PDF version of the entry
Justification Logic
https://plato.stanford.edu/archives/fall2024/entries/logic-justification/

from the Fall 2024 Edition of the

Stanford Encyclopedia of Philosophy

Library of Congress ISSN: 1095-5054

Justification Logic

First published Wed Jun 22, 2011; substantive revision Wed Jul 17, 2024

You may say, “I know that Abraham Lincoln was a tall man. ” In turn you may be asked how you know. You would almost certainly not reply semantically, Hintikka-style, that Abraham Lincoln was tall in all situations compatible with your knowledge. Instead you would more likely say, “I read about Abraham Lincoln’s height in several books, and I have seen photographs of him next to other people. ” One certifies knowledge by providing a reason, a justification. Hintikka semantics captures knowledge as true belief. Justification logics supply the missing third component of Plato’s characterization of knowledge as *justified* true belief.

---

1. Why Justification Logic?

Justification logics are epistemic logics which allow knowledge and belief modalities to be 'unfolded' into *justification terms*: instead of $\Box X$ one writes $t : X$, and reads it as "X is justified by reason t". One may think of traditional modal operators as *implicit* modalities, and justification terms as their *explicit* elaborations which supplement modal logics with finer-grained epistemic machinery. The family of justification terms has structure and operations. Choice of operations gives rise to different justification logics. For all common epistemic logics their modalities can be completely unfolded into explicit justification form. In this respect Justification Logic reveals and uses the explicit, but hidden, content of traditional Epistemic Modal Logic.

Justification logic originated as part of a successful project to provide a constructive semantics for intuitionistic logic—justification terms

abstracted away all but the most basic features of mathematical proofs. Proofs are justifications in perhaps their purest form. Subsequently justification logics were introduced into formal epistemology. This article presents the general range of justification logics as currently understood. It discusses their relationships with conventional modal logics. In addition to technical machinery, the article examines in what way the use of explicit justification terms sheds light on a number of traditional philosophical problems. The subject as a whole is still under active development.

The roots of justification logic can be traced back to many different sources, two of which are discussed in detail: epistemology and mathematical logic.

1.1 Epistemic Tradition

The properties of knowledge and belief have been a subject for formal logic at least since von Wright and Hintikka, (Hintikka 1962, von Wright 1951). Knowledge and belief are both treated as modalities in a way that is now very familiar—*Epistemic Logic*. But of Plato's three criteria for knowledge, *justified, true, belief*, (Gettier 1963, Hendricks 2005), epistemic logic really works with only two of them. Possible worlds and indistinguishability model belief—one believes what is so under all circumstances thought possible. Factivity brings a trueness component into play—if something is not so in the actual world it cannot be known, only believed. But there is no representation for the justification condition. Nonetheless, the modal approach has been remarkably successful in permitting the development of a rich mathematical theory and applications, (Fagin, Halpern, Moses, and Vardi 1995, van Ditmarsch, van der Hoek, and Kooi 2007). Still, it is not the whole picture.

The modal approach to the logic of knowledge is, in a sense, built around the universal quantifier: X is known in a situation if X is true in *all* situations indistinguishable from that one. Justifications, on the other hand, bring an existential quantifier into the picture: X is known in a situation if *there exists* a justification for X in that situation. This universal/existential dichotomy is a familiar one to logicians—in formal logics there exists a proof for a formula X if and only if X is true in all models for the logic. One thinks of models as inherently non-constructive, and proofs as constructive things. One will not go far wrong in thinking of justifications in general as much like mathematical proofs. Indeed, the first justification logic was explicitly designed to capture mathematical proofs in arithmetic, something which will be discussed further in Section 1.2.

In Justification Logic, in addition to the category of formulas, there is a second category of *justifications*. Justifications are formal terms, built up from constants and variables using various operation symbols. Constants represent justifications for commonly accepted truths—typically axioms. Variables denote unspecified justifications. Different justification logics differ on which operations are allowed (and also in other ways too). If t is a justification term and X is a formula, $t : X$ is a formula, and is intended to be read:

t is a justification for X.

One operation, common to all justification logics, is *application*, written like multiplication. The idea is, if s is a justification for $A \to B$ and t is a justification for A, then $[s \cdot t]$ is a justification for B[1]. That is, the validity of the following is generally assumed:

(1) $$s : (A \to B) \to (t : A \to [s \cdot t] : B).$$

This is the explicit version of the usual distributivity of knowledge operators, and modal operators generally, across implication:

(2) $$\Box(A \to B) \to (\Box A \to \Box B).$$

In fact, formula (2) is behind many of the problems of *logical omniscience*. It asserts that an agent knows everything that is implied by the agent's knowledge—knowledge is closed under consequence. While knowable-in-principle, knowability, is closed under consequence, the same cannot be said for any plausible version of actual knowledge. The distinction between (1) and (2) can be exploited in a discussion of the paradigmatic Red Barn Example of Goldman and Kripke; here is a simplified version of the story taken from (Dretske 2005).

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

In the first formalization of the Red Barn Example, logical derivation will be performed in a basic modal logic in which $\Box$ is interpreted as the 'belief' modality. Then some of the occurrences of $\Box$ will be externally interpreted as 'knowledge' according to the problem's description. Let B be the sentence 'the object in front of me is a barn', and let R be the sentence 'the object in front of me is red'.

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

At the metalevel, 2 is actually knowledge, whereas by the problem description, 1 is not knowledge.

3. $\Box(B \wedge R \to B)$, a knowledge assertion of a logical axiom.

Within this formalization, it appears that epistemic closure in its modal form (2) is violated: line 2, $\Box(B \wedge R)$, and line 3, $\Box(B \wedge R \to B)$ are cases of knowledge whereas $\Box B$ (line 1) is not knowledge. The modal language here does not seem to help resolving this issue.

Next consider the Red Barn Example in Justification Logic where $t : F$ is interpreted as 'I **believe** F for reason t'. Let u be a specific individual justification for belief that B, and v, for belief that $B \wedge R$. In addition, let a be a justification for the logical truth $B \wedge R \to B$. Then the list of assumptions is:

1. $u : B$, 'u is a reason to believe that the object in front of me is a barn';
2. $v : (B \wedge R)$, 'v is a reason to believe that the object in front of me is a red barn';
3. $a : (B \wedge R \to B)$.

On the metalevel, the problem description states that 2 and 3 are cases of knowledge, and not merely belief, whereas 1 is belief which is not knowledge. Here is how the formal reasoning goes:

4. $a : (B \wedge R \to B) \to (v : (B \wedge R) \to [a \cdot v] : B)$, by principle (1);
5. $v : (B \wedge R) \to [a \cdot v] : B$, from 3 and 4, by propositional logic;
6. $[a \cdot v] : B$, from 2 and 5, by propositional logic.

Notice that conclusion 6 is $[a \cdot v] : B$, and not $u : B$; epistemic closure holds. By reasoning in justification logic it was concluded that $[a \cdot v] : B$ is a case of knowledge, i.e., 'I know B for reason $a \cdot v$'. The fact that $u : B$ is not a case of knowledge does not spoil the closure principle, since the latter claims knowledge specifically for $[a \cdot v] : B$. Hence after observing a red façade, I indeed know B, but this knowledge has nothing to do with 1, which remains a case of belief rather than of knowledge. The justification logic formalization represents the situation fairly.

Tracking justifications represents the structure of the Red Barn Example in a way that is not captured by traditional epistemic modal tools. The Justification Logic formalization models what seems to be happening in such a case; closure of knowledge under logical entailment is maintained even though 'barn' is not perceptually known.[2]

1.2 Mathematical Logic Tradition

According to Brouwer, truth in constructive (intuitionistic) mathematics means the existence of a proof, cf. (Troelstra and van Dalen 1988). In 1931–34, Heyting and Kolmogorov gave an informal description of the intended proof-based semantics for intuitionistic logic (Kolmogorov 1932, Heyting 1934), which is now referred to as the *Brouwer-Heyting-Kolmogorov (BHK) semantics*. According to the BHK conditions, a formula is 'true' if it has a proof. Furthermore, a proof of a compound statement is connected to proofs of its components in the following way:

- a proof of $A \wedge B$ consists of a proof of proposition A and a proof of proposition B;
- a proof of $A \vee B$ is given by presenting either a proof of A or a proof of B;
- a proof of $A \to B$ is a construction transforming proofs of A into proofs of B;
- falsehood $\bot$ is a proposition which has no proof, $\neg A$ is shorthand for $A \to \bot$.

Kolmogorov explicitly suggested that the proof-like objects in his interpretation ("problem solutions") came from classical mathematics (Kolmogorov 1932). Indeed, from a foundational point of view it does not make much sense to understand the 'proofs' above as proofs in an intuitionistic system which these conditions are supposed to be specifying.

The fundamental value of the BHK semantics is that informally but unambiguously it suggests treating justifications, here mathematical proofs, as objects with operations.

In (Gödel 1933), Gödel took the first step towards developing a rigorous proof-based semantics for intuitionism. Gödel considered the classical modal logic $\mathsf{S4}$ to be a calculus describing properties of provability:

- Axioms and rules of classical propositional logic;
- $\Box(F \to G) \to (\Box F \to \Box G)$;
- $\Box F \to F$;
- $\Box F \to \Box\Box F$;
- Rule of necessitation: if $\vdash F$, then $\vdash \Box F$.

Based on Brouwer's understanding of logical truth as provability, Gödel defined a translation $\mathrm{tr}(F)$ of the propositional formula F in the intuitionistic language into the language of classical modal logic: $\mathrm{tr}(F)$ is obtained by prefixing every subformula of F with the provability modality $\Box$. Informally speaking, when the usual procedure of determining classical truth of a formula is applied to $\mathrm{tr}(F)$, it will test the provability (not the truth) of each of F's subformulas, in agreement with Brouwer's ideas. From Gödel's results and the McKinsey-Tarski work on topological semantics for modal logic, it follows that the translation $\mathrm{tr}(F)$ provides a proper embedding of the Intuitionistic Propositional Calculus, IPC, into $\mathsf{S4}$, i.e., an embedding of intuitionistic logic into classical logic extended by the provability operator.

(3) $$\text{If } \mathsf{IPC} \text{ proves } F, \text{ then } \mathsf{S4} \text{ proves } \mathrm{tr}(F).$$

Still, Gödel's original goal of defining intuitionistic logic in terms of classical provability was not reached, since the connection of $\mathsf{S4}$ to the usual mathematical notion of provability was not established. Moreover, Gödel noted that the straightforward idea of interpreting modality $\Box F$ as *F is provable in a given formal system T* contradicted Gödel's second incompleteness theorem. Indeed, $\Box(\Box F \to F)$ can be derived in $\mathsf{S4}$ by the rule of necessitation from the axiom $\Box F \to F$. On the other hand, interpreting modality $\Box$ as the predicate of formal provability in theory T and F as contradiction, converts this formula into a false statement that the consistency of T is internally provable in T.

The situation after (Gödel 1933) can be described by the following figure where '$X \hookrightarrow Y$' should be read as 'X is interpreted in Y'

$$\mathsf{IPC} \hookrightarrow \mathsf{S4} \hookrightarrow ? \hookrightarrow \mathit{CLASSICAL\ PROOFS}$$

In a public lecture in Vienna in 1938, Gödel observed that using the format of explicit proofs:

(4) $$t \text{ is a proof of } F.$$

can help in interpreting his provability calculus $\mathsf{S4}$ (Gödel 1938). Unfortunately, Gödel's work (Gödel 1938) remained unpublished until 1995, by which time the Gödelian logic of explicit proofs had already been rediscovered, and axiomatized as the Logic of Proofs LP and supplied with completeness theorems connecting it to both $\mathsf{S4}$ and classical proofs (Artemov 1995).

The Logic of Proofs LP became the first in the Justification Logic family. Proof terms in LP are nothing but BHK terms understood as classical proofs. With LP, propositional intuitionistic logic received the desired rigorous BHK semantics:

$$\mathsf{IPC} \hookrightarrow \mathsf{S4} \hookrightarrow \mathsf{LP} \hookrightarrow \mathit{CLASSICAL\ PROOFS}$$

For further discussion of the mathematical logic tradition, see the Section 1 of the supplementary document Some More Technical Matters.

1.3 Hyperintensionality

The *hyperintensional paradox* was formulated by Cresswell in 1975.

> It is well known that it seems possible to have a situation in which there are two propositions p and q which are logically equivalent and yet are such that a person may believe the one but not the other. If we regard a proposition as a set of possible worlds then two logically equivalent propositions will be identical, and so if 'x believes that' is a genuine sentential functor, the situation described in the opening sentence could not arise. I call this the paradox of hyperintensional contexts. Hyperintensional contexts are simply contexts which do not respect logical equivalence.

Starting with Cresswell himself, several ways of dealing with this have been proposed. Generally these involve adding more layers to familiar possible world approaches so that some way of distinguishing between logically equivalent sentences is available. Cresswell suggested that the syntactic form of sentences be taken into account. Justification Logic, in effect, takes sentence form into account through its mechanism for handling justifications for sentences. Thus Justification Logic addresses some of the central issues of hyperintensionality and, as a bonus, we automatically have an appropriate proof theory, model theory, complexity estimates and a broad variety of applications.

A good example of a hyperintensional context is the informal language used by mathematicians conversing with each other. Typically when a mathematician says he or she knows something, the understanding is that a proof is at hand. But as the following illustrates, this kind of knowledge is essentially hyperintensional.

> Fermat's Last Theorem, FLT, is logically equivalent to $0 = 0$ since both are provable, and hence denote the same proposition. However, the context of proofs distinguishes them immediately: a proof t of $0 = 0$ is not necessarily a proof of FLT, and vice versa.

To formalize mathematical speech the justification logic LP is a natural choice since $t{:}X$ was designed to have characteristics of "t *is a proof of* X."

The fact that propositions X and Y are equivalent in LP, $X \leftrightarrow Y$, does not warrant the equivalence of the corresponding justification assertions and typically $t{:}X$ and $t{:}Y$ are not equivalent, $t{:}X \not\leftrightarrow t{:}Y$.

Going further LP, and Justification Logic in general, is not only sufficiently refined to distinguish justification assertions for logically equivalent sentences, it provides a flexible machinery to connect justifications of equivalent sentences and hence to maintain constructive closure properties necessary for a quality logic system. For example, let X and Y be provably equivalent, i.e., there is a proof u of $X \leftrightarrow Y$, and so $u{:}(X \leftrightarrow Y)$ is provable in LP. Suppose also that v is a proof of X, and so $v{:}X$. It has already been mentioned that this does not mean v is a proof of Y—this is a hyperintensional context. However within the framework of Justification Logic, building on the proofs of X and of $X \leftrightarrow Y$, we can *construct* a proof term $f(u, v)$ which represents the proof of Y and so $f(u, v){:}Y$ is provable. In this respect, Justification Logic goes beyond Cresswell's expectations: logically equivalent sentences display different but constructively controlled epistemic behavior.

2. The Basic Components of Justification Logic

In this section the syntax and axiomatics of the most common systems of justification logic are presented.

2.1 The Language of Justification Logic

In order to build a formal account of justification logics one must make a basic structural assumption: *justifications are abstract objects which have structure and operations on them*. A good example of justifications is provided by formal proofs, which have long been objects of study in mathematical logic and computer science (cf. Section 1.2).

Justification Logic is a formal logical framework which incorporates epistemic assertions $t : F$, standing for 't is a justification for F'. Justification Logic does not directly analyze what it means for t to justify F beyond the format $t : F$, but rather attempts to characterize this relation axiomatically. This is similar to the way Boolean logic treats its connectives, say, disjunction: it does not analyze the formula $p \vee q$ but rather assumes certain logical axioms and truth tables about this formula.

There are several design decisions made. Justification Logic starts with the simplest base: *classical Boolean logic*, and for good reasons. Justifications provide a sufficiently serious challenge on even the simplest level. The paradigmatic examples by Russell, Goldman-Kripke, Gettier and others, can be handled with Boolean Justification Logic. The core of Epistemic Logic consists of modal systems with a classical Boolean base (K, T, K4, S4, K45, KD45, S5, etc.), and each of them has been provided with a corresponding Justification Logic companion based on Boolean logic. Finally, factivity of justifications is not always assumed. This makes it possible to capture the essence of discussions in epistemology involving matters of belief and not knowledge.

The basic operation on justifications is *application*. The *application* operation takes justifications s and t and produces a justification $s \cdot t$ such that if $s : (F \to G)$ and $t : F$, then $[s \cdot t] : G$. Symbolically,

$$s : (F \to G) \to (t : F \to [s \cdot t] : G)$$

This is a basic property of justifications assumed in combinatory logic and λ-calculi (Troelstra and Schwichtenberg 1996), Brouwer-Heyting-Kolmogorov semantics (Troelstra and van Dalen 1988), Kleene realizability (Kleene 1945), the Logic of Proofs LP, etc.

Another common operation on justifications is sum: it has been introduced to make explicit the modal logic reasoning (Artemov 1995). However, some meaningful justification logics like J^- (Artemov and Fitting 2019) or JNoC^- (Faroldi, Ghari, Lehmann, and Studer 2024) do not use the operation sum. With sum, any two justifications can safely be combined into something with broader scope. If $s : F$, then whatever evidence t may be, the combined evidence $s + t$ remains a justification for F. More properly, the operation '+' takes justifications s and t and produces $s + t$, which is a justification for everything justified by s or by t.

$$s : F \to [s + t] : F \text{ and } t : F \to [s + t] : F$$

As motivation, one might think of s and t as two volumes of an encyclopedia, and $s + t$ as the set of those two volumes. Imagine that one of the volumes, say s, contains a sufficient justification for a proposition F, i.e., $s : F$ is the case. Then the larger set $s + t$ also contains a sufficient justification for F, $[s + t] : F$. In the Logic of Proofs LP, Section 1.2, '$s + t$' can be interpreted as a concatenation of proofs s and t.

2.2 Basic Justification Logic J_0

Justification terms are built from justification variables $x, y, z, \ldots$ and justification constants $a, b, c, \ldots$ (with indices $i = 1, 2, 3, \ldots$ which are omitted whenever it is safe) by means of the operations '$\cdot$ ' and '+'. More elaborate logics considered below also allow additional operations on justifications. Constants denote atomic justifications which the system does not analyze; variables denote unspecified justifications. The Basic Logic of Justifications, J_0 is axiomatized by the following.

Classical Logic

Classical propositional axioms and the rule Modus Ponens

Application Axiom

$s : (F \to G) \to (t : F \to [s \cdot t] : G),$

Sum Axioms

$s : F \to [s + t] : F, \quad s : F \to [t + s] : F.$

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

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

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

2.3 Logical Awareness and Constant Specifications

The *Logical Awareness principle* states that logical axioms are justified *ex officio*: an agent accepts logical axioms as justified (including the ones concerning justifications). As just stated, Logical Awareness may be too strong in some epistemic situations. However Justification Logic offers the flexible mechanism of Constant Specifications to represent varying shades of Logical Awareness.

Of course one distinguishes between an assumption and a justified assumption. In Justification Logic constants are used to represent justifications of assumptions in situations where they are not analyzed any further. Suppose it is desired to postulate that an axiom A is justified for the knower. One simply postulates $e_1 : A$ for some evidence constant e_1 (with index 1). If, furthermore, it is desired to postulate that this new principle $e_1 : A$ is also justified, one can postulate $e_2 : (e_1 : A)$ for a

constant e_2 (with index 2). And so on. Keeping track of indices is not necessary, but it is easy and helps in decision procedures (Kuznets 2008). The set of all assumptions of this kind for a given logic is called a *Constant Specification*. Here is the formal definition:

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

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

where A is an axiom of $\mathcal{L}$, and $e_1, e_2, \ldots, e_n$ are similar constants with indices 1, 2, …, n. It is assumed that CS contains all intermediate specifications, i.e., whenever $e_n : e_{n-1} : \ldots : e_1 : A$ is in CS, then $e_{n-1} : \ldots : e_1 : A$ is in CS, too.

There are a number of special conditions that have been placed on constant specifications in the literature. The following are the most common.

Empty

$CS = \varnothing$. This corresponds to an absolutely skeptical agent. It amounts to working with the logic J_0.

Finite

CS is a finite set of formulas. This is a fully representative case, since any specific derivation in Justification Logic will involve only a finite set of constants.

Axiomatically Appropriate

Each axiom, including those newly acquired through the constant specification itself, have justifications. In the formal setting, for each axiom A there is a constant e_1 such that $e_1 : A$ is in CS, and if $e_n : e_{n-1} : \ldots : e_1 : A \in CS$, then $e_{n+1} : e_n : e_{n-1} : \ldots : e_1 : A \in CS$, for each $n \geq 1$. Axiomatically appropriate constant specifications are necessary for ensuring the Internalization property, discussed at the end of this section.

Total

For each axiom A and any constants $e_1, e_2, \ldots, e_n$,

$$e_n : e_{n-1} : \ldots : e_1 : A \in CS.$$

The name *TCS* is reserved for the total constant specification (for a given logic). Naturally, the total constant specification is axiomatically appropriate.

We may now specify:

Logic of Justifications with given Constant Specification:
Let CS be a constant specification. J_{CS} is the logic J_0 + CS ; the axioms are those of J_0 together with the members of CS, and the only rule of inference is *Modus Ponens*. Note that J_0 is $\mathsf{J}_\varnothing$.

Logic of Justifications:
J is the logic J_0 + **Axiom Internalization Rule**. The new rule states:

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

The latter embodies the idea of unrestricted Logical Awareness for J. A similar rule appeared in the Logic of Proofs LP, and has also been anticipated in Goldman's (Goldman 1967). Logical Awareness, as expressed by axiomatically appropriate Constant Specifications, is an explicit incarnation of the Necessitation Rule in Modal Logic: $\vdash F \Rightarrow \vdash \Box F$, but restricted to axioms. Note that J coincides with J_{TCS}.

The key feature of Justification Logic systems is their ability to internalize their own derivations as provable justification assertions within their languages. This property was anticipated in (Gödel 1938).

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

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

Proof. Induction on derivation length. Suppose $\vdash F$. If F is a member of J_0, or a member of CS, there is a constant e_n (where n might be 1) such that $e_n : F$ is in CS, since CS is axiomatically appropriate. Then $e_n : F$ is derivable. If F is obtained by *Modus Ponens* from $X \to F$ and X, then, by the Induction Hypothesis, $\vdash s : (X \to F)$ and $\vdash t : X$ for some s, t. Using the Application Axiom, $\vdash [s \cdot t] : F$.

See Section 2 of the supplementary document Some More Technical Matters for examples of concrete syntactic derivations in justification logic.

2.4 Extending Basic Justification Logic

The basic justification logic J_0, and its extension with a constant specification J_{CS}, is an explicit counterpart of the smallest normal modal logic K. A proper definition of counterpart will be given in Section 4 because the notion of *realization* is central, but some hints are already apparent at this stage of our presentation. For instance, it was noted in Section 1.1 that (1), $s{:}(A \to B) \to (t{:}A \to [s \cdot t]{:}B)$, is an explicit version of the familiar modal principle (2), $\Box(A \to B) \to (\Box A \to \Box B)$. In a similar way the first justification logic LP is an explicit counterpart of modal $\mathsf{S4}$. It turns out that many modal logics have justification logic counterparts—indeed, generally more than one. In what follows we begin by discussing some very familiar logics, leading up to $\mathsf{S4}$ and LP. Up to this point much of our original motivation applies—we have justification logics that are interpretable in arithmetic. Then we move on to a broader family of modal logics, and the arithmetic motivation is no longer applicable. The phenomenon of having a modal logic with a justification logic counterpart has turned out to be unexpectedly broad.

In almost all cases, one must add operations to the $+$ and $\cdot$ of J_0, along with axioms capturing their intended behavior. The exception is factivity, discussed next, for which no additional operations are required, though additional axioms are. It is always understood that constant specifications cover axioms from the enlarged set. We continue using the terminology of Section 2.3; for instance a constant specification is axiomatically appropriate if it meets the condition as stated there, *for all axioms including any that have been added to the original set*. Theorem 1 from Section 2.3 continues to apply to our new justification logics, and with the same proof: if we have a justification logic JL_{CS} with an axiomatically appropriate constant specification, Internalization holds.

2.5 Factivity

Factivity states that justifications are sufficient for an agent to conclude truth. This is embodied in the following.

Factivity Axiom $t : F \to F$.

The Factivity Axiom has a similar motivation to the Truth Axiom of Epistemic Logic, $\Box F \to F$, which is widely accepted as a basic property of knowledge.

Factivity of justifications is not required in basic Justification Logic systems, which makes them capable of representing both partial and factive justifications. The Factivity Axiom appeared in the Logic of Proofs LP, Section 1.2, as a principal feature of mathematical proofs. Indeed, in this setting Factivity is clearly valid: if there is a mathematical proof t of F, then F must be true.

The Factivity Axiom is adopted for justifications that lead to knowledge. However, factivity alone does not warrant knowledge, as has been demonstrated by the Gettier examples (Gettier 1963).

Logic of Factive Justifications:

- $\mathsf{JT}_0 = \mathsf{J}_0$ + Factivity;
- $\mathsf{JT} = \mathsf{J}$ + Factivity.

Systems JT_{CS} corresponding to Constant Specifications CS are defined as in Section 2.3.

2.6 Positive Introspection

One of the common principles of knowledge is identifying *knowing* and *knowing that one knows*. In a modal setting, this corresponds to $\Box F \to \Box\Box F$. This principle has an adequate explicit counterpart: the fact that an agent accepts t as sufficient evidence for F serves as sufficient evidence for $t : F$. Often such 'meta-evidence' has a physical form: a referee report certifying that a proof in 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. A *Positive Introspection* operation '!' may be added to the language for this purpose; one then assumes that given t, the agent produces a justification $!t$ of $t : F$ such that $t : F \to !t : (t : F)$. Positive Introspection in this operational form first appeared in the Logic of Proofs LP.

Positive Introspection Axiom: $t : F \to !t : (t : F)$.

We then define:

- $\mathsf{J4} := \mathsf{J}$ + Positive Introspection;
- $\mathsf{LP} := \mathsf{JT}$ + Positive Introspection.[3]

Logics $\mathsf{J4}_0, \mathsf{J4}_{CS}, \mathsf{LP}_0$, and LP_{CS} are defined in the natural way (cf. Section 2.3).

In the presence of the Positive Introspection Axiom, one can limit the scope of the Axiom Internalization Rule to internalizing axioms which are not of the form $e : A$. This is how it was done in LP: Axiom Internalization can then be emulated by using $!!e : (!e : (e : A))$ instead of $e_3 : (e_2 : (e_1 : A))$, etc. The notion of Constant Specification can also be simplified accordingly. Such modifications are minor and they do not affect the main theorems and applications of Justification Logic.

2.7 Negative Introspection

(Pacuit 2006, Rubtsova 2006) considered the *Negative Introspection* operation '?' which verifies that a given justification assertion is false. A possible motivation for considering such an operation is that the positive introspection operation '!' may well be regarded as capable of providing *conclusive* verification judgments about the validity of justification assertions $t : F$, so when t is not a justification for F, such a '!' should conclude that $\neg t : F$. This is normally the case for computer proof verifiers, proof checkers in formal theories, etc. This motivation is, however, nuanced: the examples of proof verifiers and proof checkers work with both t and F as inputs, whereas the Pacuit-Rubtsova format $?t$ suggests that the only input for '?' is a justification t, and the result $?t$ is supposed to justify propositions $\neg t : F$ uniformly for all Fs for which $t : F$ does not hold. Such an operation '?' does not exist for formal mathematical proofs since $?t$ should then be a single proof of infinitely many propositions $\neg t : F$, which is impossible. The operation '?' was, historically, the first example that did not fit into the original framework in which justifications were abstract versions of formal proofs.

Negative Introspection Axiom $\neg t : F \to ?t : (\neg t : F)$

We define the systems:

- $\mathsf{J45} = \mathsf{J4}$ + Negative Introspection;
- $\mathsf{JD45} = \mathsf{J45} + \neg t : \perp$;
- $\mathsf{JT45} = \mathsf{J45}$ + Factivity

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

2.8 Geach Logics and More

Justification logics involving ? were the first examples that went beyond sublogics of LP. More recently it has been discovered that there is an *infinite* family of modal logics that have justification counterparts, but for which the connection with arithmetic proofs is weak or missing. We discuss a single case in some detail, and sketch others.

Peter Geach proposed the axiom scheme $\Diamond\Box X \to \Box\Diamond X$. When added to axiomatic $\mathsf{S4}$ it yields an interesting logic known as $\mathsf{S4.2}$. Semantically, Geach's scheme imposes *confluence* on frames. That is, if two possible worlds, w_1 and w_2 are accessible from the same world w_0, there is a common world w_4 accessible from both w_1 and w_2. Geach's scheme was generalized in Lemmon and Scott (1977) and a corresponding notation was introduced: $\mathsf{G}^{k,l,m,n}$ is the scheme $\Diamond^k\Box^l X \to \Box^m\Diamond^n X$, where $k, l, m, n \geq 0$. Semantically these schemes correspond to generalized versions of confluence. Some people have begun referring to the schemes as *Geach schemes*, and we will follow this practice. More generally, we will call a modal logic a *Geach* logic if it can be axiomatized by adding a finite set of Geach schemes to K. The original Geach scheme is $\mathsf{G}^{1,1,1,1}$, but also note that $\Box X \to X$ is $\mathsf{G}^{0,1,0,0}$, $\Box X \to \Box\Box X$ is $\mathsf{G}^{0,1,2,0}$, $\Diamond X \to \Box\Diamond X$ is $\mathsf{G}^{1,0,1,1}$, and $X \to \Box\Diamond X$ is $\mathsf{G}^{0,0,1,1}$, so Geach logics include the most common of the modal logics. Geach logics constitute an infinite family.

Every Geach logic has a justification counterpart. Consider the original Geach logic, with axiom scheme $\mathsf{G}^{1,1,1,1}$, $\Diamond\Box X\rightarrow\Box\Diamond X$ added to a system for $\mathsf{S4}$—the system $\mathsf{S4.2}$ mentioned above. We build a justification counterpart for $\mathsf{S4.2}$ axiomatically by starting with LP. Then we add two function symbols, f and g, each two-place, and adopt the following axiom scheme, calling the resulting justification logic $\mathsf{J4.2}$.

$$\neg f(t,u){:}\neg t{:}X\rightarrow g(t,u){:}\neg u{:}\neg X$$

There is some informal motivation for this scheme. In LP, because of the axiom scheme $t{:}X\rightarrow X$, we have provability of $(t{:}X\wedge u{:}\neg X)\rightarrow\bot$ for any t and u, and thus provability of $\neg t{:}X\vee\neg u{:}\neg X$. In any context one of the disjuncts must hold. The scheme above is equivalent to $f(t,u){:}\neg t{:}X\vee g(t,u){:}\neg u{:}\neg X$, which informally says that in any context we have means for computing a justification for the disjunct that holds. It is a strong assumption, but not implausible at least in some circumstances.

A realization theorem connects $\mathsf{S4.2}$ and $\mathsf{J4.2}$, though it is not known if this has a constructive proof.

As another example, consider $\mathsf{G}^{1,2,2,1}$, $\Diamond\Box\Box X\rightarrow\Box\Box\Diamond X$, or equivalently $\Box\neg\Box\Box X\vee\Box\Box\neg\Box X$. It has as a corresponding justification axiom scheme the following, where f, g, and h are three-place function symbols.

$$f(t,u,v){:}\neg t{:}u{:}X\vee g(t,u,v){:}h(t,u,v){:}\neg v{:}\neg X$$

An intuitive interpretation for f, g, and h is not as clear as it is for $\mathsf{G}^{1,1,1,1}$, but formally things behave quite well.

Even though the Geach family is infinite, these logics do not cover the full range of logics with justification counterparts. For instance, the normal modal logic using the axiom scheme $\Box(\Box X\rightarrow X)$, sometimes called *shift reflexivity*, is not a Geach logic, but it does have a justification counterpart. Add a one-place function symbol k to the machinery building up

justification terms, and adopt the justification axiom scheme $k(t){:}(t{:}X\rightarrow X)$. A Realization Theorem holds; this is shown in Fitting (2014b). We speculate that all logics axiomatized with Sahlquist formulas will have justification counterparts, but this remains a conjecture at this point.

3. Semantics

The now-standard semantics for justification logic originates in (Fitting 2005)—the models used are generally called *Fitting models* in the literature, but will be called *possible world justification models* here. Possible world justification models are an amalgam of the familiar possible world semantics for logics of knowledge and belief, due to Hintikka and Kripke, with machinery specific to justification terms, introduced by Mkrtychev in (Mkrtychev 1997), (cf. Section 3.4).

3.1 Single-Agent Possible World Justification Models for J

To be precise, a semantics for J_{CS}, where CS is any constant specification, is to be defined. Formally, a *possible world justification logic model* for J_{CS} is a structure $\mathcal{M}=\langle\mathcal{G},\mathcal{R},\mathcal{E},\mathcal{V}\rangle$. Of this, $\langle\mathcal{G},\mathcal{R}\rangle$ is a standard K frame, where $\mathcal{G}$ is a set of possible worlds and $\mathcal{R}$ is a binary relation on it. $\mathcal{V}$ is a mapping from propositional variables to subsets of $\mathcal{G}$, specifying atomic truth at possible worlds.

The new item is $\mathcal{E}$, an *evidence function*, which originated in (Mkrtychev 1997). This maps justification terms and formulas to sets of worlds. The intuitive idea is, if the possible world Γ is in $\mathcal{E}(t,X)$, then t is *relevant* or *admissible* evidence for X at world Γ . One should not think of relevant evidence as conclusive. Rather, think of it as more like evidence that can be admitted in a court of law: this testimony, this document is something a jury should examine, something that is pertinent, but something whose truth-determining status is yet to be considered. Evidence functions must meet certain conditions, but these are discussed a bit later.

Given a J_{CS} possible world justification model $\mathcal{M} = \langle \mathcal{G}, \mathcal{R}, \mathcal{E}, \mathcal{V} \rangle$, truth of formula X at possible world Γ is denoted by $\mathcal{M}, \Gamma \Vdash X$, and is required to meet the following standard conditions:

For each $\Gamma \in \mathcal{G}$:

1. $\mathcal{M}, \Gamma \Vdash P$ iff $\Gamma \in \mathcal{V}(P)$ for P a propositional letter;
2. it is not the case that $\mathcal{M}, \Gamma \Vdash \bot$;
3. $\mathcal{M}, \Gamma \Vdash X \rightarrow Y$ iff it is not the case that $\mathcal{M}, \Gamma \Vdash X$ or $\mathcal{M}, \Gamma \Vdash Y$.

These just say that atomic truth is specified arbitrarily, and propositional connectives behave truth-functionally at each world. The key item is the next one.

1. $\mathcal{M}, \Gamma \Vdash (t : X)$ if and only if $\Gamma \in \mathcal{E}(t, X)$ and, for every $\Delta \in \mathcal{G}$ with $\Gamma \mathcal{R} \Delta$, we have that $\mathcal{M}, \Delta \Vdash X$.

This condition breaks into two parts. The clause requiring that $\mathcal{M}, \Delta \Vdash X$ for every $\Delta \in \mathcal{G}$ such that $\Gamma \mathcal{R} \Delta$ is the familiar Hintikka/Kripke condition for X to be believed, or be believable, at Γ. The clause requiring that $\Gamma \in \mathcal{E}(t, X)$ adds that t should be relevant evidence for X at Γ. Then, informally, $t : X$ is true at a possible world if X is believable at that world in the usual sense of epistemic logic, and t is relevant evidence for X at that world.

It is important to realize that, in this semantics, one might not believe something for a particular reason at a world either because it is simply not believable, or because it is but the reason is not appropriate.

Some conditions must still be placed on evidence functions, and the constant specification must also be brought into the picture. Suppose one is given s and t as justifications. One can combine these in two different ways: simultaneously use the information from both; or use the information from just one of them, but first choose which one. Each gives rise to a basic operation on justification terms, $\cdot$ and $+$, introduced axiomatically in Section 2.2.

Suppose s is relevant evidence for an implication and t is relevant evidence for the antecedent. Then s and t together provides relevant evidence for the consequent. The following condition on evidence functions is assumed:

$$\mathcal{E}(s, X \rightarrow Y) \cap \mathcal{E}(t, X) \subseteq \mathcal{E}(s \cdot t, Y)$$

With this condition added, the validity of

$$s : (X \rightarrow Y) \rightarrow (t : X \rightarrow [s \cdot t] : Y)$$

is secured.

If s and t are items of evidence, one might say that something is justified by one of s or t, without bothering to specify which, and this will still be evidence. The following requirement is imposed on evidence functions.

$$\mathcal{E}(s, X) \cup \mathcal{E}(t, X) \subseteq \mathcal{E}(s + t, X)$$

Not surprisingly, both

$$s : X \rightarrow [s + t] : X$$

and

$$t : X \rightarrow [s + t] : X$$

now hold.

Finally, the Constant Specification CS should be taken into account. Recall that constants are intended to represent reasons for basic assumptions that are accepted outright. A model $\mathcal{M} = \langle \mathcal{G}, \mathcal{R}, \mathcal{E}, \mathcal{V} \rangle$ *meets* Constant Specification CS provided: if $c : X \in CS$ then $\mathcal{E}(c,X) = \mathcal{G}$.

> **Possible World Justification Model** A possible world justification model for J_{CS} is a structure $\mathcal{M} = \langle \mathcal{G}, \mathcal{R}, \mathcal{E}, \mathcal{V} \rangle$ satisfying all the conditions listed above, and meeting Constant Specification CS.

Despite their similarities, possible world justification models allow a fine-grained analysis that is not possible with Kripke models. See Section 3 of the supplementary document Some More Technical Matters for more details.

3.2 Weak and Strong Completeness

A formula X is *valid* in a particular model for J_{CS} if it is true at all possible worlds of the model. Axiomatics for J_{CS} was given in Sections 2.2 and 2.3. A completeness theorem now takes the expected form.

> **Theorem 2**: A formula X is provable in J_{CS} if and only if X is valid in all J_{CS} models.

The completeness theorem as just stated is sometimes referred to as *weak* completeness. It maybe a bit surprising that it is significantly easier to prove than completeness for the modal logic K. Comments on this point follow. On the other hand it is very general, working for all Constant Specifications.

In (Fitting 2005) a stronger version of the semantics was also introduced. A model $\mathcal{M} = \langle \mathcal{G}, \mathcal{R}, \mathcal{E}, \mathcal{V} \rangle$ is called *fully explanatory* if it meets the following condition. For each $\Gamma \in \mathcal{G}$, if $\mathcal{M}, \Delta \Vdash X$ for all $\Delta \in \mathcal{G}$ such that $\Gamma \mathcal{R} \Delta$, then $\mathcal{M}, \Gamma \Vdash t : X$ for some justification term t. Note that the condition, $\mathcal{M}, \Delta \Vdash X$ for all $\Delta \in \mathcal{G}$ such that $\Gamma \mathcal{R} \Delta$, is the usual condition for X being believable at Γ in the Hintikka/Kripke sense. So, fully explanatory really says that if a formula is believable at a possible world, there is a justification for it.

Not all weak models meet the fully explanatory condition. Models that do are called *strong* models. If constant specification CS is rich enough so that an Internalization theorem holds, then one has completeness with respect to strong models meeting CS. Indeed, in an appropriate sense completeness with respect to strong models is equivalent to being able to prove Internalization.

The proof of completeness with respect to strong models bears a close similarity to the proof of completeness using canonical models for the modal logic K. In turn, strong models can be used to give a semantic proof of the Realization Theorem (cf. Section 4).

3.3 The Single-Agent Family

So far a possible world semantics for one justification logic has been discussed, for J, the counterpart of K. Now things are broadened to encompass justification analogs of other familiar modal logics.

Simply by adding reflexivity of the accessibility relation $\mathcal{R}$ to the conditions for a model in Section 3.1, one gains the validity of $t{:}X \to X$ for every t and X, and obtains a semantics for JT, the justification logic analog of the modal logic T, the weakest logic of knowledge. Indeed, if $\mathcal{M}, \Gamma \Vdash t{:}X$ then, in particular, X is true at every state accessible from Γ. Since the accessibility relation is required to be reflexive, $\mathcal{M}, \Gamma \Vdash X$. Weak and strong completeness theorems are provable using the same machinery that applied in the case of J, and a semantic proof of a Realization

Theorem connecting JT and T is also available. The same applies to the logics discussed below.

For a justification analog of $\mathsf{K4}$ an additional unary operator '!' is added to the term language, see Section 2.5. Recall this operator maps justifications to justifications, where the idea is that if t is a justification for X, then $!t$ should be a justification for $t{:}X$. Semantically this adds conditions to a model $\mathcal{M} = \langle \mathcal{G}, \mathcal{R}, \mathcal{E}, \mathcal{V} \rangle$, as follows.

First, of course, $\mathcal{R}$ should be transitive, but not necessarily reflexive. Second, a monotonicity condition on evidence functions is required:

$$\text{If } \Gamma\mathcal{R}\Delta \text{ and } \Gamma \in \mathcal{E}(t, X) \text{ then } \Delta \in \mathcal{E}(t, X)$$

And finally, one more evidence function condition is needed.

$$\mathcal{E}(t, X) \subseteq \mathcal{E}(!t, t{:}X)$$

These conditions together entail the validity of $t{:}X \rightarrow !t{:}t{:}X$ and produce a semantics for $\mathsf{J4}$, a justification analog of $\mathsf{K4}$, with a Realization Theorem connecting them. Adding reflexivity leads to a logic that is called LP for historical reasons.

We have discussed justification logics that are sublogics of LP, corresponding to sublogics of the modal logic $\mathsf{S4}$. The first examples that went beyond LP were those discussed in Section 2.7, involving a negative introspection operator, '?'. Models for justification logics that include this operator add three conditions. First R is symmetric. Second, one adds a condition that has come to be known as *strong evidence*: $\mathcal{M}, \Gamma \Vdash t{:}X$ for all $\Gamma \in \mathcal{E}(t, X)$. Finally, there is a condition on the evidence function:

$$\overline{\mathcal{E}(t, X)} \subseteq \mathcal{E}(?t, \neg t{:}X)$$

If this machinery is added to that for $\mathsf{J4}$ we get the logic $\mathsf{J45}$, a justification counterpart of $\mathsf{K45}$. Axiomatic soundness and completeness

can be proved. In a similar way, related logics $\mathsf{JD45}$ and $\mathsf{JT45}$ can be formulated semantically. A Realization Theorem taking the operator ? into account was shown in (Rubtsova 2006).

Moving to Geach logics as introduced in Section 2.8, a semantic model for $\mathsf{J4.2}$ can also be specified. Suppose $G = \langle \mathcal{G}, \mathcal{R}, \mathcal{E}, \mathcal{V} \rangle$ is an LP model. We add the following requirements. First, the frame must be convergent, as with $\mathsf{S4.2}$. Second, as with ?, $\mathcal{E}$ must be a *strong* evidence function. And third, $\mathcal{E}(f(t, u), \neg t{:}X) \cup \mathcal{E}(g(t, u), \neg u{:}\neg X) = \mathcal{G}$. Completeness and soundness results follow in the usual way.

In a similar way every modal logic axiomatized by Geach schemes in this family has a justification counterpart, with a Fitting semantics and a realization theorem connecting the justification counterpart with the corresponding modal logic. In particular, this tells us that the justification logic family is infinite, and certainly much broader than it was originally thought to be. It is also the case that some modal logics not previously considered, and not in this family, have justification counterparts as well. Investigating the consequences of all this is still work in progress.

3.4 Single World Justification Models

Single world justification models were developed considerably before the more general possible world justification models we have been discussing, (Mkrtychev 1997). Today they can most simply be thought of as possible world justification models that happen to have a single world. The completeness proof for J and the other justification logics mentioned above can easily be modified to establish completeness with respect to single world justification models, though of course this was not the original argument. What completeness with respect to single world justification models tells us is that information about the possible world structure of justification models can be completely encoded by the admissible evidence function, at least for the logics discussed so far. Mkrtychev used single world justification models to establish decidability of LP, and others have made fundamental use of them in setting complexity bounds for justification logics, as well as for showing conservativity results for justification logics of belief (Kuznets 2000, Kuznets 2008, Milnikel 2007, Milnikel 2009). Complexity results have further been used to address the problem of logical omniscience (Artemov and Kuznets 2014).

3.5 Ontologically Transparent Semantics

The formal semantics for Justification Logic described above in 3.1–3.4 defines truth value at a given world Γ the same way it is done in Awareness Models: $t{:}F$ holds at Γ iff

1. F holds at all worlds accessible from Γ and
2. t is admissible evidence for F according to the given evidence function.

In addition, there is a different kind of semantics, so-called modular semantics, which focuses on making more transparent the ontological status of justifications. Within modular semantics propositions receive the usual classical truth values and justifications are interpreted syntactically as sets of formulas. We retain a classical interpretation $*$ of the propositional formulas Fm, which, in the case of a single world, reduces to

$$*: Fm \mapsto \ \{0,1\}$$

i.e., each formula gets a truth value 0 (false) or 1 (true), with the usual Boolean conditions: $\Vdash A \rightarrow B$ iff $\nVdash A$ or $\Vdash B$, etc. The principal issue is

how to interpret justification terms. For *sets of formulas* X and Y, we define

$$X \cdot Y = \{F \mid G \rightarrow F \in X \text{ and } G \in Y \text{ for some } G\}.$$

Informally, $X \cdot Y$ is the result of applying *Modus Ponens* once between all members of X and of Y (in that order). Justification terms *Tm* are interpreted as subsets of the set of formulas:

$$*: Tm \mapsto \ 2^{Fm}$$

such that

$$(s \cdot t)^* \supseteq s^* \cdot t^* \ \text{ and } \ (s+t)^* \supseteq s^* \cup t^*.$$

These conditions correspond to the basic justification logic J; other systems require additional closure properties of $*$. Note that whereas propositions in modular models are interpreted semantically, as truth values, justifications are interpreted syntactically, as sets of formulas. This is a principal hyperintensional feature: a modular model may treat distinct formulas F and G as equal in the sense that $F^* = G^*$, but still be able to distinguish justification assertions $t{:}F$ and $t{:}G$, for example when $F \in t^*$ but $G \notin t^*$ yielding $\Vdash t{:}F$ but $\nVdash t{:}G$. In the general possible world setting, formulas are interpreted classically as subsets of the set W of possible worlds,

$$*: Fm \mapsto \ 2^W,$$

and justification terms are interpreted syntactically as sets of formulas at each world

$$*: W \times Tm \mapsto \ 2^{Fm}.$$

Soundness and completeness of Justification Logic systems with respect to modular models have been demonstrated in (Artemov 2012; Kuznets and Studer 2012).

3.6 Connections with Awareness Models

The logical omniscence problem is that in epistemic logics all tautologies are known and knowledge is closed under consequence, which is unreasonable. In Fagin and Halpern (1988) a simple mechanism for avoiding the problems was introduced. One adds to the usual Kripke model structure an awareness function $\mathcal{A}$ indicating for each world which formulas the agent is aware of at this world. Then a formula is taken to be known at a possible world Γ if 1) the formula is true at all worlds accessible from Γ (the Kripkean condition for knowledge) and 2) the agent is aware of the formula at Γ. Awareness functions can serve as a practical tool for blocking knowledge of an arbitrary set of formulas. However as logical structures, awareness models can exhibit unusual behavior due to the lack of natural closure properties. For example, the agent can know $A \wedge B$ but be aware of neither A nor B and hence not know either.

Possible world justification logic models use a forcing definition reminiscent of the one from the awareness models: for any given justification t the justification assertion $t{:}F$ holds at world Γ iff 1) F holds at all worlds Δ accessible from Γ and 2) t is admissible evidence for F at Γ, $\Gamma \in \mathcal{E}(t, F)$. The principal difference is in the operations on justifications and corresponding closure conditions on admissible evidence function $\mathcal{E}$ in Justification Logic models, which may hence be regarded as a dynamic version of awareness models which necessary closure properties specified. This idea has been explored in Sedlár (2013) which worked with the language of LP, thinking of it as a multi-agent modal logic, and taking justification terms as agents (more properly, actions of agents). This shows that Justification Logic models absorb the usual epistemic themes of awareness, group agency and dynamics in a natural way.

4. Realization Theorems

The natural modal epistemic counterpart of the evidence assertion $t : F$ is $\Box F$, read *for some x, x*:F. This observation leads to the notion of *forgetful projection* which replaces each occurrence of $t : F$ by $\Box F$ and hence converts a Justification Logic sentence S to a corresponding Modal Logic sentence S^o. The forgetful projection extends in the natural way from sentences to logics.

Obviously, different Justification Logic sentences may have the same forgetful projection, hence S^o loses certain information that was contained in S. However, it is easily observed that the forgetful projection always maps valid formulas of Justification Logic (e.g., axioms of J) to valid formulas of a corresponding Epistemic Logic (K in this case). The converse also holds: any valid formula of Epistemic Logic is the forgetful projection of some valid formula of Justification Logic. This follows from the Correspondence Theorem 3.

> **Theorem 3**: $\mathsf{J}^o = \mathsf{K}$.

This correspondence holds for other pairs of Justification and Epistemic systems, for instance $\mathsf{J4}$ and $\mathsf{K4}$, or LP and $\mathsf{S4}$, and many others. In such extended form, the Correspondence Theorem shows that major modal logics such as $\mathsf{K}, \mathsf{T}, \mathsf{K4}, \mathsf{S4}, \mathsf{K45}, \mathsf{S5}$ and some others have exact Justification Logic counterparts.

At the core of the Correspondence Theorem is the following Realization Theorem.

> **Theorem 4**: There is an algorithm which, for each modal formula F derivable in K, assigns evidence terms to each occurrence of modality in F in such a way that the resulting formula F^r is derivable in J. Moreover, the realization assigns evidence variables to the negative

occurrences of modal operators in F, thus respecting the existential reading of epistemic modality.

Known realization algorithms which recover evidence terms in modal theorems use cut-free derivations in the corresponding modal logics. Alternatively, the Realization Theorem can be established semantically by Fitting's method or its proper modifications. In principle, these semantic arguments also produce realization procedures which are based on exhaustive search.

It would be a mistake to draw the conclusion that **any** modal logic has a reasonable Justification Logic counterpart. For example the logic of formal provability, GL, (Boolos 1993) contains the *Löb Principle*:

(5) $$\Box(\Box F \to F) \to \Box F,$$

which does not seem to have an epistemically acceptable explicit version. Consider, for example, the case where F is the propositional constant $\bot$ for *false*. If an analogue of Theorem 4 would cover the Löb Principle there would be justification terms s and t such that $x : (s : \bot \to \bot) \to t : \bot$. But this is intuitively false for factive justification. Indeed, $s : \bot \to \bot$ is an instance of the Factivity Axiom. Apply Axiom Internalization to obtain $c : (s : \bot \to \bot)$ for some constant c. This choice of c makes the antecedent of $c : (s : \bot \to \bot) \to t : \bot$ intuitively true and the conclusion false[4]. In particular, the Löb Principle (5) is not valid for the proof interpretation (cf. (Goris 2007) for a full account of which principles of GL are realizable).

The Correspondence Theorem gives fresh insight into epistemic modal logics. Most notably, it provides a new semantics for the 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 can be read as *there is a witness (proof, justification) for F*.

Justification semantics plays a similar role in Modal Logic to that played by Kleene realizability in Intuitionistic Logic. In both cases, the intended semantics is **existential**: the Brouwer-Heyting-Kolmogorov interpretation of Intuitionistic Logic (Heyting 1934, Troelstra and van Dalen 1988, van Dalen 1986) and Gödel's provability reading of $\mathsf{S4}$ (Gödel 1933, Gödel 1938). In both cases there is a possible-world semantics of **universal** character which is a highly potent and dominant technical tool. It does not, however, address the existential character of the intended semantics. It took Kleene realizability (Kleene 1945, Troelstra 1998) to reveal the computational semantics of Intuitionistic Logic and the Logic of Proofs to provide exact BHK semantics of proofs for Intuitionistic and Modal Logic.

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

See Section 4 of the supplementary document Some More Technical Matters for more on Realization Theorems.

5. Generalizations

So far in this article only single-agent justification logics, analogous to single-agent logics of knowledge, have been considered. Justification

Logic can be thought of as logic of *explicit* knowledge, related to more conventional logics of *implicit* knowledge. A number of systems beyond those discussed above have been investigated in the literature, involving multiple agents, or having both implicit and explicit operators, or some combination of these.

5.1 Mixing Explicit and Implicit Knowledge

Since justification logics provide explicit justifications, while conventional logics of knowledge provide an implicit knowledge operator, it is natural to consider combining the two in a single system. The most common joint logic of explicit and implicit knowledge is $\mathsf{S4LP}$ (Artemov and Nogina 2005). The language of $\mathsf{S4LP}$ is like that of LP, but with an implicit knowledge operator added, written either $\mathbf{K}$ or $\Box$. The axiomatics is like that of LP, combined with that of $\mathsf{S4}$ for the implicit operator, together with a connecting axiom, $t : X \rightarrow \Box X$, anything that has an explicit justification is knowable.

Semantically, possible world justification models for LP need no modification, since they already have all the machinery of Hintikka/Kripke models. One models the $\Box$ operator in the usual way, making use of just the accessibility relation, and one models the justification terms as described in Section 3.1 using both accessibility and the evidence function. Since the usual condition for $\Box X$ being true at a world is one of the two clauses of the condition for $t : X$ being true, this immediately yields the validity of $t : X \rightarrow \Box X$, and soundness follows easily. Axiomatic completeness is also rather straightforward.

In $\mathsf{S4LP}$ both implicit and explicit knowledge is represented, but in possible world justification model semantics a single accessibility relation serves for both. This is not the only way of doing it. More generally, an explicit knowledge accessibility relation could be a proper extension of

that for implicit knowledge. This represents the vision of explicit knowledge as having stricter standards for what counts as known than that of implicit knowledge. Using different accessibility relations for explicit and implicit knowledge becomes necessary when these epistemic notions obey different logical laws, e.g., $\mathsf{S5}$ for implicit knowledge and LP for explicit. The case of multiple accessibility relations is commonly known in the literature as Artemov-Fitting models, but will be called multi-agent possible world models here. (cf. Section 5.2).

Curiously, while the logic $\mathsf{S4LP}$ seems quite natural, a Realization Theorem has been problematic for it: no such theorem can be proved if one insists on what are called *normal* realizations (Kuznets 2010). Realization of implicit knowledge modalities in $\mathsf{S4LP}$ by explicit justifications which would respect the epistemic structure remains a major challenge in this area.

Interactions between implicit and explicit knowledge can sometimes be rather delicate. As an example, consider the following mixed principle of negative introspection (again $\Box$ should be read as an implicit epistemic operator),

(6) $$\neg t : X \rightarrow \Box \neg t : X.$$

From the provability perspective, it is the right form of negative introspection. Indeed, let $\Box F$ be interpreted as *F is provable* and $t : F$ as *t is a proof of F* in a given formal theory T, e.g., in Peano Arithmetic PA. Then (6) states a provable principle. Indeed, if t is not a proof of F then, since this statement is decidable, it can be established inside T, hence in T this sentence is provable. On the other hand, the proof p of 't is not a proof of F' depends on both t and F, $p = p(t, F)$ and cannot be computed given t only. In this respect, $\Box$ cannot be replaced by any specific proof term depending on t only and (6) cannot be presented in an entirely explicit justification-style format.

The first examples of explicit/implicit knowledge systems appeared in the area of provability logic. In (Sidon 1997, Yavorskaya (Sidon) 2001), a logic LPP was introduced which combined the logic of provability GL with the logic of proofs LP, but to ensure that the resulting system had desirable logical properties some *additional operations* from outside the original languages of GL and LP were added. In (Nogina 2006, Nogina 2007) a complete logical system, GLA, for proofs and provability was offered, in the sum of the *original languages* of GL and LP. Both LPP and GLA enjoy completeness relative to the class of arithmetical models, and also relative to the class of possible world justification models.

Another example of a provability principle that cannot be made completely explicit is the Löb Principle (5). For each of LPP and GLA, it is easy to find a proof term $l(x)$ such that

(7) $$x : (\Box F \rightarrow F) \rightarrow l(x) : F$$

holds. However, there is no realization which makes all *three* $\Box$ s in (5) explicit. In fact, the set of realizable provability principles is the intersection of GL and $\mathsf{S4}$ (Goris 2007).

5.2 Multi-Agent Possible World Justification Models

In *multi-agent possible world justification models* multiple accessibility relations are employed, with connections between them, (Artemov 2006). The idea is, there are multiple agents, each with an implicit knowledge operator, and there are justification terms, which each agent understands. Loosely, everybody understands explicit reasons; these amount to *evidence-based common knowledge*.

An n-agent possible world justification model is a structure $\langle \mathcal{G}, \mathcal{R}_1, \ldots, \mathcal{R}_n, \mathcal{R}, \mathcal{E}, \mathcal{V}\rangle$ meeting the following conditions. $\mathcal{G}$ is a set of possible worlds. Each of $\mathcal{R}_1,\ldots,\mathcal{R}_n$ is an accessibility relation, one for each agent.

These may be assumed to be reflexive, transitive, or symmetric, as desired. They are used to model implicit agent knowledge for the family of agents. The accessibility relation $\mathcal{R}$ meets the LP conditions, reflexivity and transitivity. It is used in the modeling of explicit knowledge. $\mathcal{E}$ is an evidence function, meeting the same conditions as those for LP in Section 3.3. $\mathcal{V}$ maps propositional letters to sets of worlds, as usual. There is a special condition imposed: for each $i = 1, \ldots, n$, $\mathcal{R}_i \subseteq \mathcal{R}$.

If $\mathcal{M} = \langle \mathcal{G}, \mathcal{R}_1, \ldots, \mathcal{R}_n, \mathcal{R}, \mathcal{E}, \mathcal{V}\rangle$ is a multi-agent possible world justification model a truth-at-a-world relation, $\mathcal{M}, \Gamma \Vdash X$, is defined with most of the usual clauses. The ones of particular interest are these:

- $\mathcal{M}, \Gamma \Vdash K_i X$ if and only if, for every $\Delta \in \mathcal{G}$ with $\Gamma \mathcal{R}_i \Delta$, we have that $\mathcal{M}, \Delta \Vdash X$.
- $\mathcal{M}, \Gamma \Vdash t : X$ if and only if $\Gamma \in \mathcal{E}(t, X)$ and, for every $\Delta \in \mathcal{G}$ with $\Gamma \mathcal{R} \Delta$, we have that $\mathcal{M}, \Delta \Vdash X$.

The condition $\mathcal{R}_i \subseteq \mathcal{R}$ entails the validity of $t : X \rightarrow K_i X$, for each agent i. If there is only a single agent, and the accessibility relation for that agent is reflexive and transitive, this provides another semantics for $\mathsf{S4LP}$. Whatever the number of agents, each agent accepts explicit reasons as establishing knowledge.

A version of LP with two agents was introduced and studied in (Yavorskaya (Sidon) 2008), though it can be generalized to any finite number of agents. In this, each agent has its own set of justification operators, variables, and constants, rather than having a single set for everybody, as above. In addition some limited communication between agents may be permitted, using a new operator that allows one agent to verify the correctness of the other agent's justifications. Versions of both single world and more general possible world justification semantics were created for the two-agent logics. This involves a straightforward extension of the notion of an evidence function, and for possible world justification models, using two accessibility relations. Realization theorems have been proved syntactically, though presumably a semantic proof would also work.

Multi-agent models (where each agent has its own set of justification operators) with explicit and implicit knowledge can be used to epistemically analyze zero-knowledge proofs (Lehnherr, Ognjanovic, and Studer 2022). Zero-knowledge proofs are protocols by which one agent (the prover) can prove to another agent (the verifier) that the prover has certain knowledge (e.g., knows a password) without conveying any information beyond the mere fact of the possession of knowledge (e.g., without revealing the password). The following formulas can be used to describe the situation after the execution of the protocol, where the term s justifies the verifier's knowledge that results from the protocol:

$$s :_V K_P F,$$

meaning the protocol yields a justification s to the verifier V that the prover P knows F; and

$$\neg s :_V t :_P F \text{ for any term t},$$

i.e., for no term t the protocol justifies that the verifier could know that t justifies the prover's knowledge of F. That is, the protocol justifies that the prover knows F but it does not justify any possible evidence for that knowledge.

There has been some exploration of the role of public announcements in multi-agent justification logics (Renne 2008, Renne 2009).

There is more on the notion of evidence-based common knowledge in Section 5 of the supplementary document Some More Technical Matters.

Besides multi-agent epistemic logics, there are other justification logics that feature two types of terms. Kuznets, Marin, and Strassburger (2021) introduce an explicit version of constructive modal logic. There, the $\Box$-modality is realized by proof terms like in LP. To realize the $\Diamond$-modality, a second kind of terms is introduced, which are called witness terms. In constructive modal logic, the formula $\Diamond F$ means F *is consistent*. In its realization $s : F$, the witness term s represents an abstract witnessing model for the formula F.

Another example is dyadic deontic logic (DDL), which can be axiomatized by two modalities $\Box$ and $\bigcirc$. The formula $\Box F$ means F *is settled true*, and the conditional $\bigcirc(F/G)$ means F *is obligatory given* G. Faroldi, Rohani, and Studer (2023) consider an explicit version of DDL. Again, $\Box F$ is realized by a proof term as in LP, whereas $\bigcirc(F/G)$ is realized by making use of a new type of terms that represent deontic reasons.

6. Russell's Example: Induced Factivity

There is a technique for using Justification Logic to analyze different justifications for the same fact, in particular when some of the justifications are factive and some are not. To demonstrate the technique consider a well-known example:

> If a man believes that the late Prime Minister's last name began with a 'B,' he believes what is true, since the late Prime Minister was Sir Henry Campbell Bannerman[5]. But if he believes that Mr. Balfour was the late Prime Minister[6], he will still believe that the late Prime Minister's last name began with a 'B,' yet this belief, though true, would not be thought to constitute knowledge. (Russell 1912)

As in the Red Barn Example, discussed in Section 1.1, here one has to deal with two justifications for a true statement, one of which is correct and one of which is not. Let B be a sentence (propositional atom), w be a designated justification variable for the wrong reason for B and r a designated justification variable for the right (hence factive) reason for B. Then, Russell's example prompts the following set of assumptions[7]:

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

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

1. $r : B$ (assumption)
2. $r : B \to B$ (assumption)
3. B (from 1 and 2 by Modus Ponens)
4. $B \to (w : B \to B)$ (propositional axiom)
5. $w : B \to B$ (from 3 and 4 by Modus Ponens)

However, this derivation utilizes the fact that r is a factive justification for B to conclude $w : B \to B$, which constitutes a case of 'induced factivity' for $w : B$. The question is, how can one distinguish the 'real' factivity of $r : B$ from the 'induced factivity' of $w : B$? Some sort of evidence-tracking is needed here, and Justification Logic is an appropriate tool. The natural approach is to consider the set of assumptions **without** $r : B$, i.e.,

$$\mathcal{S} = \{w : B, r : B \to B\}$$

and establish that factivity of w, i.e., $w : B \to B$ is not derivable from $\mathcal{S}$. Here is a possible world justification model $\mathcal{M} = (\mathcal{G}, \mathcal{R}, \mathcal{E}, \mathcal{V})$ in which $\mathcal{S}$ holds but $w : B \to B$ does not:

- $\mathcal{G} = \{\mathbf{1}\}$,
- $\mathcal{R} = \varnothing$,
- $\mathcal{V}(B) = \varnothing$ (and so not-$\mathbf{1} \Vdash B$),

- $\mathcal{E}(t, F) = \{\mathbf{1}\}$ for all pairs (t, F) except (r, B),and
- $\mathcal{E}(r, B) = \varnothing$.

It is easy to see that the closure conditions *Application* and *Sum* on $\mathcal{E}$ are fulfilled. At $\mathbf{1}, w : B$ holds, i.e.,

$$\mathbf{1} \Vdash w : B$$

since w is admissible evidence for B at $\mathbf{1}$ and there are no possible worlds accessible from $\mathbf{1}$. Furthermore,

$$\text{not-}\mathbf{1} \Vdash r : B$$

since, according to $\mathcal{E}, r$ is not admissible evidence for B at $\mathbf{1}$. Hence:

$$\mathbf{1} \Vdash r : B \to B$$

On the other hand,

$$\text{not-}\mathbf{1} \Vdash w : B \to B$$

since B does not hold at $\mathbf{1}$.

7. Self-referentiality of justifications

The Realization algorithms sometimes produce Constant Specifications containing self-referential justification assertions $c : A(c)$, that is, assertions in which the justification (here c) occurs in the asserted proposition (here $A(c)$).

Self-referentiality of justifications is a new phenomenon which is not present in the conventional modal language. In addition to being intriguing epistemic objects, such self-referential assertions provide a special challenge from the semantical viewpoint because of the built-in vicious circle. Indeed, to evaluate c one would expect first to evaluate A and then assign a justification object for A to c. However, this cannot be done since

A contains c which is yet to be evaluated. The question of whether or not modal logics can be realized without using self-referential justifications was a major open question in this area.

The principal result by Kuznets in (Brezhnev and Kuznets 2006) states that self-referentiality of justifications is unavoidable in realization of $\mathsf{S4}$ in LP. The current state of things is given by the following theorem due to Kuznets:

> **Theorem 5**: Self-referentiality can be avoided in realizations of modal logics K and D. Self-referentiality cannot be avoided in realizations of modal logics $\mathsf{T}, \mathsf{K4}, \mathsf{D4}$ and $\mathsf{S4}$.

This theorem establishes that a system of justification terms for $\mathsf{S4}$ will necessarily be self-referential. This creates a serious, though not directly visible, constraint on provability semantics. In the Gödelian context of arithmetical proofs, the problem was coped with by a general method of assigning arithmetical semantics to self-referential assertions $c : A(c)$ stating that c is a proof of $A(c)$. In the Logic of Proofs LP it was dealt with by a non-trivial fixed-point construction.

Self-referentiality gives an interesting perspective on Moore's Paradox. See Section 6 of the supplementary document Some More Technical Matters for details.

The question of the self-referentiality of BHK-semantics for intuitionistic logic IPC has been answered by Junhua Yu (Yu 2014). Extending Kuznets' method, he established

> **Theorem 6**: Each LP realization of the intuitionistic law of double negation $\neg\neg(\neg\neg p \to p)$ requires self referential constant specifications.

More generally, Yu has proved that any double negation of a classical tautology (by Glivenko's Theorem all of them are theorems of IPC) needs self-referential constant specifications for its realization in LC. Another example of unavoidable self-referentiality was found by Yu in the purely implicational fragment of IPC. This suggests that the BHK semantics of intuitionistic logic (even just of intuitionistic implication) is intrinsically self-referential and needs a fixed-point construction to connect it to formal proofs in PA or similar systems. This might explain, in part, why any attempt to build provability BHK semantics in a direct inductive manner without self-referentiality was doomed to failure.

8. Quantifiers in Justification Logic

While the investigation of propositional Justification Logic is far from complete, there has also been some work on first-order versions. Quantified versions of Modal Logic already offer complexities beyond standard first-order logic. Quantification has an even broader field to play when Justification Logics are involved. Classically one quantifies over 'objects,' and models are equipped with a domain over which quantifiers range. Modally one might have a single domain common to all possible worlds, or one might have separate domains for each world. The role of the Barcan formula is well-known here. Both constant and varying domain options are available for Justification Logic as well. In addition there is a possibility that has no analog for Modal Logic: one might quantify over justifications themselves.

Initial results concerning the possibility of Quantified Justification Logic were notably unfavorable. The arithmetical provability semantics for the Logic of Proofs LP, naturally generalizes to a first-order version with conventional quantifiers, and to a version with quantifiers over proofs. In both cases, axiomatizability questions were answered negatively.

Theorem 7: The first-order logic of proofs is not recursively enumerable (Artemov and Yavorskaya (Sidon) 2001). The logic of proofs with quantifiers over proofs is not recursively enumerable (Yavorsky 2001).

Although an arithmetic semantics is not possible, in (Fitting 2008) a possible world semantics, and an axiomatic proof theory, was given for a version of LP with quantifiers ranging over justifications. Soundness and completeness were proved. At this point possible world semantics separates from arithmetic semantics, which may or may not be a cause for alarm. It was also shown that $\mathsf{S4}$ embeds into the quantified logic by translating $\Box Z$ as "there exists a justification x such that $x : Z^*$," where Z^* is the translation of Z. While this logic is somewhat complicated, it has found applications, e.g., in (Dean and Kurokawa 2009b) it is used to analyze the Knower Paradox, though objections have been raised to this analysis in (Arlo-Costa and Kishida 2009).

A First-Order Logic of Proofs, FOLP, with quantifiers over individual variables, has been presented in Artemov and Yavorskaya (Sidon) (2011). In FOLP proof assertions are represented by formulas of the form $t{:}_X A$ where X is a finite set of individual variables that are considered global parameters open for substitution. All occurrences of variables from X that are free in A are also free in $t{:}_X A$. All other free variables of A are considered local and hence bound in $t{:}_X A$. For example, if $A(x, y)$ is an atomic formula, then in $p{:}_{\{x\}} A(x, y)$ variable x is free and variable y is bound. Likewise, in $p{:}_{\{x,y\}} A(x, y)$ both variables are free, and in $p{:}_{\emptyset} A(x, y)$ neither x nor y is free.

Proofs (justifications) are represented by proof terms which do not contain individual variables. In addition to LP operations there is one more series of operations on proof terms, $\mathsf{gen}_x(t)$, corresponding to generalization over individual variable x. The new axiom that governs this operation is

$t{:}_X A \rightarrow \mathsf{gen}{:}_x(\mathsf{t})_X \forall x A$, with $x \notin X$. The complete list of FOLP principles along with realization of First-Order $\mathsf{S4}$ can be found in Artemov and Yavorskaya (Sidon) (2011). A semantics for FOLP has been developed in Fitting (2014a).

9. Historical Notes

The initial Justification Logic system, the Logic of Proofs LP, was introduced in 1995 in (Artemov 1995) (cf. also (Artemov 2001)) where such basic properties as Internalization, Realization, arithmetical completeness, were first established. LP offered an intended provability semantics for Gödel's provability logic $\mathsf{S4}$, thus providing a formalization of Brouwer-Heyting-Kolmogorov semantics for intuitionistic propositional logic. Epistemic semantics and completeness (Fitting 2005) were first established for LP. Symbolic models and decidability for LP are due to Mkrtychev (Mkrtychev 1997). Complexity estimates first appeared in (Brezhnev and Kuznets 2006, Kuznets 2000, Milnikel 2007). A comprehensive overview of all decidability and complexity results can be found in (Kuznets 2008). Systems J, $\mathsf{J4}$, and JT were first considered in (Brezhnev 2001) under different names and in a slightly different setting. $\mathsf{JT45}$ appeared independently in (Pacuit 2006) and (Rubtsova 2006), and $\mathsf{JD45}$ in (Pacuit 2006). The logic of uni-conclusion proofs has been found in (Krupski 1997). A more general approach to common knowledge based on justified knowledge was offered in (Artemov 2006). Game semantics of Justification Logic and Dynamic Epistemic Logic with justifications were studied in (Renne 2008, Renne 2009). Connections between Justification Logic and the problem of logical omniscience were examined in (Artemov and Kuznets 2009, Artemov and Kuznets 2014, Wang 2009). The name *Justification Logic* was introduced in (Artemov 2008), in which Kripke, Russell, and Gettier examples were formalized; this formalization has been used for the resolution of paradoxes, verification, hidden assumption analysis, and eliminating redundancies. In (Dean and Kurokawa 2009a), Justification Logic was used for the analysis of Knower and Knowability paradoxes.

The first two monographs on Justification Logic were published in 2019 (Artemov and Fitting 2019, Kuznets and Studer 2019).

Bibliography

Antonakos, E., 2007. "Justified and Common Knowledge: Limited Conservativity", in S. Artemov and A. Nerode (eds.), *Logical Foundations of Computer Science, International Symposium, LFCS 2007, New York, NY, USA, June 4–7, 2007, Proceedings* (Lecture Notes in Computer Science: Volume 4514), Berlin: Springer, pp. 1–11.

Arlo-Costa, H. and K. Kishida, 2009. "Three proofs and the Knower in the Quantified Logic of Proofs", in *Formal Epistemology Workshop / FEW 2009. Proceedings*, Carnegie Mellon University, Pittsburgh, PA, USA.

Artemov, S., 1995. "Operational modal logic", Technical Report MSI 95–29, Cornell University.

–––, 2001. "Explicit provability and constructive semantics", *The Bulletin of Symbolic Logic*, 7(1): 1–36.

–––, 2006. "Justified common knowledge", *Theoretical Computer Science*, 357 (1–3): 4–22.

–––, 2008. "The logic of justification", *The Review of Symbolic Logic*, 1(4): 477–513.

–––, 2012. "The Ontology of Justifications in the Logical Setting." *Studia Logica*, 100(1–2): 17–30.

Artemov, S. and M. Fitting, 2019. *Justification Logic: Reasoning with Reasons*, New York: Cambridge University Press.

Artemov, S. and R. Kuznets, 2009. "Logical omniscience as a computational complexity problem", in A. Heifetz (ed.), *Theoretical Aspects of Rationality and Knowledge, Proceedings of the Twelfth Conference* (TARK 2009), ACM Publishers, pp. 14–23.

–––, 2014. "Logical omniscience as infeasibility", *Annals of Pure and Applied Logic*, 165(1): 6–25.

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

Artemov, S. and T. Yavorskaya (Sidon), 2001. "On first-order logic of proofs", *Moscow Mathematical Journal*, 1(4): 475–490.

–––, 2011. "First-Order Logic of Proofs." TR–2011005, City University of New York, Ph.D. Program in Computer Science.

Boolos, G., 1993. *The Logic of Provability*, Cambridge: Cambridge University Press.

Brezhnev, V., 2001. "On the logic of proofs", in K. Striegnitz (ed.), *Proceedings of the Sixth ESSLLI Student Session, 13th European Summer School in Logic, Language and Information* (ESSLLI'01), pp. 35–46.

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

Cubitt, R. P. and R. Sugden, 2003. "Common knowledge, salience and convention: A reconstruction of David Lewis' game theory", *Economics and Philosophy*, 19: 175–210.

Dean, W. and H. Kurokawa, 2009a. "From the Knowability Paradox to the existence of proofs", *Synthese*, 176(2): 177–225.

–––, 2009b. "Knowledge, proof and the Knower", in A. Heifetz (ed.), *Theoretical Aspects of Rationality and Knowledge, Proceedings of the Twelfth Conference* (TARK 2009), ACM Publications, pp. 81–90.

Dretske, F., 2005. "Is Knowledge Closed Under Known Entailment? The Case against Closure", in M. Steup and E. Sosa (eds.), *Contemporary Debates in Epistemology*, Oxford: Blackwell, pp. 13–26.

Fagin, R., and J. Y. Halpern, 1988. "Belief, Awareness, and Limited Reasoning." *Artificial Intelligence*, 34: 39–76.

Fagin, R., J. Halpern, Y. Moses, and M. Vardi, 1995. *Reasoning About Knowledge*, Cambridge, MA: MIT Press.

Faroldi, F. L. G., M. Ghari, E. Lehmann, and T. Studer, 2024. “Consistency and permission in deontic justification logic”, *Journal of Logic and Computation*, 34(4): 640–664.

Faroldi, F. L. G., A. Rohani, and T. Studer, 2023. “Conditional Obligations in Justification Logic”, in H.H. Hansen, A. Scedrov, R. de Queiroz (eds), *Logic, Language, Information, and Computation, WoLLIC 2023* (Lecture Notes in Computer Science: Volume 13923), Cham: Springer, pp. 178–193.

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

–––, 2006. “A replacement theorem for **LP**”, Technical Report TR-2006002, Department of Computer Science, City University of New York.

–––, 2008. “A quantified logic of evidence”, *Annals of Pure and Applied Logic*, 152(1–3): 67–83.

–––, 2009. “Realizations and **LP**”, *Annals of Pure and Applied Logic*, 161(3): 368–387.

–––, 2014a. “Possible World Semantics for First Order Logic of Proofs.” *Annals of Pure and Applied Logic* 165: 225–40.

–––, 2014b. “Justification Logics and Realization.” TR-2014004, City University of New York, Ph.D. Program in Computer Science.

Gettier, E., 1963. “Is Justified True Belief Knowledge?” *Analysis*, 23: 121–123.

Girard, J.-Y., P. Taylor, and Y. Lafont, 1989. *Proofs and Types* (Cambridge Tracts in Computer Science: Volume 7), Cambridge: Cambridge University Press.

Gödel, K., 1933. “Eine Interpretation des intuitionistischen Aussagenkalkuls”, *Ergebnisse Math. Kolloq.*, 4: 39–40. English translation in: S. Feferman *et al*. (eds.), *Kurt Gödel Collected Works* (Volume 1), Oxford and New York: Oxford University Press and Clarendon Press, 1986, pp. 301–303.

–––, 1938. “Vortrag bei Zilsel/Lecture at Zilsel’s” (*1938a), in S. Feferman, J. J. Dawson, W. Goldfarb, C. Parsons, and R. Solovay (eds.), *Unpublished Essays and Lectures* (Kurt Gödel Collected Works: Volume III), Oxford: Oxford University Press, 1995, pp. 86–113.

Goldman, A., 1967. “A causal theory of meaning”, *The Journal of Philosophy*, 64: 335–372.

Goodman, N., 1970. “A theory of constructions is equivalent to arithmetic”, in J. Myhill, A. Kino, and R. Vesley (eds.), *Intuitionism and Proof Theory*, Amsterdam: North-Holland, pp. 101–120.

Goris, E., 2007. “Explicit proofs in formal provability logic”, in S. Artemov and A. Nerode (eds.), *Logical Foundations of Computer Science, International Symposium, LFCS 2007, New York, NY, USA, June 4–7, 2007, Proceedings* (Lecture Notes in Computer Science: Volume 4514), Berlin: Springer, pp. 241–253.

Lehnherr, D., Z. Ognjanovic, and T. Studer, 2022. “A logic of interactive proofs”, *Journal of Logic and Computation*, 32(8): 1645–1658.

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

Heyting, A., 1934. *Mathematische Grundlagenforschung. Intuitionismus. Beweistheorie*, Berlin: Springer.

Hintikka, J., 1962. *Knowledge and Belief*, Ithaca: Cornell University Press.

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

Kolmogorov, A., 1932. “Zur Deutung der Intuitionistischen Logik”, *Mathematische Zeitschrift*, 35: 58–65. English translation in V.M. Tikhomirov (ed.), *Selected works of A.N. Kolmogorov. Volume I: Mathematics and Mechanics*, Dordrecht: Kluwer, 1991, pp. 151–158.

Kreisel, G., 1962. “Foundations of intuitionistic logic”, in E. Nagel, P. Suppes, and A. Tarski (eds.), *Logic, Methodology and Philosophy of Science. Proceedings of the 1960 International Congress*, Stanford: Stanford University Press, pp. 198–210.

–––, 1965. “Mathematical logic”, in T. Saaty (ed.), *Lectures in Modern Mathematics III*, New York: Wiley and Sons, pp. 95–195.

Krupski, V., 1997. “Operational logic of proofs with functionality condition on proof predicate”, in S. Adian and A. Nerode (eds.), *Logical Foundations of Computer Science, 4th International Symposium, LFCS’97, Yaroslavl, Russia, July 6–12, 1997, Proceedings* (Lecture Notes in Computer Science: Volume 1234), Berlin: Springer, pp. 167–177.

Kurokawa, H., 2009. “Tableaux and Hypersequents for Justification Logic”, in S. Artemov and A. Nerode (eds.), *Logical Foundations of Computer Science, International Symposium, LFCS 2009, Deerfield Beach, FL, USA, January 3–6, 2009, Proceedings* (Lecture Notes in Computer Science: Volume 5407), Berlin: Springer, pp. 295–308.

Kuznets, R., 2000. “On the Complexity of Explicit Modal Logics”, in P. Clote and H. Schwichtenberg (eds.), *Computer Science Logic, 14th International Workshop, CSL 2000, Annual Conference of the EACSL, Fischbachau, Germany, August 21–26, 2000, Proceedings* (Lecture Notes in Computer Science: Volume 1862), Berlin: Springer, pp. 371–383.

–––, 2008. *Complexity Issues in Justification Logic*, Ph. D. dissertation, Computer Science Department, City University of New York Graduate Center.

–––, 2010. “A note on the abnormality of realizations of **S4LP**”, in K. Brünnler and T. Studer (eds.), *Proof, Computation, Complexity PCC 2010, International Workshop, Proceedings*, IAM Technical Reports IAM-10-001, Institute of Computer Science and Applied Mathematics, University of Bern.

Kuznets, R., S. Marin, and L. Strassburger, 2021. “Justification logic for constructive modal logic”, *Journal of Applied Logics*, 8(8): 2313–2332.

Kuznets, R. and T. Studer, 2012. “Justifications, Ontology, and Conservativity”, in *Advances in Modal Logic* (Volume 9), Thomas Bolander, Torben Braüner, Silvio Ghilardi, and Lawrence Moss (eds.), London: College Publications, 437–58.

–––, 2019. *Logics of Proofs and Justifications*, London: College Publications.

Lemmon, E. J., and Dana S. Scott, 1977. *The “Lemmon Notes”: An Introduction to Modal Logic* (American Philosophical Quarterly Monograph 11), Oxford: Blackwell.

McCarthy, J., M. Sato, T. Hayashi, and S. Igarishi, 1978. “On the model theory of knowledge”, Technical Report STAN-CS-78-667, Department of Computer Science, Stanford University.

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

–––, 2009. “Conservativity for Logics of Justified Belief”, in S. Artemov and A. Nerode (eds.), *Logical Foundations of Computer Science, International Symposium, LFCS 2009, Deerfield Beach, FL, USA, January 3–6, 2009, Proceedings* (Lecture Notes in Computer Science: Volume 5407), Berlin: Springer, pp. 354–364.

Mkrtychev, A., 1997. “Models for the Logic of Proofs”, in S. Adian and A. Nerode (eds.), *Logical Foundations of Computer Science, 4th International Symposium, LFCS’97, Yaroslavl, Russia, July 6–12, 1997, Proceedings* (Lecture Notes in Computer Science: Volume 1234), Berlin: Springer, pp. 266–275.

Nogina, E., 2006. “On logic of proofs and provability”, in *2005 Summer Meeting of the Association for Symbolic Logic, Logic Colloquium’05,* *Athens, Greece* (July 28–August 3, 2005), *The Bulletin of Symbolic Logic*, 12(2): 356.

–––, 2007. "Epistemic completeness of **GLA**", in *2007 Annual Meeting of the Association for Symbolic Logic, University of Florida, Gainesville, Florida* (March 10–13, 2007), *The Bulletin of Symbolic Logic*, 13(3): 407.

Pacuit, E., 2006. "A Note on Some Explicit Modal Logics", Technical Report PP–2006–29, Institute for Logic, Language and Computation, University of Amsterdam.

Plaza, J., 2007. "Logics of public communications", *Synthese*, 158(2): 165–179.

Renne, B., 2008. *Dynamic Epistemic Logic with Justification*, Ph. D. thesis, Computer Science Department, CUNY Graduate Center, New York, NY, USA.

–––, 2009. "Evidence Elimination in Multi-Agent Justification Logic", in A. Heifetz (ed.), *Theoretical Aspects of Rationality and Knowledge, Proceedings of the Twelfth Conference (TARK 2009)*, ACM Publications, pp. 227–236.

Rose, G., 1953. "Propositional calculus and realizability", *Transactions of the American Mathematical Society*, 75: 1–19.

Rubtsova, N., 2006. "On Realization of **S5**-modality by Evidence Terms", *Journal of Logic and Computation*, 16(5): 671–684.

Russell, B., 1912. *The Problems of Philosophy*, Oxford: Oxford University Press.

Sedlár, Igor. 2013. "Justifications, Awareness and Epistemic Dynamics", in S. Artemov and A. Nerode (eds.), *Logical Foundations of Computer Science* (Lecture Notes in Computer Science: 7734), Berlin/Heidelberg: Springer, 307–18.

Sidon, T., 1997. "Provability logic with operations on proofs", in S. Adian and A. Nerode (eds.), *Logical Foundations of Computer Science, 4th International Symposium, LFCS'97, Yaroslavl, Russia, July 6–12, 1997, Proceedings* (Lecture Notes in Computer Science: Volume 1234), Berlin: Springer, pp. 342–353.

Troelstra, A., 1998. "Realizability", in S. Buss (ed.), *Handbook of Proof Theory*, Amsterdam: Elsevier, pp. 407–474.

Troelstra, A. and H. Schwichtenberg, 1996. *Basic Proof Theory*, Amsterdam: Cambridge University Press.

Troelstra, A. and D. van Dalen, 1988. *Constructivism in Mathematics* (Volumes 1, 2), Amsterdam: North–Holland.

van Dalen, D., 1986. "Intuitionistic logic", in D. Gabbay and F. Guenther (eds.), *Handbook of Philosophical Logic* (Volume 3), Bordrecht: Reidel, pp. 225–340.

van Ditmarsch, H., W. van der Hoek, and B. Kooi (eds.), 2007. *Dynamic Epistemic Logic* (Synthese Library: Volume 337), Berlin: Springer..

von Wright, G., 1951. *An Essay in Modal Logic*, Amsterdam: North-Holland.

Wang, R.-J., 2009. "Knowledge, Time, and Logical Omniscience", in H. Ono, M. Kanazawa, and R. de Queiroz (eds.), *Logic, Language, Information and Computation, 16th International Workshop, WoLLIC 2009, Tokyo, Japan, June 21–24, 2009, Proceedings* (Lecture Notes in Artificial Intelligence: Volume 5514), Berlin: Springer, pp. 394–407.

Yavorskaya (Sidon), T., 2001. "Logic of proofs and provability", *Annals of Pure and Applied Logic*, 113 (1–3): 345–372.

–––, 2008. "Interacting Explicit Evidence Systems", *Theory of Computing Systems*, 43(2): 272–293.

Yavorsky, R., 2001. "Provability logics with quantifiers on proofs", *Annals of Pure and Applied Logic*, 113 (1–3): 373–387.

Yu, J., 2014. "Self-Referentiality of Brouwer-Heyting-Kolmogorov semantics", *Annals of Pure and Applied Logic*, 165: 371–388.

Academic Tools

- How to cite this entry.
- Preview the PDF version of this entry at the Friends of the SEP Society.
- Look up topics and thinkers related to this entry at the Internet Philosophy Ontology Project (InPhO).
- Enhanced bibliography for this entry at PhilPapers, with links to its database.

Other Internet Resources

- Justification Logic Bibliography, a complete bibliography of material on justification logic up to February 2014. Maintained by Roman Kuznets.

Related Entries

belief, formal representations of | logic: modal | logic: provability

Acknowledgments

Beginning with the 2024 update, Thomas Studer has taken responsibility for updating and maintaining this entry.

Some More Technical Matters

1. Mathematical Logic Tradition

Several well-known mathematical notions which appeared prior to Justification Logic have sometimes been perceived as related to the BHK idea: Kleene realizability (Troelstra 1998), Curry-Howard isomorphism (Girard, Taylor, and Lafont 1989, Troelstra and Schwichtenberg 1996), Kreisel-Goodman theory of constructions (Goodman 1970, Kreisel 1962, Kreisel 1965), just to name a few. These interpretations have been very instrumental for understanding intuitionistic logic, though none of them qualifies as the *BHK* semantics.

Kleene realizability revealed a fundamental *computational content* of formal intuitionistic derivations, however it is still quite different from the intended *BHK* semantics. Kleene realizers are computational programs rather then proofs. The predicate ‘*r realizes F* ’ is not decidable, which leads to some serious deviations from intuitionistic logic. Kleene realizability is not adequate for the intuitionistic propositional calculus IPC. There are realizable propositional formulas not derivable in IPC (Rose 1953).[8]

The Curry-Howard isomorphism transliterates natural derivations in IPC to typed λ-terms thus providing a generic functional reading for logical derivations. However the foundational value of this interpretation is limited since, as proof objects, Curry-Howard λ-terms denote nothing but derivations in IPC itself and thus yield a circular provability semantics for the latter.

An attempt to formalize the *BHK* semantics directly was made by Kreisel in his theory of constructions (Kreisel 1962, Kreisel 1965). The original variant of the theory was inconsistent; difficulties already occurred at the propositional level. In (Goodman 1970) this was fixed by introducing a stratification of constructions into levels, which ruined the *BHK* character of this semantics. In particular, a proof of $A \to B$ was no longer a construction that could be applied to any proof of A

2. Logical Awareness and Constant Specifications

Two examples in J are presented, showing modal theorems of K, and realizations for them. In the examples indices on constants have been omitted.

Example 1. This shows how to build a justification of a conjunction from justifications of the conjuncts. In traditional modal language, this principle is formalized as $\Box A \wedge \Box B \to \Box(A \wedge B)$. In J this idea is expressed in the more precise justification language.

1. $A \to (B \to (A \wedge B))$ (propositional axiom)
2. $c : (A \to (B \to (A \wedge B)))$ (from 1 by Axiom Internalization)
3. $x : A \to [c \cdot x] : (B \to (A \wedge B))$ (from 2 by Application and Modus Ponens)
4. $x : A \to (y : B \to [[c \cdot x] \cdot y] : (A \wedge B))$ (from 3 by Application and propositional reasoning)
5. $x : A \wedge y : B \to [[c \cdot x] \cdot y] : (A \wedge B)$ (from 5 by propositional reasoning)

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

$x : A \wedge y : B \to [[c \cdot x] \cdot y] : (A \wedge B)$, given $c : (A \to (B \to (A \wedge B)))$.

Example 2. This example shows how to build a justification of a disjunction from justifications of either of the disjuncts. In the usual modal language this is represented by $\Box A \vee \Box B \to \Box(A \vee B)$. Here is the corresponding result in J.

1. $A \to (A \vee B)$ (classical logic)
2. $a : (A \to (A \vee B))$ (from 1 by Axiom Internalization)
3. $x : A \to [a \cdot x] : (A \vee B)$ (from 2 by Application and Modus Ponens)
4. $B \to (A \vee B)$ (by classical logic)
5. $b : (B \to (A \vee B))$ (from 4 by Axiom Internalization)
6. $y : B \to [b \cdot y] : (A \vee B)$ (from 5 by Application and Modus Ponens)
7. $[a \cdot x] : (A \vee B) \to [a \cdot x + b \cdot y] : (A \vee B)$ (by Sum)
8. $[b \cdot y] : (A \vee B) \to [a \cdot x + b \cdot y] : (A \vee B)$ (by Sum)
9. $(x : A \vee y : B) \to [a \cdot x + b \cdot y] : (A \vee B)$ (from 3, 6, 7, and 8 by propositional reasoning)

The complete reading of the result of this derivation is:

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

3. Single-Agent Possible World Models for J

Here are two models in each of which $x : P \to x : (P \wedge Q)$ is not valid. (P and Q are atomic and x is a justification variable.) It was pointed out that a formula $t : X$ might fail at a possible world either because X is not believable there (it is false at some accessible world), or because t is not an appropriate reason for X. The two models illustrate both versions.

First, consider the model $\mathcal{M}$ having a single state, Γ, accessible to itself, and with an evidence function such that $\mathcal{E}(x, Z)$ is Γ, for every formula Z. In this model, x serves as‘universal’ evidence. Use a valuation such that $\mathcal{V}(P) = \Gamma$ and $\mathcal{V}(Q) = \varnothing$. Then one has $\mathcal{M}, \Gamma \Vdash x : P$ but not-$\mathcal{M}, \Gamma \Vdash x : (P \wedge Q)$ because, even though x serves as universal evidence, $P \wedge Q$ is not believable at Γ in the Hintikka/Kripke sense because Q is not true.

Next consider the model $\mathcal{N}$, again having a single state Γ accessible to itself. This time take $\mathcal{V}$ to be the mapping assigning Γ to every propositional letter. But also, set $\mathcal{E}(x,P)=\Gamma, \mathcal{E}(x,Z)=\varnothing$ for $Z\neq P$, and otherwise $\mathcal{E}$ doesn't matter for this example. Then of course both P and $P\wedge Q$ are believable at Γ, but $\mathcal{N},\Gamma\Vdash x:P$ and not-$\mathcal{N},\Gamma\Vdash x:(P\wedge Q)$, the latter because x does not serve as evidence for $P\wedge Q$ at Γ.

In Hintikka/Kripke models, believability and knowability are essentially semantic notions, but the present treatment of evidence is more of a syntactic nature. For example, the model $\mathcal{N}$ above also invalidates $x:P\to x:(P\wedge P)$. At first glance this is surprising, since in any standard logic of knowledge or belief $\Box P\to\Box(P\wedge P)$ is valid. But, just because x serves as evidence for P, it need not follow that it also serves as evidence for $P\wedge P$. The formulas are syntactically different, and effort is needed to recognize that the later formula is a redundant version of the former. To take this to an extreme, consider the formula $x:P\to x:(P\wedge P\wedge\ \ldots\ \wedge P)$, where the consequent has as many conjuncts as there are elementary particles in the universe! In brief, Hintikka/Kripke style knowledge is knowledge of *propositions*, but justification terms justify *sentences*.

4. Realization Theorems

Here is an example of an S4-derivation realized as an LP-derivation in the style of the Realization theorem. There are two columns in the table below. The first is a Hilbert-style S4-derivation of a modal formula $\Box A\vee\Box B\to\Box(\Box A\vee B)$. The second column displays corresponding steps of an LP-derivation of a formula:

$$x:A\vee y:B\to(a\cdot!x+b\cdot y):(x:A\vee B)$$

with constant specification

$$\{a:(x:A\to x:A\vee B), b:(B\to x:A\vee B)\}.$$

Comparing derivations in S4 and LP

	Derivation in S4	Derivation in LP
1.	$\Box A\to\Box A\vee B$	$x:A\to x:A\vee B$
2.	$\Box(\Box A\to\Box A\vee B)$	$a:(x:A\to x:A\vee B)$
3.	$\Box\Box A\to\Box(\Box A\vee B)$	$!x:x:A\to(a\cdot!x):(x:A\vee B)$
4.	$\Box A\to\Box\Box A$	$x:A\to!x:x:A$
5.	$\Box A\to\Box(\Box A\vee B)$	$x:A\to(a\cdot!x):(x:A\vee B)$
5′.		$(a\cdot!x):(x:A\vee B)\to(a\cdot!x+b\cdot y):(x:A\vee B)$
5″.		$x:A\to(a\cdot!x+b\cdot y):(x:A\vee B)$
6.	$B\to\Box A\vee B$	$B\to x:A\vee B$
7.	$\Box(B\to\Box A\vee B)$	$b:(B\to x:A\vee B)$
8.	$\Box B\to\Box(\Box A\vee B)$	$y:B\to(b\cdot y):(x:A\vee B)$
8′.		$(b\cdot y):(x:A\vee B)\to(a\cdot!x+b\cdot y):(x:A\vee B)$
8″.		$y:B\to(a\cdot!x+b\cdot y):(x:A\vee B)$
9.	$\Box A\vee\Box B\to\Box(\Box A\vee B)$	$x:A\vee y:B\to(a\cdot!x+b\cdot y):(x:A\vee B)$

Extra steps $5', 5'', 8'$, and $8''$ are needed in the LP case to reconcile different internalized proofs of the same formula: $(a\cdot!x):(x:A\vee B)$ and $(b\cdot y):(x:A\vee B)$. The resulting realization respects Skolem's idea that negative occurrences of existential quantifiers (here over proofs hidden in the modality of provability) are realized by free variables whereas positive occurrences are realized by functions of those variables.

Proof theory plays an important role in the study of Justification Logics. Axiom systems we represented in this article, and a sequent calculus was introduced in (Artemov 1995, Artemov 2001). It has the curious disadvantage that it is cut free, but does not have the subformula property —no version with the subformula property is known. More recently, other kinds of proof procedures for Justification Logics have been created. Kurokawa uses hypersequents to provide a comprehensive proof-theoretical treatment of major systems of Justification Logic, (Kurokawa 2009), including those which combine implicit and explicit knowledge. These systems are notoriously hard to analyze and Kurokawa's results constitute a remarkable fundamental contribution to this area.

Realizations have been investigated for their own sake, (Fitting 2009). The basic results all are algorithmic in nature. For example, in Modal Logic a Replacement Theorem holds just as it does classically: if $X \equiv X'$ is provable in a normal modal logic then so is $F \equiv F'$ where F is a formula and F' is like F except that some subformula occurrence X has been replaced with X'. Indeed this can be strengthened to establish that if $X \to X'$ is provable, and X has a *positive* designated subformula occurrence in F, while F' replaces that occurrence with X', then $F \to F'$ is provable. This does not carry over in a simple way to Justification Logic. Roughly speaking, an occurrence of X in a formula F of Justification Logic may be within the scope of various justification terms, and in F' these will need to be 'adjusted' to take into account a justification for $X \to X'$. It turns out this is not simple, but an algorithm for doing so has been developed, (Fitting 2006, Fitting 2009).

Here is another result having a similar proof. In a modal sequent calculus argument a typical step is the following.

$$\frac{S_1 \to S_2, X \qquad S_1 \to S_2, Y}{S_1 \to S_2, X \wedge Y}$$

The notion of a realization easily extends from formulas to sequents. Suppose both of the premise sequents above have realizations, and one would like a realization for the consequent sequent. The problem is, the two premise realizations may be quite different. Algorithmic machinery for merging them has been developed that does exactly this. This

algorithm, in turn, serves as part of a new algorithm for the Realization Theorem itself.

Finally, all initial realization proofs were constructive and were based on some kind of cut-free proof system. But non-constructive, semantic arguments have been developed, which allow the extension of realization machinery substantially. For instance, it is now known that the family of modal logics having justification counterparts is infinite. Artemov and Fitting 2019 contains a detailed investigation along these lines.

5. Multi-Agent Justification Models

To simplify the language, one can use the forgetful projection which replaces explicit knowledge assertions $t : X$ by $\mathbf{J}X$ where $\mathbf{J}$ stands for so-called *justified common knowledge* modality. Modality $\mathbf{J}$ is a stronger version of common knowledge: $\mathbf{J}X$ states all agents share sufficient evidence for X. In a formal setting, in Kripke models, $\mathbf{J}X \to \mathbf{C}X$, but not necessarily $\mathbf{C}X \to \mathbf{J}X$ (Artemov 2006).

Informally, the traditional common knowledge modality (Fagin, Halpern, Moses, and Vardi 1995) is represented by the condition

$$\mathbf{C}X \Leftrightarrow X \wedge EX \wedge E^2X \wedge \ldots \wedge E^nX \wedge \ldots$$

whereas for the justified common knowledge operator $\mathbf{J}$ one has

$$\mathbf{J}X \Rightarrow X \wedge EX \wedge E^2X \wedge \ldots \wedge E^nX \wedge \ldots$$

Justified common knowledge has the same modal principles as McCarthy's common knowledge (McCarthy, Sato, Hayashi, and Igarishi 1978). In (Cubitt and Sugden 2003) a case is made that David Lewis' version of common knowledge (more properly, belief) is not identified with unlimited iteration of knowledge operators, but is much closer to justified common knowledge. (See the encyclopedia article on Common Knowledge). A good example of such a Lewis-McCarthy-Artemov justified common knowledge assertion, $\mathbf{J}X$, which is stronger than the usual common knowledge, $\mathbf{C}X$, is provided by situations following a public announcement of X (Plaza 2007) after which X holds at all states, not only at reachable states. Note that public announcements are the usual means for attaining common knowledge, and they lead to justified common knowledge $\mathbf{J}$ rather than the usual common knowledge $\mathbf{C}$.

The axiomatic description of justified common knowledge $\mathbf{J}$ is significantly simpler than that of $\mathbf{C}$. According to (Antonakos 2007), in the standard epistemic scenarios, justified common knowledge $\mathbf{J}$ is conservative with respect to the usual common knowledge $\mathbf{C}$ and hence provides a lighter alternative to the latter.

6. Self-referentiality of Justifications

Let us consider an example which was suggested by the well-known *Moore's paradox*:

It will rain but I don't believe that it will.

If R stands for *it will rain*, then a modal formalization is:

$$M = R \wedge \neg\Box R.$$

The Moore sentence M is easily satisfiable, hence consistent, e.g., whenever the weather forecast wrongly shows "no rain". However, it is impossible to know Moore's sentence because

$$\neg\Box M = \neg\Box(R \wedge \neg\Box R)$$

holds in any modal logic containing T . Here is a derivation.

1. $(R \wedge \neg\Box R) \to R$ (logical axiom)
2. $\Box((R \wedge \neg\Box R) \to R)$ (Necessitation)
3. $\Box(R \wedge \neg\Box R) \to \Box R$, (from 2 by Distribution)
4. $\Box(R \wedge \neg\Box R) \to (R \wedge \neg\Box R)$ (Factivity in T)
5. $\Box(R \wedge \neg\Box R) \to \neg\Box R$ (from 4 in Boolean logic)
6. $\neg\Box(R \wedge \neg\Box R)$ (from 3 and 5 in Boolean logic)

Furthermore, here is how this derivation is realized in LP.

1. $(R \wedge \neg[c \cdot x] : R) \to R$ (logical axiom)
2. $c : ((R \wedge \neg[c \cdot x] : R) \to R)$ (Constant Specification)
3. $x : (R \wedge \neg[c \cdot x] : R) \to [c \cdot x] : R$ (from 2 by Application)
4. $x : (R \wedge \neg[c \cdot x] : R) \to (R \wedge \neg[c \cdot x] : R)$ (Factivity)
5. $x : (R \wedge \neg[c \cdot x] : R) \to \neg[c \cdot x] : R$ (from 4 by Boolean logic)
6. $\neg x : (R \wedge \neg[c \cdot x] : R)$ (from 3 and 5 in Boolean logic)

Note that Constant Specification in line 2 is self-referential.

Notes to Justification Logic

1. For better readability brackets '[', ']' will be used in terms, and parentheses '(', ')' in formulas. Both will be avoided when it is safe.

2. One could devise a formalization of the Red Barn Example in a bi-modal language with distinct modalities for knowledge and belief. However, it seems that such a resolution must involve reproducing justification-tracking arguments in a way that obscures, rather than reveals, the truth. Such a bi-modal formalization would distinguish $u : B$ from $[a \cdot v] : B$ not because they have different reasons (which reflects the true epistemic structure of the problem), but rather because the former is labelled 'belief' and the latter 'knowledge.' But what if one needs to keep track of a larger number of different unrelated reasons? By introducing a multiplicity of distinct modalities and then imposing various assumptions

governing the inter-relationships between these modalities, one would essentially end up with a reformulation of the language of Justification Logic itself (with distinct terms replaced by distinct modalities). This suggests that there may not be a satisfactory 'halfway point' between the modal language and the language of Justification Logic, at least inasmuch as one tries to capture the essential structure of examples involving the deductive nature of knowledge.

3. In our notation, LP can be assigned the name $\mathsf{JT4}$. However, in virtue of the fundamental role played by LP in the history of Justification Logic, the name LP has been preserved for this system.

4. To be precise, one must substitute c for x everywhere in s and t .

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

6. Which was false in 1912.

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

Notes to the Supplement

8. Kleene himself denied any connection of his realizability with the BHK interpretation.