge C(\texttt{Jones}), \tag{37}$$

and the assumption that (35) is justified for Smith is formalized as

$$u{:}[J(\texttt{Jones}) \wedge C(\texttt{Jones})]. \tag{38}$$

In this case, the set of explicitly made non-logical assumptions is

1. $u{:}[J(\texttt{Jones}) \wedge C(\texttt{Jones})]$, *assumption (38)*;
2. $\neg J(\texttt{Jones})$ *(Jones does not get the job)*;
3. $J(\texttt{Smith})$ *(Smith gets the job)*;
4. $C(\texttt{Smith})$ *(Smith has coins)*.

Condition (36) naturally formalizes as

$$[J(\texttt{Jones}) \rightarrow C(\texttt{Jones})] \wedge [J(\texttt{Smith}) \rightarrow C(\texttt{Smith})]. \tag{39}$$

The claim that (39) is justified for Smith is formalized as

$$t{:}\{[J(\texttt{Jones}) \rightarrow C(\texttt{Jones})] \wedge [J(\texttt{Smith}) \rightarrow C(\texttt{Smith})]\} \tag{40}$$

for some justification term t.

We show that the assumptions 1–4 above do not suffice for proving (40).

Proposition 3 *For any justification term t, formula (40) is not derivable in* qfJ *from assumptions 1–4.*

Proof. Suppose (40) is derivable in qfJ from assumptions 1–4. Then, by the Deduction Theorem, qfJ would derive

$$\textit{`Conjunction of 1–4'} \rightarrow \textit{(40)}. \tag{41}$$

It now suffices to build a Fitting qfJ-model (Figure 3) where (41) does not hold at a certain world.

At $\mathbf{0}$, all assumptions 1–4 hold, but (40) is false at $\mathbf{0}$ for all t's. Indeed, (39) is false at $\mathbf{1}$, since its conjunct

$$J(\texttt{Smith}) \rightarrow C(\texttt{Smith})$$

is false at $\mathbf{1}$, and $\mathbf{1}$ is accessible from $\mathbf{0}$. $\square$

$$\begin{tikzcd}[row sep=large]
\textit{maximal } \mathcal{E} & \overset{\mathbf{1}}{\bullet} & J(\texttt{Smith}), J(\texttt{Jones}), C(\texttt{Jones}) \\
\textit{maximal } \mathcal{E} & \underset{\mathbf{0}}{\bullet} \arrow[u] & J(\texttt{Smith}), C(\texttt{Jones}), C(\texttt{Smith})
\end{tikzcd}$$

Figure 3: Counter-model for Case I without uniqueness

10.5 Streamlined Case I: No Coins/Pockets Are Needed

In this subsection, we show that references to coins and pockets, as well as definite descriptions, are redundant for making the point in Gettier example Case I. Here is a simpler, streamlined case based on the same material.

Smith has strong evidence for the proposition:
(d) Jones will get the job.
Proposition (d) entails:
(e) Either Jones or Smith will get the job.
Let us suppose that Smith sees the entailment from (d) to (e), and accepts (e) on the grounds of (d), for which he has strong evidence. In this case, Smith is clearly justified in believing that (e) is true. But imagine further that unknown to Smith, he himself, not Jones, will get the job. Then
1) (e) is true,
2) Smith believes that (e) is true, and
3) Smith is justified in believing that (e) is true.
But it is equally clear that Smith does not know that (e) is true. . . .

In this version, the main assumption is

$$\textit{Smith has a strong evidence that Jones gets the job.} \tag{42}$$

Its straightforward formalization is

$$v{:}J(\texttt{Jones}). \tag{43}$$

The claim is that

$$\textit{Smith is justified in believing that either Jones or Smith will get the job.} \tag{44}$$

The natural formalization of the claim

$$t{:}[J(\texttt{Jones}) \vee J(\texttt{Smith})]. \tag{45}$$

The set of formal assumptions is

$$v{:}J(\texttt{Jones}), \quad J(\texttt{Smith}), \quad \neg J(\texttt{Jones}).$$

It is easy now to derive (45) in qfJ from assumption (43).

1. $v{:}J(\texttt{Jones})$, *assumption (43)*;
2. $J(\texttt{Jones})\rightarrow J(\texttt{Jones})\vee J(\texttt{Smith})$, *propositional axiom*;
3. $c{:}[J(\texttt{Jones})\rightarrow J(\texttt{Jones})\vee J(\texttt{Smith})]$, *from 2, by Axiom Internalization R4*;
4. $c{:}[J(\texttt{Jones})\rightarrow J(\texttt{Jones})\vee J(\texttt{Smith})]\rightarrow[v{:}J(\texttt{Jones})\rightarrow(c\cdot v){:}(J(\texttt{Jones})\vee J(\texttt{Smith}))]$, *Axiom A2*;
5. $(c\cdot v){:}[J(\texttt{Jones})\vee J(\texttt{Smith})]$, *from 4, 3, and 1, by Modus Ponens twice.*

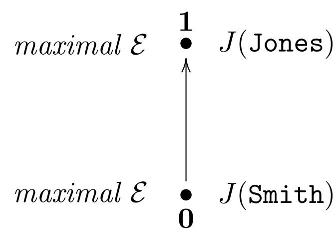

Figure 4: Natural Fitting model for the streamlined Case I

At the actual world $\mathbf{0}$, both hold:

$$J(\texttt{Jones})\vee J(\texttt{Smith})\ \text{(meaning \textit{(e) is true})}$$

and

$$(c\cdot v){:}[J(\texttt{Jones})\vee J(\texttt{Smith})]\ \text{(meaning \textit{(e) is justified})}.$$

The desired Gettier-style point is made on the same material but without the unnecessary use of quantifiers, definite descriptions, coins, and pockets.

It is fair to note, however, that Gettier example Case II in [29] does not have these kinds of redundancies and is logically similar to the streamlined version of Case I presented above.

11 Gettier Example and Factivity

Theorem 12 *Gettier assumptions (25) in Case I are inconsistent in Justification Logic systems with factive justifications.*

Proof. Here is an obvious derivation of a contradiction in qfJT from (25):

$u{:}[(\texttt{Jones}=\iota x J(x))\wedge C(\texttt{Jones})]$, *by (24)*;
$\texttt{Jones}=\iota x J(x)$, *by the Factivity Axiom and some propositional logic*;
$(\texttt{Jones}=\iota x J(x))\rightarrow J(\texttt{Jones})$, *an assumed natural property of definite descriptions*;
$J(\texttt{Jones})$, *by Modus Ponens. This contradicts the condition* $\neg J(\texttt{Jones})$ *from (25).* $\Box$

The question is, what we have learned about Justification, Belief, Knowledge, and other epistemic matters?

Within the domain of formal epistemology, we now have a basic logic machinery to study justifications and their connections with Belief and Knowledge. Formalizing Gettier is a case study that demonstrates the method.

We show that Gettier reasoning was formally correct, with some hidden assumptions related to definite descriptions. Gettier examples belong to the area of Justification Logic dealing with partial justifications and are inconsistent within Justification Logic systems of factive justifications and knowledge. All this, perhaps, does not come as a surprise to epistemologists. However, these observations show that models provided by Justification Logic behave in a reasonable manner.

For epistemology, these developments are furthering the study of justification, e.g., the search for the 'fourth condition' of the JTB definition of knowledge. Justification Logic provides systematic examples of epistemological principles such as Application, Monotonicity, Logical Awareness, and their combinations, which look plausible, at least, within the propositional domain. Further discussion on these and other Justification Logic principles could be an interesting contribution to this area.

12 Conclusions

Justification Logic extends the logic of knowledge by the formal theory of justification. Justification Logic has roots in mainstream epistemology, mathematical logic, computer science, and artificial intelligence. It is capable of formalizing a significant portion of reasoning about justifications. In particular, we have seen how to formalize Kripke, Russell, and Gettier examples in Justification Logic. This formalization has been used for the resolution of paradoxes, verification, hidden assumption analysis, and eliminating redundancies.

Among other known applications of Justification Logic, so far there are

- intended provability semantics for Gödel's provability logic $\mathsf{S4}$ with the Completeness Theorem ([2; 4]);
- formalization of Brouwer-Heyting-Kolmogorov semantics for intuitionistic propositional logic with the Completeness Theorem ([2; 4]);
- a general definition of the Logical Omniscience property, rigorous theorems that evidence assertions in Justification Logic are not logically omniscient ([10]). This provides a general framework for treating the problem of logical omniscience;
- an evidence-based approach to Common Knowledge (so-called Justified Common Knowledge) which provides a rigorous semantics to McCarthy's 'any fool knows' systems ([1; 5; 47]). Justified Common Knowledge offers formal systems which are less restrictive than the usual epistemic logics with Common Knowledge [5].
- analysis of Knower and Knowability paradoxes ([17; 18]).

It remains to be seen to what extent Justification Logic can be useful for analysis of empirical, perceptual, and *a priori* types of knowledge. From the perspective of Justification Logic, such knowledge may be considered as justified by constants (i.e., atomic justifications). Apparently, further discussion is needed here.

13 Acknowledgements

The author is very grateful to Walter Dean, Mel Fitting, Vladimir Krupski, Roman Kuznets, Elena Nogina, Tudor Protopopescu, and Ruili Ye, whose advice helped with this paper. Many thanks to Karen Kletter for editing this text. Thanks to audiences at the CUNY Graduate Center, Bern University, the Collegium Logicum in Vienna, and the 2nd International Workshop on Analytic Proof Systems for comments on earlier versions of this paper.

References

[1] E. Antonakos. Justified and Common Knowledge: Limited Conservativity. In S. Artemov and A. Nerode, editors, *Logical Foundations of Computer Science. International Symposium, LFCS 2007, New York, NY, USA, June 2007, Proceedings*, volume 4514 of *Lecture Notes in Computer Science*, pages 1–11. Springer, 2007.

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

[3] S. Artemov. Understanding constructive semantics. Spinoza Lecture for European Association for Logic, Language and Information, Utrecht, August 1999.

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

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

[6] S. Artemov. On two models of provability. In Dov M. Gabbay, Michael Zakharyaschev, and Sergei S. Goncharov, editors, *Mathematical Problems from Applied Logic II*, pages 1–52. Springer - New York, 2007.

[7] S. Artemov. Symmetric Logic of Proofs. In A. Avron, N. Dershowitz, and A. Rabinovich, editors, *Pillars of Computer Science, Essays Dedicated to Boris (Boaz) Trakhtenbrot on the Occasion of His 85th Birthday*, volume 4800 of *Lecture Notes in Computer Science*, pages 58–71. Springer, February 2008.

[8] S. Artemov and L. Beklemishev. Provability logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, 2nd ed.*, volume 13, pages 189–360. Springer, Dordrecht, 2005.

[9] S. Artemov, E. Kazakov, and D. Shapiro. Epistemic logic with justifications. Technical Report CFIS 99-12, Cornell University, 1999.

[10] S. Artemov and R. Kuznets. Logical omniscience via proof complexity. In *Computer Science Logic 2006*, volume 4207, pages 135–149. Springer Lecture Notes in Computer Science, 2006.

[11] 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.

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

[13] S. Artemov and T. Strassen. Functionality in the basic logic of proofs. Technical Report IAM 93-004, Department of Computer Science, University of Bern, Switzerland, 1993.

[14] G. Boolos. *The Logic of Provability*. Cambridge University Press, Cambridge, 1993.

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

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

[17] W. Dean and H. Kurokawa. From the knowability paradox to the existence of proofs. Manuscript (submitted to *Synthese*), 2007.

[18] W. Dean and H. Kurokawa. The knower paradox and the quantified logic of proofs. In Alexander Hieke, editor, *Austrian Ludwig Wittgenstein Society*, volume 31, August 2008.

[19] F. Dretske. Conclusive reasons. *Australasian Journal of Philosophy*, 49:1–22, 1971.

[20] F. Dretske. Is knowledge closed under known entailment? the case against closure. In M. Steup and E. Sosa, editors, *Contemporary Debates in Epistemology*, pages 13–26. Blackwell, 2005.

[21] R. Fagin and J. Halpern. Belief, awareness, and limited reasoning: Preliminary report. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence (IJCAI-85)*, pages 491–501, 1985.

[22] R. Fagin and J. Halpern. Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34(1):39–76, 1988.

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

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

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

[26] M. Fitting. Intensional Logic. Stanford Encyclopedia of Philosophy (http://plato.stanford.edu), February 2007.

[27] M. Fitting and R.L. Mendelsohn. *First-order Modal Logic*. Kluwer Academic, 1998.

[28] G. Frege. On sense and reference. In P. Geach and M. Black, editors, *Translations of the Philosophical Writings of Gottlob Frege*. Oxford: Blackwell, 1952.

[29] E. Gettier. Is Justified True Belief Knowledge? *Analysis*, 23:121–123, 1963.

[30] K. Gödel. Eine Interpretation des intuitionistischen Aussagenkalkuls. *Ergebnisse Math. Kolloq.*, 4:39–40, 1933. English translation in: S. Feferman et al., editors, *Kurt Gödel Collected Works, Vol. 1*, pages 301–303. Oxford University Press, Oxford, Clarendon Press, New York, 1986.

[31] K. Gödel. Vortrag bei Zilsel/Lecture at Zilsel's (*1938a). In Solomon Feferman, John W. Dawson, Jr., Warren Goldfarb, Charles Parsons, and Robert M. Solovay, editors, *Unpublished essays and lectures*, volume III of *Kurt Gödel Collected Works*, pages 86–113. Oxford University Press, 1995.

[32] A. Goldman. A causal theory of knowing. *The Journal of Philosophy*, 64:335–372, 1967.

[33] E. Goris. Explicit Proofs in Formal Provability Logic. In S. Artemov and A. Nerode, editors, *Logical Foundations of Computer Science. International Symposium, LFCS 2007, New York, NY, USA, June 2007, Proceedings*, volume 4514 of *Lecture Notes in Computer Science*, pages 241–253. Springer, 2007.

[34] V.F. Hendricks. Active Agents. *Journal of Logic, Language and Information*, 12(4):469–495, 2003.

[35] V.F. Hendricks. *Mainstream and Formal Epistemology*. New York: Cambridge University Press, 2005.

[36] A. Heyting. *Mathematische Grundlagenforschung. Intuitionismus. Beweistheorie*. Springer, Berlin, 1934.

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

[38] J. Hintikka. Impossible possible worlds vindicated. *Journal of Philosophical Logic*, 4:475–484, 1975.

[39] S. Kleene. On the interpretation of intuitionistic number theory. *The Journal of Symbolic Logic*, 10(4):109–124, 1945.

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

[41] V.N. Krupski. The single-conclusion proof logic and inference rules specification. *Annals of Pure and Applied Logic*, 113(1-3):181–206, 2001.

[42] V.N. Krupski. Referential logic of proofs. *Theoretical Computer Science*, 357(1):143–166, 2006.

[43] 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.

[44] R. Kuznets. *Complexity Issues in Justification Logic*. PhD thesis, CUNY Graduate Center, 2008. http://kuznets.googlepages.com/PhD.pdf.

[45] K. Lehrer and T. Paxson. Knowledge: undefeated justified true belief. *The Journal of Philosophy*, 66:1–22, 1969.

[46] S. Luper. The epistemic closure principle. Stanford Encyclopedia of Philosophy, 2005.

[47] J. McCarthy, M. Sato, T. Hayashi, and S. Igarishi. On the model theory of knowledge. Technical Report STAN-CS-78-667, Stanford University, 1978.

[48] J.-J. Ch. Meyer and W. van der Hoek. *Epistemic Logic for AI and Computer Science*. Cambridge, 1995.

[49] 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.

[50] 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.

[51] Y. Moses. Resource-bounded knowledge. In M. Vardi, editor, *Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge, March 7–9, 1988, Pacific Grove, California*, pages 261–276. Morgan Kaufmann Pbl., 1988.

[52] S. Neale. *Descriptions*. Cambridge: MIT Press Books, 1990.

[53] R. Nozick. *Philosophical Explanations*. Harvard University Press, 1981.

[54] E. Pacuit. A note on some explicit modal logics. 5th Panhellenic Logic Symposium, Athens, July 2005.

[55] E. Pacuit. A note on some explicit modal logics. Technical Report PP-2006-29, University of Amsterdam. ILLC Publications, 2006.

[56] R. Parikh. Knowledge and the problem of logical omniscience. In Z. Ras and M. Zemankova, editors, ISMIS-87 *(International Symposium on Methodology for Intellectual Systems)*, pages 432–439. North-Holland, 1987.

[57] N. Rubtsova. Evidence-based knowledge for S5. Logic Colloquium 2005, Athens, August 2005.

[58] N. Rubtsova. Evidence Reconstruction of Epistemic Modal Logic $\mathsf{S5}$. In *Computer Science - Theory and Applications. CSR 2006*, volume 3967 of *Lecture Notes in Computer Science*, pages 313–321. Springer, 2006.

[59] B. Russell. On Denoting. *Mind*, 14:479–493, 1905.

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

[61] B. Russell. *Introduction to Mathematical Philosophy*. London: George Allen and Unwin, 1919.

[62] R.C. Stalnaker. Knowledge, Belief and Counterfactual Reasoning in Games. *Economics and Philosophy*, 12:133–163, 1996.

[63] A.S. Troelstra. Realizability. In S. Buss, editor, *Handbook of Proof Theory*, pages 407–474. Elsevier, 1998.

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

[65] A.S. Troelstra and D. van Dalen. *Constructivism in Mathematics, Vols 1, 2*. North–Holland, Amsterdam, 1988.

[66] D. van Dalen. Intuitionistic logic. In D. Gabbay and F. Guenther, editors, *Handbook of Philosophical Logic. Volume 3*, pages 225–340. Reidel, 1986.

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

[68] 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.