2007

TR-2007019: Justification Logic

Sergei Artemov

Justification Logic

Sergei Artemov

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

October 27, 2007

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.

As a case study, we formalize Gettier examples in Justification Logic and reveal hidden assumptions and redundancies in Gettier reasoning. 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.

1 Introduction

Plato's celebrated account of Knowledge as Justified True Belief in the dialogues *Theaetetus* and *Meno* was widely accepted until 1963 when a paper by Edmund Gettier [26] opened the door to a broad philosophical discussion of the subject (cf. [17; 29; 42; 49; 57] and many others).

Meanwhile, commencing from seminal works [34; 62], 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 is adopted:

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

The resulting *Epistemic Logic* has been remarkably successful in terms of developing a rich mathematical theory and applications (cf. [20; 44], and other sources). However, the notion of justification, which has been an essential component of epistemic studies, was notoriously absent in the mathematical models of knowledge within the epistemic logic framework. This deficiency is displayed most prominemtly, in the *Logical Omniscience* defect of the modal logic of knowledge (cf. [18;

19; 35; 47; 51]). In the provability domain, the absence of an adequate description of the logic of justifications (here mathematical proofs) remained an impediment to both the formalizing of Brouwer-Heyting-Kolmorogorov semantics of proofs and providing a long-anticipated exact provability semantics for Gödel's provability logic $\mathsf{S4}$ ([3; 4; 6; 61]). This lack of a justification component has, perhaps, contributed to a certain gap between epistemic logic and mainstream epistemology ([31; 32]). We wish 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 argue that justifications have structure which can be formalized and studied by logical methods. We describe basic logical principles for justifications and relate them to mainstream epistemology and formal epistemology. We establish a general result (the Correspondence Theorem) stating that behind each major epistemic modal logic system, there is a robust system of justifications. This renders a new, evidence-based foundation for epistemic modal logic.*

Justification Logic extends the Logic of Knowledge in two major ways. It adds a long-anticipated formal notion of justification to epistemic logic/formal epistemology which makes logic more expressive. It validates the choice of logical principles in epistemic logic by providing them with a natural justification semantics.

The formal theory of justification has roots in the mathematical theory of proofs. It can be traced back to Brouwer-Heyting-Kolmogorov informal semantics for intuitionistic logic [60], typed combinatory logic (cf. [59]), Kleene realizability semantics for intuitionistic logic [36], Gödel's logic of provability [27], and Logic of Proofs [2; 4; 28].

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

$$t \textit{ is a justification of } F. \tag{2}$$

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

$$t \textit{ is accepted by agent as a justification of } F. \tag{3}$$

We believe that the language and tools of Justification Logic accommodate both readings of $t{:}F$.

Moreover, the Justification Logic is general enough to incorporate other semantics, not necessarily terminologically related to justifications or proofs. For example, $t{:}F$ can be read as

$$t \textit{ is a sufficient resource for } F. \tag{4}$$

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

$$F \textit{ satisfies conditions } t. \tag{5}$$

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.

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 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 even on the simplest Boolean base. After 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., Gettier examples), can be handled with the 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 always assumed to be factive.

4. In this paper, we consider the case of one agent only, although several examples of multi-agent Justification Logic systems have already been developed ([5; 12; 56]).

Formal logical methods do not directly solve philosophical problems, but rather provide a tool for analyzing assumptions and making sure 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 paper [26] and look at some classical post-Gettier sources to determine what logical principles in the given Justification Logic format (propositional Boolean logic with justification assertions $t\!:\!F$) can be extracted from there. 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 [17; 29; 42; 49; 57] may be readily formulated in the Boolean Justification Logic. The aforementioned papers are written in the 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 [28] and developed in full in [2; 4]. 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 *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.*

Axioms and rules of LP reflect the properties of the proof predicates in formal mathematical theories.

I. **Axioms of classical propositional logic**

Standard axioms of classical logic, e.g., **A1-A10** from [37]

II. **Axioms of the Logic of Proofs** LP

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

III. **Rules of inference**

$F,\ F\to G\vdash G$	(*Modus Ponens*)
$\vdash c{:}A$, where A is an axiom from I and II, and c is a proof constant	(*Axiom Internalization*)

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 $\vdash p{:}F$.

LP gives a complete axiomatization of the notion of mathematical proof with natural interpretations of operations '$\cdot$', '+', and '!'. We discuss these operations below in a more general epistemic setting.

2.2 Gettier Examples

Gettier in [26] described two situations, Case I and Case II, that were supposed to provide examples of justified true belief 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 [26].

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 S is justified in believing P, ... and S deduces Q from P ..., then S is justified in believing Q.*

Here is its natural formalization:

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

S deduces Q from P — "there is a deduction of $P \to Q$ available to S";

S 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:

$$\textit{if} \vdash F,\ \textit{then} \vdash s{:}F\ \textit{for some}\ s \tag{6}$$

and the Application Axiom:

$$s{:}(P \to Q) \to (x{:}P \to (s \cdot x){:}Q). \tag{7}$$

Indeed, suppose $x{:}P$ 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 $(s \cdot x){:}Q$.

2.3 Goldman's Reliabilism

Goldman in [29] offered the 'fourth condition' to be added to the Justified True Belief definition of knowledge. According to [29],

> *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 privides the Factivity Axiom for 'knowledge-producing' justifications

$$t{:}F \to F \qquad \textit{(Factivity Axiom)}. \tag{8}$$

The Factivity Axiom is assumed for *factive justifications* (systems JT, LP, JT45 below) but not for partial justifications (systems J, J4, J45, J45D).

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 LP and all other Justification Logic systems considered in this paper:

$$\textit{if F is valid, then one could construct a justification } p \textit{ such that } p{:}F \textit{ is valid.} \tag{9}$$

Internalization is usually represented in an equivalent (in the presence of the Completeness Theorem) form as a meta-rule (6). 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.'

In particular, Internalization entails that all explicit assumptions are justified for an agent. For example, if the Factivity Axiom $t : F \rightarrow F$ is assumed, then $s : (t : F \rightarrow F)$ should also hold for some justification s. However, Justication Logic provides a flexible approach to Internalization by offering the mechanism of *Constant Specifications* (Section 3.4) which can keep Internalization under control. In particular, each Justification Logic system has the option to model an absolutely skeptical agent who does not have justifications even for logical axioms. Such an agent may not believe in the Factivity Axiom $t{:}F \rightarrow F$ even if it is assumed true.

Principles (8) and (9) by no means represent Goldman's reliabilism in full, since it uses a much more elaborate natural language. However, the format and methods of Justification Logic could be helpful here too.

2.4 Lehrer and Paxson's Indefeasibility Condition

Lehrer and Paxson in [42] 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 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. Another plausible reading of 'further truth' could be 'further evidence,' and it leads to a fundamental principle of Justification Logic; 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

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 suggests 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

$$s{:}F \rightarrow (s+t){:}F \qquad \textit{(Monotonicity Axiom).} \tag{10}$$

Note that all principles (6), (7), (8), (9), and (10) are LP-compliant.

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: **partial 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 $s{:}(F \rightarrow G)$ and $t{:}F$, then $(s \cdot t){:}G$. Symbolically

$$s{:}(F \rightarrow G) \rightarrow (t{:}F \rightarrow (s \cdot t){:}G). \tag{11}$$

This is a basic property of justifications assumed in combinatory logic, λ-calculi, BHK-semantics, Kleene realizability, the Logic of Proofs LP, etc. Application corresponds to the closure under the known entailment principle

$$\mathbf{K}(F \rightarrow G) \rightarrow (\mathbf{K}F \rightarrow \mathbf{K}G), \tag{12}$$

which is widely accepted in formal epistemology. However, the principle (12) smuggles the *logical omniscience* defect into modal epistemic logic, which does not have the capacity to measure how hard it is to attain knowledge [18; 19; 35; 47; 51]. Justification Logic provides natural means of escaping logical omniscience by keeping track of the size of evidence terms [10].

3.2 Monotonicity of Justification

The *Monotonicity property* of justifications has been expressed by the operation **sum** '+' which can be read from (10). 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 t.

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

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.

From the technical point of view, assuming Monotonicity is necessary for connecting Jusitification Logic to Epistemic modal logic, Correspondence Theorem 11 holds only in the presence of Monotonicity. 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; 39; 40].

3.3 Basic Justification Logic J_0

Justification terms (polynomials) are built from justification variables $x, y, z, \ldots$ and justification constants $a, b, c, \ldots$ 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* $s{:}(F \rightarrow G) \rightarrow (t{:}F \rightarrow (s \cdot t){:}G)$,

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

J_0 is the logic of partial (not 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

$$u_1{:}A_1 \wedge u_2{:}A_2 \wedge \ldots \wedge u_n{:}A_n \ \rightarrow\ t{:}B.$$

Such conclusions have a natural reading:

If u_i are justifications for A_i for all $i = 1, 2, \ldots, n$, then t is a justification for B.

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

3.4 Logical Awareness and Justification Logic J

Logical Awareness principle states that logical axioms are justified *ex officio*. An agent accepts logical axioms (including the ones concerning justifications) as justified at any level of internalization. The natural way of formalizing this principle is postulating that

$$\text{for each evidence constant } c \text{ and for each axiom } A,\ c : A \text{ is again an axiom.} \tag{13}$$

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

A similar principle appeared in the Logic of Proofs LP; it has also been anticipated in Goldman's ([29]). Logical Awareness is an explicit reincarnation of the Necessitation Rule in modal epistemic logic:

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

applied to axioms.

Logic of Partial Justifications $\mathsf{J} = \mathsf{J}_0 + \mathrm{R4}$, where

R4. *Axiom Internalization Rule: for each axiom A and each constant c, $c{:}A$ is again an axiom.*

Constant Specifications. An inseparable part of a derived formula in a Justification Logic system is the list of *constant specifications* of form $c{:}A$ introduced in a derivation by the Rule R4. Indeed, one cannot properly read a Justification Logic theorem F without given specifications of justification constants occurring in F. Let us consider some basic examples of derivations in J.

Example 1 This example shows how to build a justification of a conjunction from justifications of both of the conjuncts.

1. $A\rightarrow(B\rightarrow(A\wedge B))$, *by A1*;
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 4, by propositional reasoning.*

The derived formula 5 contains constant c which was introduced in line 2 by R4. The constant specification here is

$$\{c{:}[A\rightarrow(B\rightarrow(A\wedge B))]\},$$

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.

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.*

Constant specification:

$$\{a{:}[A\rightarrow(A\vee B)],\quad b{:}[B\rightarrow(A\vee B)]\}.$$

The complete reading of the result of this derivation is

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

Note that the set of constant specifications associated with a given derivation is *closed downward*:

$$\textit{if } c_1{:}c_2{:}A \textit{ is in } CS\textit{, then } c_2{:}A \textit{ is in } CS \textit{ as well.}$$

Let us introduce a general definition of *Constant Specification* as a *downward closed set of formulas of form* $c{:}A$ *where* c *is a constant and* A *is an axiom.* Examples of Constant Specifications:

- *empty*: $CS = \emptyset$;
- *axiomatically appropriate*: for any axiom A, there is a constant c such that $c{:}A \in CS$;
- *Total Constant Specification*: $TCS = \{c{:}A \mid A \textit{ is any axiom and } c \textit{ is any constant}\}$.

Constant Specifications provide an alternative way to axiomatize Justification Logic systems and a flexible devise to keep justification assignments in check. For any Justification Logic L, by L_{CS} we understand L without rule R4 but with CS to be an additional set of axioms. In particular, $\mathsf{J}_{\emptyset}$ is the same system as J_0, J_{TCS} is J, etc.

Obviously,

$$\mathsf{L}_{CS} \textit{ proves } F \textit{ iff } \mathsf{L} \textit{ proves } F \textit{ with } \mathrm{R4} \textit{ limited to producing only formulas from } CS\ .$$

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 the specification of constants involved.

Theorem 1 J *enjoys the Deduction Theorem, closure under substitutions, and Internalization:*

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

Proof. The Deduction Theorem holds because J contains propositional axioms and *modus ponens* as the only rule of inference. The closure property follows from the schema-style formulation of the systems.

Internalization can be established by induction on derivation length. Suppose $\vdash F$. If F is an axiom, then by the Axiom Internalization rule *R4*, $c{:}F$ is again an axiom. If F is obtained by *Modus Ponens* from $X \rightarrow F$ and X, then, by the Induction Hypothesis, $\vdash s{:}(X \rightarrow F)$ 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. □

4 Basic Epistemic Semantics

The standard epistemic semantics for J has been provided by a proper adaptation of Fitting models [22] and Mkrtychev models [46].

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

An admissible evidence function satisfies the closure conditions with respect to operations '$\cdot$' and '$+$':

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

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

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

Given a model $\mathcal{M} = (W, R, \mathcal{A}, \Vdash)$, 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 $u \not\Vdash F$, etc.);

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

Note that an admissible evidence function $\mathcal{A}$ may be regarded as a Halpern-Moses awareness function [20] equipped with the structure of justifications.

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

For each given world $u \in W$ we define the *minimal admissible evidence function $\mathcal{A}$ respecting CS* at u. We first postulate that $u \in \mathcal{A}(c, A)$ for all formulas $c{:}A$ from CS and $\mathcal{A}(t, F) = \emptyset$ for all other t, F. Then we use the following completion steps:

- **Application**: *if $u \in \mathcal{A}(s, F \to G)$ and $u \in \mathcal{A}(t, F)$, then add u to $\mathcal{A}(s \cdot t, G)$.*

- **Sum**: *if $u \in \mathcal{A}(s, F)$ or $u \in \mathcal{A}(t, F)$, then add u to $\mathcal{A}(s+t, F)$.*

Obviously, for any CS and any model with a minimal admissible evidence function $\mathcal{A}$ respecting CS at each world, for each t and F, either $\mathcal{A}(t, F) = W$ (in this case we say that $\mathcal{A}(t, F)$ *holds in a model*) or $\mathcal{A}(t, F) = \emptyset$ ($\mathcal{A}(t, F)$ *does not hold in a model*).

Theorem 2 *For any Constant Specification CS,* J_{CS} *is sound and complete with respect to the corresponding class of Fitting models.*

Proof.

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

Soundness is straightforward. It suffices to check that all axioms of J are true at each node of any J-model. Let us check the axioms of J.

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

Axiom Internalization. Induction on R4. If F is an axiom from A1–A3, then it has already been proven to hold at each node. Suppose $c{:}A \in CS$ and A holds at each node. Then, by the definition of CS-model, $\mathcal{A}(c, A)$ holds at each node, hence $c{:}A$ does as well.

To establish completeness, we use standard canonical model construction. The canonical model $\mathcal{M} = (W, R, \mathcal{A}, \Vdash)$ for J is defined relative to a given Constant Specification CS:

- 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 $\Gamma^\sharp = \{F \mid t{:}F \in \Gamma \text{ for some } t\}$;
- $\mathcal{A}(s, F) = \{\Gamma \in W \mid s{:}F \in \Gamma\}$;
- $\Gamma \Vdash p$ 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 the 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, $\Gamma \in \mathcal{A}(t, G)$ by the definition of $\mathcal{A}$. Hence $\Gamma \Vdash t{:}G$, i.e., $\Gamma \Vdash F$.

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

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 ([20; 44]), extend it to a maximal consistent set Γ containing $\neg F$. 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 formally stronger.

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

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

Indeed, let $\Gamma \in \mathcal{A}(t,F)$. By the definition of $\mathcal{A}$, $t{:}F \in \Gamma$, 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:

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

This property holds in canonical models associated with axiomatically appropriate constant specifications *CS* (i.e. those which contain formulas $c : A$ for each axiom A). For such *CS*'s, the internalization property holds: if G is provable in a logic with an axiomatically appropriate *CS*, then $t{:}G$ is also provable there for some term t. Here is the proof of the Fully Explanatory property for canonical models with axiomatically appropriate constant specifications[2]. 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.

The Mkrtychev semantics is a predecessor of the Fitting semantics ([46]). *Mrktychev models* are Fitting models with a single world, and the proof of Theorem 2 can be easily modified to establish completeness of J with respect to Mkrtychev models.

Theorem 3 *For any Constant Specification CS,* J_{CS} *is sound and complete with respect to the corresponding class of Mrktychev models.*

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 $\Gamma \in \mathcal{A}(t,F)$, i.e., $t{:}F \in \Gamma$. This simplification immediately makes the Truth Lemma true.

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. □

Note that Mkrtychev models built in Theorem 3 are not reflexive, and have 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.

[2]This proof for LP was offered by Fitting in [22].

Theorem 3 shows that the information about Kripke structure in Fitting models can be completely encoded by the admissible evidence function. Mkrtychev models play an important theoretical role in Justification Logic [7; 16; 38; 41; 45]. On the other hand, as we will see in Section 6, Fitting models can be useful a counter-models with desired properties as they take into account both epistemic Kripke structure and evidence structure. The main application area of Fitting-style models is Justification Logic with both epistemic modalities and justification assertions (cf. [5; 12]).

As a corollary, we get a nice characterization of minimal admissible evidence functions.

Corollary 1 (cf. [38]) *Let $\mathcal{A}$ be the minimal admissible evidence function respecting CS. Then for any model $\mathcal{M}$*

$$\mathcal{A}(t,F) \text{ holds in } \mathcal{M} \text{ if and only if } \mathsf{J}_{CS} \vdash t{:}F.$$

Proof. 'If' is shown by the model argument. Suppose $\mathcal{A}(t,F)$ does not hold for the minimal admissible evidence function $\mathcal{A}$ respecting CS. Consider a Mkrtychev model for J_{CS} with the reflexive accessibility relation and with this $\mathcal{A}$ as an admissible evidence function. Obviously, $t{:}F$ is false in this model, since t is not an admissible evidence for F there. Contradiction.

'Only if' is established by induction on the number of Application and Sum steps in building $\mathcal{A}$ from CS. The base is obvious since for any $c{:}A \in CS$, $c{:}A$ is derivable in J_{CS}. Suppose the claim holds for assumptions of the Application completion step $\mathcal{A}(s, F\rightarrow G)$ and $\mathcal{A}(t,F)$. By Induction Hypothesis, both $s{:}(F\rightarrow G)$ and $t{:}F$ are derivable in J_{CS}. By Application Axiom and *Modus Ponens*, $(s{\cdot}t){:}G$ is derivable in J_{CS}. The case of Sum is similar.

Note that this proof uses only the 'soundness" part of Theorem 2. □

5 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 6, 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 (with operations '·' 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. qfJ (standing for 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.

The system qfJ is qfJ_0 + R4, where

R4. *Axiom Internalization Rule: for each axiom* A *and each constant* c, $c{:}A$ *is again an axiom.*

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 and Internalization hold for* qfJ.

Proposition 2 [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*

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

Proof. Taking into account Example 1, it suffices to establish

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

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 $s(u)$. □

However, Frege cases ([25]) can be represented in qfJ: 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

$$(x = y) \rightarrow s{:}[P(x) \leftrightarrow P(y)] \tag{15}$$

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

We define qfJ-*models* as the usual extensional[3] first-order Kripke models equipped with the admissible evidence function. A model is $(W, \{D_w\}, R, \mathcal{A}, \Vdash)$ such that the following properties hold.

- W is an nonempty set of worlds.
- $\{D_w\}$ is the collection of nonempty domains D_w, one for each $w \in W$.
- R is the binary (accessibility) relation on W.
- $\mathcal{A}$ is the admissible evidence function which for each justification term t, and formula F returns the set of worlds $\mathcal{A}(t, F) \subseteq W$. Informally, these are the worlds where t is admissible evidence for F. We also assume that $\mathcal{A}$ satisfies the usual closure properties *Application* and *Sum* (Section 4).
- $\Vdash$ 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 emulating 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 $w \in \mathcal{A}(t, F)$.

The notion of a model respecting given constant specification is directly transfered from the case of J-models, presented in Section 4. A direct analogue of Corollary 1 also holds for qfJ.

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

Theorem 4 *For any Constant Specification CS,* qfJ_{CS} *is sound with respect to the corresponding class of epistemic models.*

Corollary 2 *Let CS be a Constant Specification and $\mathcal{A}$ be the minimal admissible evidence function respecting CS. Then for any model $\mathcal{M}$*

$$\mathcal{A}(t, F) \textit{ holds in } \mathcal{M} \textit{ if and only if } \mathsf{qfJ}_{CS} \vdash t{:}F.$$

[3]A model is *extensional* if equality is interpreted as the identity in the model.

Proof. Similar to Corollary 1. □

Now we are 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 (15) with the total constant specification

$$TCS = \{c{:}A \mid c \textit{ is any constant and } A \textit{ is any axiom}\}.$$

Obviously, the maximal $\mathcal{A}$ (i.e. $\mathcal{A}(t, F)$ holds at each world for any t and F) respects TCS.

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.

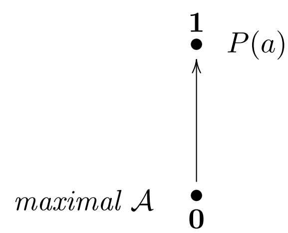

Figure 1: Fitting counter-model for unjustified substitution

- $W = \{\mathbf{0}, \mathbf{1}\}$; $R = \{(\mathbf{0}, \mathbf{1})\}$; $D_0 = D_1 = \{a, b\}$;
- $\mathbf{1} \Vdash P(a)$ 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 $\mathbf{1}$;
- $\mathcal{A}$ is maximal at $\mathbf{0}$.

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

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

Note that other instances of unjustified substitution for different, not necessarily atomic, $P(x)$ and $P(y)$ require different counter-models.

Figure 2 displays Mkrtychev counter-model, which uses an approach which is specific to Justification Logic where a belief assertion fails because the agent does not have evidence for it.

Figure 2: Mkrtychev counter model for unjustified substitution

- $W = \{\mathbf{0}\}$; $R = \emptyset$; $D_0 = \{a\}$; the truth value of P does not matter;

- x and y are interpreted as a;

- $\mathcal{A}$ is the minimal admissible evidence function respecting *TCS*.

To establish that

$$\mathbf{0} \not\Vdash x = y \rightarrow s{:}[P(x) \leftrightarrow P(y)],$$

it suffices to show that

$$\mathbf{0} \not\Vdash s{:}[P(x) \leftrightarrow P(y)].$$

To do that we must check that $\mathcal{A}(s, P(x) \leftrightarrow P(y))$ does not hold in the model. By Corollary 2, it suffices to show that

$$\mathsf{qfJ} \not\vdash s{:}[P(x) \leftrightarrow P(y)].$$

By soundness Theorem 4, it now suffices to find a qfJ-counter-model for $s{:}[P(x) \leftrightarrow P(y)]$. Consider a first-order counter-model $\mathcal{M}$ for $P(x) \leftrightarrow P(y)$. As in Corollary 1, make $\mathcal{M}$ a single-world qfJ-model by adding a reflexive accessibility relation and the minimal admissible evidence function respecting *TCS*. In this model, since $P(x) \leftrightarrow P(y)$ is false and the model is reflexive, $s{:}[P(x) \leftrightarrow P(y)]$ is false too, for any justification s.

In Mkrtychev counter-model,

$$\mathbf{0} \Vdash P(x) \leftrightarrow P(y),$$

which reflects the fact that there are situations in which $P(x)$ and $P(y)$ may be equivalent but at the same time an agent may simply not be aware of this fact. Note that Mkrtychev counter-model will also suffice to refute any instance of (15) for any formula P (not necessarily atomic) where $P(x) \leftrightarrow P(y)$ is not a tautology.

6 Formalization of Gettier Examples

We consider Gettier's Case I in detail; Case II is much simpler logically and can be given a 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[4];

- u be a justification variable.

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

6.1 Natural Models for Case I

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

$$\textit{the man who will get the job.} \qquad (16)$$

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 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. [23]). 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.

1. It the actual world $\mathbf{0}$, $J(\mathtt{Smith})$, $C(\mathtt{Smith})$, and $C(\mathtt{Jones})$[5] hold and $J(\mathtt{Jones})$ does not hold.

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

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

   or, in the logic form,

   $$(\mathtt{Jones}{=}\iota x J(x)) \ \wedge \ C(\mathtt{Jones})$$

   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 $C(\mathtt{Smith})$ is false; we use $\mathbf{1}$ to represent this possibility.

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

4. We assume the Logical Awareness principle according to which the agents (the observer/reader, Smith) follow all logical axioms and postulates of qfJ. In particular, $\mathcal{M}$ agrees with the Axiom Internalization Rule R4, hence the admissible evidence function $\mathcal{A}$ respects the total constant specification

   $$\mathit{TCS} = \{c{:}A \mid c \textit{ is a constant and } A \textit{ is an axiom}\}.$$

   In fact, a small finite fragment of *TCS* suffices to reproduce Gettier's reasoning in Case I.

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

5. Smith has a *strong evidence of (d)*, which we will represent by introducing a justification variable u such that

$$u{:}[(\mathtt{Jones}\!=\!\iota x J(x)) \wedge C(\mathtt{Jones})] \tag{17}$$

holds at the actual world $\mathbf{0}$. We further assume that the admissible evidence function $\mathcal{A}$ respects both *TCS* and the justification assertion (17), which yields

$$\mathbf{0} \in \mathcal{A}(u, (\mathtt{Jones}\!=\!\iota x J(x)) \wedge C(\mathtt{Jones})).$$

6. Since no other features of the model may be read off directly from the Case I description, no other structure features are assumed of $\mathcal{M}$. In particular, $\mathcal{A}$ is the minimal admissible evidence function at $\mathbf{0}$ that respects *TCS* and (17).

These observations lead to the following model $\mathcal{M}$ on Figure 3. At each location, the admissible evidence function $\mathcal{A}$ is the minimal one respecting *TCS* and conditions explicitly mentioned next to this location.

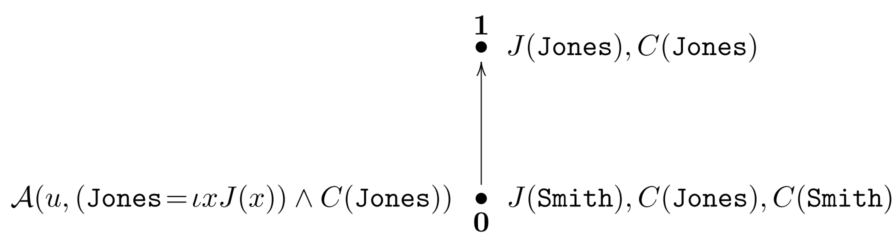

Figure 3: Natural Fitting model for Gettier Case I

- $W = \{\mathbf{0}, \mathbf{1}\}$; $\quad R = \{(\mathbf{0}, \mathbf{1})\}$;
- $D_{\mathbf{0},\mathbf{1}} = \{\mathtt{Jones}, \mathtt{Smith}\}$, Jones is interpreted as Jones and 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 Smith and at $\mathbf{1}$ as Jones;
- $\mathcal{A}$ at $\mathbf{0}$ is the minimal evidence function respecting the total constant specification *TCS* and (17), $\mathcal{A}$ at $\mathbf{1}$ is the minimal evidence function respecting *TCS*.

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

$$J(\mathtt{Smith}),\quad C(\mathtt{Smith}),\quad C(\mathtt{Jones}),\quad \neg J(\mathtt{Jones}),\quad u{:}[(\mathtt{Jones}\!=\!\iota x J(x)) \wedge C(\mathtt{Jones})]. \tag{18}$$

It follows from the Soundness Theorem 4 that assumptions (18) provide a sound description of the actual world:

Proposition 3 $\mathsf{qfJ} + (18) \vdash F$ *entails* $\mathbf{0} \Vdash F$.

Example 3 The description of a model by (18) is not complete. For example, conditions (18) do not specifically indicate whether $t{:}C(\texttt{Smith})$ holds at the actual world for some t whereas it is clear from the model that $\mathbf{0} \nVdash t{:}C(\texttt{Smith})$ for any t since $\mathbf{1} \nVdash C(\texttt{Smith})$ and $\mathbf{1}$ is accessible from $\mathbf{0}$. $\mathcal{M}$ provides a **complete** formal description of Case I: every ground proposition F is either true or false at the 'actual' world $\mathbf{0}$ of the model.

Fitting model $\mathcal{M}$ is not the only natural model for Case I: Figure 4 shows the *natural Mkrtychev model* $\mathcal{M}'$ which looks simpler and also does the job.

$$\mathcal{A}(u, (\texttt{Jones} = \iota x J(x)) \wedge C(\texttt{Jones})) \;\underset{\mathbf{0}}{\bullet}\; J(\texttt{Smith}), C(\texttt{Jones}), C(\texttt{Smith})$$

Figure 4: Natural Mkrtychev model for Gettier Case I

- $W' = \{\mathbf{0}\}$; $\quad R' = \emptyset$;
- $D'_{\mathbf{0}} = \{\texttt{Jones}, \texttt{Smith}\}$, $\texttt{Jones}$ is interpreted as $\texttt{Jones}$ and $\texttt{Smith}$ as $\texttt{Smith}$;
- $\mathbf{0} \Vdash' J(\texttt{Smith}),\ C(\texttt{Jones}),\ C(\texttt{Smith}), \neg J(\texttt{Jones})$;
- $\iota x J(x)$ is interpreted as $\texttt{Smith}$;
- the admissible evidence function is the same as in model $\mathcal{M}$.

It is easy to notice that $\mathcal{M}'$ is obtained from $\mathcal{M}$ by deleting the belief world and relying on the admissible evidence function only to decide which propositions hold at the actual world $\mathbf{0}$. This works: Mkrtychev model $\mathcal{M}'$ is equivalent to Fitting model $\mathcal{M}$ as far as the truth relation at the actual world is concerned.

Proposition 4 *For any proposition F,*

$$\mathbf{0} \Vdash F \quad \textit{if and only if} \quad \mathbf{0} \Vdash' F.$$

Proof. Induction on F. It suffices to check atomic formulas and justification assertions, since each proposition is their Boolean combination. Atomic formulas have the same truth value at the actual worlds of $\mathcal{M}$ and $\mathcal{M}'$. Consider a proposition $t{:}F$. If $\mathsf{qfJ} + (17) \vdash t{:}F$, then both $\mathbf{0} \Vdash t{:}F$ and $\mathbf{0} \Vdash' t{:}F$ (soundness property, similar to Proposition 3). If $\mathsf{qfJ} + (17) \nvdash t{:}F$, then $\mathcal{A}(t, F)$ does not hold at $\mathbf{0}$ of both models $\mathcal{M}$ and $\mathcal{M}'$ (similar to Corollary 1). In this case, both $\mathbf{0} \nVdash t{:}F$ and $\mathbf{0} \nVdash' t{:}F$. $\square$

Which model is better, $\mathcal{M}$ or $\mathcal{M}'$? Mkrtychev model $\mathcal{M}'$ looks simpler, but this may be a deceiving simplicity. If one needs to show that $\mathbf{0} \nVdash t{:}C(\texttt{Smith})$ for any t, then it is immediate in $\mathcal{M}$, but requires a special reasoning in $\mathcal{M}'$ (e.g., to show that $\mathsf{qfJ} + (17) \not\vdash t{:}C(\mathtt{Smith})$, for which one might want to use a model like $\mathcal{M}$).

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,

$$t{:}C(\iota x J(x)) \tag{19}$$

is derivable in qfJ from assumptions of Case I.

Theorem 5 *Gettier's conclusion $t{:}C(\iota x J(x))$ is derivable in qfJ from assumptions (18) of Case I. $t{:}C(\iota x J(x))$ 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 show now 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. $s{:}\{[\mathtt{Jones} = \iota x J(x) \wedge C(\mathtt{Jones})] \rightarrow C(\iota x J(x))\}$, *by Internalization, from 2;*
4. $u{:}[\mathtt{Jones} = \iota x J(x) \wedge C(\mathtt{Jones})] \rightarrow (s \cdot u){:}C(\iota x J(x))$, *by Axiom A2 and Modus Ponens, from 3;*

At this point we invoke our assumptions (18)

5. $u{:}[\mathtt{Jones} = \iota x J(x) \wedge C(\mathtt{Jones})]$, *an assumption from (18);*
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,

$$\mathsf{qfJ} + (18) \vdash (s \cdot u){:}C(\iota x J(x))$$

and, by Proposition 3,

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

□

So, Gettier's claim that *Smith is justified in believing that 'The man who will get the job has ten coins in his pocket'* has been confirmed in our formal model of Case I.

6.2 Eliminating Definite Description, Russell's style

We can eliminate definite descriptions from Case I using, e.g., Russell's translation (cf. [24; 48; 54; 55]) of definite descriptions. According to Russell, $C(\iota x J(x))$ contains a hidden *uniqueness assumption* and reads as

$$\exists x[J(x) \wedge \forall y(J(y) \rightarrow y = x) \wedge C(x)], \tag{20}$$

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

$$J(\mathtt{Jones}) \wedge \forall y(J(y) \rightarrow y = \mathtt{Jones}). \tag{21}$$

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

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

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[6] to

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

Now, (21) is equivalent to

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

and the whole Gettier proposition (d) collapses to

$$J(\texttt{Jones}) \wedge \neg J(\texttt{Smith}) \wedge C(\texttt{Jones}). \tag{22}$$

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

$$v{:}[J(\texttt{Jones}) \wedge \neg J(\texttt{Smith}) \wedge C(\texttt{Jones})], \tag{23}$$

for some justification variable v.

Smith's justified belief

'the man who will get the job has coins,' (24)

according to Russell, should read as

$$\exists x[J(x) \wedge \forall y(J(y) \rightarrow y = x) \wedge C(x)]. \tag{25}$$

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 (25) is equivalent to

$$[J(\texttt{Jones}) \wedge \neg J(\texttt{Smith}) \wedge C(\texttt{Jones})] \vee [J(\texttt{Smith}) \wedge \neg J(\texttt{Jones}) \wedge C(\texttt{Smith})]. \tag{26}$$

Finally, the formalization of (24) in our language amounts to stating that for some justification term p,

$$p{:}\{[J(\texttt{Jones}) \wedge \neg J(\texttt{Smith}) \wedge C(\texttt{Jones})] \vee [J(\texttt{Smith}) \wedge \neg J(\texttt{Jones}) \wedge C(\texttt{Smith})]\}. \tag{27}$$

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

Theorem 6 *Gettier's claim (27) is derivable in* qfJ *from the assumption (23) 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 (22) is a disjunct of (26). A derivation of (27) from (23) in qfJ reduces now to repeating steps of Example 2, which shows how to derive a justified disjunction from its justified disjunct. □

Figures 5 and 6 show the natural Fitting and Mkrtychev modes for Case I under Russell-style simplifications.

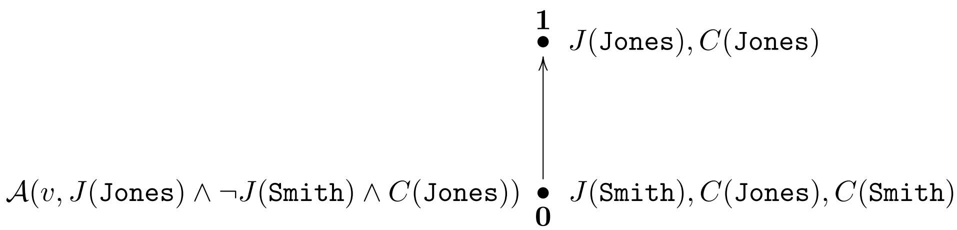

Figure 5: Natural Fitting model for Case I with eliminated definite description

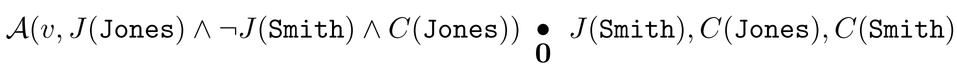

Figure 6: Natural Mkrtychev model for Case I with eliminated definite description

Comment 1 One can see clearly the essence of Gettier's example. In proposition (26), one of two disjuncts is justified but false, whereas the other disjunct is unjustified but true. The resulting disjunction (26) is both justified and true, but not really known to Smith.

6.3 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

$$\textit{Jones will get the job, and Jones has ten coins in his pocket,} \qquad (28)$$

and proposition (e) as

$$\textit{A man who will get the job has ten coins in his pocket.} \tag{29}$$

Then a fair formalization of (28) would be

$$J(\texttt{Jones}) \wedge C(\texttt{Jones}), \tag{30}$$

and the assumption that (28) is justified for Smith is formalized as

$$u{:}[J(\texttt{Jones}) \wedge C(\texttt{Jones})]. \tag{31}$$

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

1. $u{:}[J(\texttt{Jones}) \wedge C(\texttt{Jones})]$, *assumption (31)*;
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 (29) naturally formalizes as

$$[J(\texttt{Jones}) \rightarrow C(\texttt{Jones})] \wedge [J(\texttt{Smith}) \rightarrow C(\texttt{Smith})]. \tag{32}$$

The claim that (32) is justified for Smith is formalized as

$$t{:}\{[J(\texttt{Jones}) \rightarrow C(\texttt{Jones})] \wedge [J(\texttt{Smith}) \rightarrow C(\texttt{Smith})]\} \tag{33}$$

for some justification term t.

We show that the assumptions 1–4 above do not suffice for proving (33).

Proposition 5 *For any justification term t, formula (33) is not derivable in Justification Logic* qfJ *from assumptions 1–4.*

Proof. Suppose (33) is derivable in qfJ from assumptions 1–4. Then, by the Deduction Theorem, qfJ would derive

$$\textit{`Conjunction of 1–4 '} \rightarrow \textit{(33)}. \tag{34}$$

It now suffices to build a Fitting qfJ-model (Figure 7) where (34) does not hold at a certain world.

At **0**, all assumptions 1–4 hold, but (33) is false at **0** for all t's. Indeed, (32) is false at **1**, since its conjunct

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

is false at **1**, and **1** is accessible from **0**. □

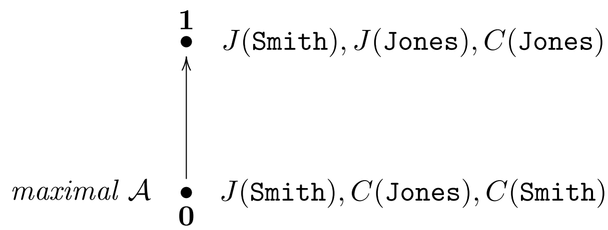

Figure 7: Counter-model for Case I without uniqueness

6.4 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

$$\textit{Smith has a strong evidence that Jones gets the job.} \tag{35}$$

Its straightforward formalization is

$$v{:}J(\texttt{Jones}). \tag{36}$$

The claim is that

$$\textit{Smith is justified in believing that either Jones or Smith will get the job.} \tag{37}$$

The natural formalization of the claim

$$t{:}[J(\texttt{Jones}) \vee J(\texttt{Smith})]. \tag{38}$$

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 (38) in qfJ from assumption 1.

1. $v{:}J(\mathtt{Jones})$, *assumption (36)*;
2. $J(\mathtt{Jones}) \rightarrow J(\mathtt{Jones}) \vee J(\mathtt{Smith})$, *propositional axiom*;
3. $c{:}[J(\mathtt{Jones}) \rightarrow J(\mathtt{Jones}) \vee J(\mathtt{Smith})]$, *from 2, by Axiom Internalization R4*;
4. $c{:}[J(\mathtt{Jones}) \rightarrow J(\mathtt{Jones}) \vee J(\mathtt{Smith})] \rightarrow [v{:}J(\mathtt{Jones}) \rightarrow (c{\cdot}v){:}(J(\mathtt{Jones}) \vee J(\mathtt{Smith}))]$, *Axiom A2*;
5. $(c{\cdot}v){:}[J(\mathtt{Jones}) \vee J(\mathtt{Smith})]$, *from 4, 3, and 1, by Modus Ponens twice.*

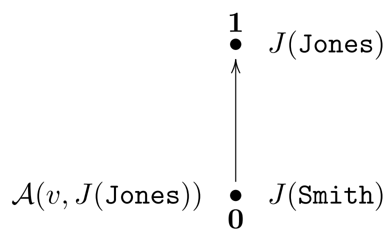

Figure 8: Natural Fitting model for the streamlined Case I

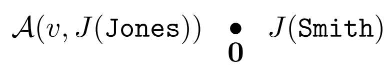

Figure 9: Natural Mkrtychev model for the streamlined Case I

At the actual world **0**, both hold:

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

and

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

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

It is fair to notice, however, that Gettier example Case II in [26] does not have these kinds of redundancies and is logically similar to the streamlined version of Case I presented above.

7 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

$$t{:}F \rightarrow F \tag{39}$$

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

$$\mathbf{K}F \rightarrow F \tag{40}$$

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

The Factivity Axiom first appeared in the Logic of Proofs LP as a principal feature of mathematical proofs. Indeed, in this setting (39) is valid: if there is a mathematical proof t of F, then F must be true.

Gettier examples demonstrate that a partial (i.e., not factive) justification is not sufficient to entail knowledge regardless of whether the belief is true. There should be a vehicle that connects the truth of F with the knowledge of F, and this vehicle should sufficiently reflect what is required for F to be true to warrant an agent's belief in F. Such a vehicle is what we call a factive justification.

Logic of Factive Justifications:

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

with

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

Systems JT_{CS} corresponding to Constant Specifications CS are defined as in Section 4.

JT-models are J-models with reflexive R's. The reflexivity condition makes each possible world accessible from itself which exactly corresponds to the Factivity Axiom.

Theorem 7 *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 $\Gamma R \Gamma$. □

Theorem 8 *Gettier assumptions (18) in Case I are inconsistent in Justification Logic systems with factive justifications.*

Proof. Here is an obvious derivation of a contradiction in qfJT from (18):

$u{:}[(\mathtt{Jones}=\iota x J(x)) \wedge C(\mathtt{Jones})]$, *by (17)*;
$\mathtt{Jones}=\iota x J(x)$, *by the Factivity Axiom and some propositional logic*;
$(\mathtt{Jones}=\iota x J(x)) \rightarrow J(\mathtt{Jones})$, *an assumed natural property of definite descriptions*;
$J(\mathtt{Jones})$, *by Modus Ponens. This contradicts the condition* $\neg J(\mathtt{Jones})$ *from (18).* □

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. Gettier examples 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 in this area.

7.1 Beyond Gettier: Justified True Belief With Wrong Reasons

Here is an example of a more subtle epistemic scenario which can be meaningfully analyzed in Justification Logic.

Smith has a strong piece of evidence (b) that he will obtain a faculty position (proposition F) based on the fact that his book on epistemology is universally admired. In addition, Smith has good reason (v) to believe that his earlier Silicon Valley experience is also quite sufficient to win this job. In fact, the hiring committee could not care less about epistemology and Smith gets the job based on his Silicon Valley experience. So Smith's belief that F based on 'b' is a case of Justified True Belief, but not knowledge.

Here Smith's belief is both justified and true, but the justification 'b' is not the one which is responsible for knowledge.

At first glance, it looks like Justification Logic cannot rule out b as a factive justification for F in this setting. Indeed, in J from the natural set of assumptions

$$\{b{:}F,\ v{:}F,\ v{:}F \rightarrow F\}, \tag{41}$$

it follows that $b{:}F \rightarrow F$ and even $t{:}(b{:}F \rightarrow F)$ for an appropriate t by simple propositional reasoning. Let us establish the former.

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

However, this derivation utilizes the fact that v is a factive justification for F to conclude $b{:}F \to F$.

When analyzing evidence b, we may exclude influence from evidence v by considering only *the class $\mathcal{P}$ of justification polynomials that do not contain v.* The set of assumptions (41) collapses to

$$\{b{:}F\}. \tag{42}$$

It is now an easy exercise in Justification Logic to show that b is not a factive justification if we are barred from assumptions involving v. First, we notice that logic J over $\mathcal{P}$ (axioms and rules of J which do not contain v) is sound (and complete) with respect to J-models which admissible evidence functions and forcing relations deal only with terms from $\mathcal{P}$. To establish that, it is sufficient to exclude v from the language of J in Sections 3.3 and 4. Now we claim that $b{:}F \to F$ is not provable in J over $\mathcal{P}$ from $b{:}F$, where b is a justification variable and F is a propositional variable.

Figure 10 presents an easy J-counter-model for $b{:}F \to F$ with the maximal admissible function $\mathcal{A}$ over $\mathcal{P}$: $\mathcal{A}(s, G)$ holds at $\mathbf{0}$ for any $s \in \mathcal{P}$ with terms from $\mathcal{P}$ and any G.

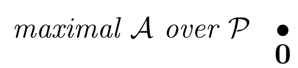

Figure 10: Counter-model for factivity of $b{:}F$

At $\mathbf{0}$, $b{:}F$ holds, since b is an admissible evidence for F at $\mathbf{0}$ and there are no possible worlds accessible from $\mathbf{0}$. On the other hand, $\mathbf{0} \not\Vdash b{:}F \to F$ since F does not hold at $\mathbf{0}$.

8 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.

8.1 Positive Introspection

One of the fundamental 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 the adequate explicit counterpart: the fact that 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 $t{:}F \rightarrow !t{:}(t{:}F)$. Positive Introspection in this operational form first appeared in the Logic of Proofs LP [2; 4].

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

and

$$\mathsf{LP} = \mathsf{JT} + \mathrm{A5},^{7}$$

with

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

Note that in the presence of the Positive Introspection Axiom, one could limit the scope of the Axiom Internalization Rule to making internalization of axioms which are not yet in the form $c{:}A$. This is how it has been done in LP: the Axiom Internalization can then be emulated by using $!!c{:}(!c{:}(c{:}A))$ instead of $a{:}(b{:}(c{:}A))$, etc. Such modifications are minor and they do not affect the main theorems and applications of Justification Logic.

8.2 Negative Introspection

Pacuit and Rubtsova considered in [50; 52; 53] 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 giving conclusive verification judgments about 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 for computer proof verifiers, proof checkers in formal theories, etc. There is a nuance in this motivation though: the examples of proof verifiers and proof checkers work with both t and F as an input, 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[8]. Anyway, 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

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

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

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

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

and

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

Theorem 9 J4, LP, J45, JD45, *and* JT45 *enjoy the Deduction Theorem, closure under substitution, and Internalization:*

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

Proof. Similar to Theorem 1. □

8.3 More Epistemic Models

Now we define epistemic models for other Justification Logic systems.

- J4-models are J-models with *transitive* R and two additional conditions:

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

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

- LP-models are J4-models with *reflexive* R (these are the original Fitting models);

- J45-models are J4-models satisfying conditions:

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

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

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

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

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

- JT45-models are 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 JT45-model, hence they could be included into the formulation of the Completeness Theorem.

Theorem 10 *Each of the logics* J4, LP, J45, JD45, JT45 *for any Constant Specification is sound and complete with respect to the corresponding class of epistemic models.*

Proof. We will follow the footsteps 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 $u \in \mathcal{A}(t, F)$ and $v \Vdash F$ for each v such that uRv. By the closure condition, $u \in \mathcal{A}(!t, t{:}F)$, and it remains to check that $v \Vdash t{:}F$. By monotonicity of $\mathcal{A}$, $v \in \mathcal{A}(t, F)$. 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 $\mathcal{A}$ is monotone, and the additional closure condition on $\mathcal{A}$ holds.

Monotonicity. Suppose $\Gamma R \Delta$ and $\Gamma \in \mathcal{A}(t, F)$, 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., $\Delta \in \mathcal{A}(t, F)$.

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

2. LP. This is the well-studied case of the Logic of Proofs, cf. [22].

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 Strong Evidence condition, $u \notin \mathcal{A}(t, F)$. By Negative Introspection closure, $u \in \mathcal{A}(?t, \neg t{:}F)$. 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 $\Gamma \notin \mathcal{A}(t, F)$. Then $t:F \notin \Gamma$. By maximality, $\neg t{:}F \in \Gamma$. By Negative Introspection Axiom $?t{:}(\neg t{:}F) \in \Gamma$, hence $\Gamma \in \mathcal{A}(?t, \neg t{:}F)$.

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., $\Delta \in \mathcal{A}(t, F)$. By Stability, $\Gamma \in \mathcal{A}(t, F)$ hence $t{:}F \in \Gamma$ and $F \in \Gamma^\sharp$. By the definition of R, $F \in \Delta'$.

4. $\mathsf{JD45}$. Similar to $\mathsf{J45}$.

5. $\mathsf{JT45}$. For soundness it suffices to check the Factivity Axiom, which easily follows from the reflexivity of R. For completeness, follow the footsteps 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; 41; 45; 46]), and epistemic semantics and completeness [21; 22] were first established for LP.

A fair amount of work has already been done on Jusification Logics other then LP. Systems J, $\mathsf{J4}$, and JT were first considered in [15] under different names and in a slightly different setting[9]. $\mathsf{JT45}$

[9] [15] also considered variants of Justification Logic systems which in our notations would be called "JD" and "JD4."

appeared independently in [50] and [52; 53], and JD45 in [50]. J45 has, perhaps, first been considered in this paper. Systems combining epistemic modalities and justifications were studied in [5; 11; 12].

Mkrtychev semantics for J, JT, and J4 with Completeness Theorem were found in [41]. Complexity bounds for LP and J4 were found [41; 45].

9 Forgetful Projection and the Correspondence Theorem

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

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

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** '$\rightsquigarrow$' that replaces each occurrence of $t{:}F$ by $\Box F$.

Example: the sentence

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

can be regarded as a Skolem-style version of

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

which can be read as

$$\Box P \to \Box Q,$$

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):

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

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

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

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 11. We assume that '$\rightsquigarrow$' is naturally extended from sentences to logics.

Theorem 11 [Consolidated Correspondence Theorem]

1. J $\rightsquigarrow$ K
2. JT $\rightsquigarrow$ T
3. J4 $\rightsquigarrow$ K4
4. LP $\rightsquigarrow$ S4
5. J45 $\rightsquigarrow$ K45
6. JD45 $\rightsquigarrow$ KD45
7. JT45 $\rightsquigarrow$ 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 11 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 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 proof of the Realization Theorem for 4 is given in [2; 4], cases 1–3 are covered in [15]. The Realization Theorem for 6 is established in [50], 7 is covered in [50; 52; 53], and the proof for 5 is very similar to [52; 53] and can be safely omitted here. $\Box$

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

9.1 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 for modal logic to that played by Kleene realizability for intuitionistic logic. In this respect, modal logic can be compared to intuitionistic logic. In both cases, the intended semantics was **existential**: the Brouwer-Heyting-Kolmogorov interpretation of intuitionistic logic ([33; 60; 61]) and Gödel's provability reading of S4 ([27; 28]). 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 [36; 58], the Logic of Proofs [2; 4], and other approaches to provide exact formalizations of computational and BHK semantics 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 off the scope of formal logical methods. The Correspondence Theorem tells us that justifications are compatible with the 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 [22; 50; 52; 53]. In principle, these results also produce realization procedures which are based on exhaustive search.

It would be a mistake to draw a 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*

$$\Box(\Box F \rightarrow F) \rightarrow \Box F, \tag{43}$$

which does not seem to have an epistemically acceptable explicit version. Let us consider, for example, a case when F is the propositional constant $\bot$ for *false*. A Skolem-style reading of (43) suggests that there are justification terms s and t such that

$$x{:}(s{:}\bot \rightarrow \bot) \rightarrow t{:}\bot. \tag{44}$$

This is intuitively false for factive justification though. Indeed, $s{:}\bot \rightarrow \bot$ is the Factivity Axiom. Apply Axiom Internalization R4 to obtain $c{:}[s{:}\bot \rightarrow \bot]$ for some constant c. This choice of c makes the antecedent of (44) intuitively true and the conclusion of (44) false[10]. In particular, (43) is not valid for proof interpretation (cf. [30] for a total account of which principles of GL are realizable).

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

10 Conclusions

1. Justification Logic has roots in mainstream epistemology, mathematical logic, and CS/AI. Justification Logic is capable of formalizing a significant portion of reasoning about justifications, both partial and factive. In particular, we have seen how to formalize Gettier examples in Justification Logic. This formalization has been used for verification, hidden assumption analysis, and eliminating redundancies.

2. Justification Logic provides justification-based foundations for epistemic modal logic. According to the Correspondence Theorem 11,

F is known

can be read as

there is a sufficient justification of F

for major systems of epistemic modal logic.

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

- intended provability semantics for Gödel's provability logic S4 with the Completeness Theorem ([2; 4]);
- formalization of Brouwer-Heyting-Kolmogorov semantics for intuitionistic propositional logic with the Completeness Theorem ([2; 4]);
- general definition of 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;
- 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; 43]). Justified Common Knowledge offers formal systems which are less restrictive than the usual epistemic logics with Common Knowledge [5].

4. 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.

11 Acknowledgements

The author is very grateful to Walter Dean, Mel Fitting, Vladimir Krupski, Roman Kuznets, Elena Nogina, and Tudor Protopopescu, 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, 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. Technical Report TR-2007016, CUNY Ph.D. Program in Computer Science, 2007.

[8] S. Artemov and L. Beklemishev. Provability logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, 2nd ed.*, volume 13, pages 229–403. Kluwer, Dordrecht, 2004.

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

[18] 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.

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

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

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

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

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

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

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

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

[27] 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.

[28] K. Gödel. Vortrag bei Zilsel, 1938. In S. Feferman, editor, *Kurt Gödel Collected Works. Volume III*, pages 86–113. Oxford University Press, 1995.

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

[30] 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.

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

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

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

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

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

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

[37] S. Kleene. *Introduction to Metamathematics*. Van Norstrand, 1952.

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

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

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

[41] 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.

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

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

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

[45] 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.

[46] 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.

[47] 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.

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

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

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

[51] 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.

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

[53] N. Rubtsova. On Realization of S5-modality by Evidence Terms. *Journal of Logic and Computation*, 16:671–684, 2006.

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

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

[56] T. Yavorskaya (Sidon). Multi-agent Explicit Knowledge. In J. Harrison D. Grigoriev 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.

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

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

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

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

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

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