Collegium Logicum

Justification Logic

Sergei Artemov & Elena Nogina
(in collaboration with Melvin Fitting)

Vienna, April 26, 2011

Why Justification Logic

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*.

Why Justification Logic

A paradigmatic illustration of the logical omniscience problem is a mathematician who knows the axioms of Peano Arithmetic PA but does not know all their logical consequences, i.e., all theorems of PA. Normally, the mathematician is interested in a specific easily formulated open problem *O*, say "there are infinitely many twin primes," about which she does not know whether it is true. Awareness models offer a dubious analysis by assuming that the mathematician is not aware of *O* which is plain false: the mathematician is very well aware of the problem, may be even obsessed with it, she just does not have a proof of it.

Awareness models' resolution does not reach the core of the matter: *logical omniscience is a complexity issue*.

What is 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.

What is Justification Logic

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.

What is Justification Logic

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. If t is a justification term and X is a formula, $t{:}X$ is a formula, and is intended to be read

$$t \textit{ is a justification for } X.$$

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

$$s{:}(A \rightarrow B) \rightarrow (t{:}A \rightarrow [s \cdot t]{:}B). \tag{1}$$

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

$$\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B). \tag{2}$$

Red Barn example

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."*

Red Barn example

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. $\Box(B \wedge R \rightarrow B)$, 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, $\Box(B \wedge R \rightarrow B)$ are cases of knowledge whereas $\Box B$ (line 1) is not knowledge. The modal language here does not seem to help resolving this issue.

Red Barn example

Next consider this scenario in Justification Logic where $t{:}F$ is interpreted as 'I **believe** F for reason t.' 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 is a red barn';
3. $a{:}(B \wedge R \rightarrow B)$, '$a$ is some default justification for the conjunction axiom.'

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

4. $a{:}(B \wedge R \rightarrow B) \rightarrow (v{:}(B \wedge R) \rightarrow [a \cdot v]{:}B)$, by principle (1);
5. $v{:}(B \wedge R) \rightarrow [a \cdot v]{:}B$, from 3 and 4, by propositional logic;
6. $[a \cdot v]{:}B$, from 2 and 5, by propositional 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 $[a \cdot v]{:}B$. The justification logic formalization represents the situation fairly.

Sum of justifications

The information-preserving property of justifications is expressed using the operation *sum* '+.' 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.

$$s{:}F \rightarrow [s+t]{:}F \quad \text{and} \quad t{:}F \rightarrow [s+t]{:}F.$$

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.

Basic Justification Logic J

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 $s{:}(F \rightarrow G) \rightarrow (t{:}F \rightarrow [s \cdot t]{:}G)$,

Monotonicity Axioms $s{:}F \rightarrow [s + t]{:}F$, $s{:}F \rightarrow [t + s]{:}F$.

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

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

Basic Justification Logic J

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

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

where A is an axiom of $\mathcal{L}$, and $e_1, e_2, \ldots, e_n$ are similar constants with indices $1, 2, \ldots, n$. 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.

Basic Justification Logic J

Logic of Justifications with given Constant Specification Let CS be a constant specification. J_{CS} is the logic $\mathsf{J}_0 + CS$; 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 $\mathsf{J}_\emptyset$.

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}.

Internalization

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:*

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

Proof. Induction on derivation length. Suppose $\vdash F$. If F is 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 $X \rightarrow F$ and X, then, by the Induction Hypothesis, $\vdash s{:}(X \rightarrow F)$ and $\vdash t{:}X$ for some s, t. Using the Application Axiom, $\vdash [s \cdot t]{:}F$. $\square$

Basic Justification Logic J

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. $A \to (A \vee B)$, *by 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 Monotonicity*
8. $[b \cdot y]{:}(A \vee B) \to [a \cdot x + b \cdot y]{:}(A \vee B)$, *by Monotonicity*
9. $(x{:}A \vee y{:}B) \to [a \cdot x + b \cdot y]{:}(A \vee B)$ *from 3, 6, 7, 8.*

The complete reading of the result of this derivation is

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

Further principles (optional)

Factivity Axiom $t{:}F \to F$.

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

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

$\mathsf{JT} = \mathsf{J} + \text{Factivity}.$

$\mathsf{J4} = \mathsf{J} + \text{Positive Introspection},$

$\mathsf{LP} = \mathsf{JT} + \text{Positive Introspection}.$

$\mathsf{J45} = \mathsf{J4} + \text{Negative Introspection},$

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

$\mathsf{JT45} = \mathsf{J45} + \text{Factivity},$

Evidence Function

A *Fitting model* is a structure $\mathcal{M} = \langle \mathcal{G}, \mathcal{R}, \mathcal{E}, \mathcal{V} \rangle$. Of this, $\langle \mathcal{G}, \mathcal{R}, \mathcal{V} \rangle$ is a standard K model. 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.

Given a model $\mathcal{M} = \langle \mathcal{G}, \mathcal{R}, \mathcal{E}, \mathcal{V} \rangle$, truth of formula X at possible world Γ is required to meet the standard Boolean conditions. The key item is

$$\Gamma \Vdash (t{:}X) \textit{ if and only if } \Gamma \in \mathcal{E}(t, X) \textit{ and, for every } \Delta \in \mathcal{G},\ \Delta \Vdash X.$$

Informally, $t{:}X$ is true at a 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.

Singleton models

Mkrtychev models were developed considerably before Fitting models, (Mkrtychev 1997). Today they can most simply be thought of as Fitting models with 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 Mkrtychev models, though of course this was not the original argument. What completeness with respect to Mkrtychev models tells us is that information about the possible world structure of Fitting models can be completely encoded by the admissible evidence function, at least for the logics discussed so far. Mkrtychev used these 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 2009).

Forgetful projection

Forgetful projection 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.

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:

$$\mathsf{J}^o = \mathsf{K}.$$

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 K, T, $\mathsf{K4}$, $\mathsf{S4}$, $\mathsf{K45}$, $\mathsf{S5}$ and some others have exact Justification Logic counterparts.

Realization Theorem

The Correspondence Theorem is based on the **Realization Theorem**:

> *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 (K, T, K4, S4). Alternatively, the Realization Theorem can be established semantically (K45, S5).

Realization Theorem

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, contains the *Löb Principle*

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

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 the Realization Theorem 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 (to be precise, one must substitute c for x everywhere in s and t).

Justified Belief and Knowledge

Justification Logic 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.

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. This new 'justification' 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.

Justification semantics plays a similar role in Modal Logic to that played by Kleene realizability in Intuitionistic Logic. It took Kleene realizability to reveal the computational semantics of Intuitionistic Logic and the Logic of Proofs to provide the semantics of proofs for Intuitionistic and Modal Logic.

Knowledge + justifications

The most common joint logic of explicit and implicit knowledge is S4LP (Artemov and Nogina 2005). The language of 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 S4 for the implicit operator, together with a connecting axiom, $t{:}X \to \Box X$, anything that has an explicit justification is knowable.

Semantically, Fitting models for LP need no modification, since they already have all the machinery of Hintikka/Kripke models. While the logic 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. Realization of implicit knowledge modalities in S4LP by explicit justifications which would respect the epistemic structure remains a major challenge in this area.

Tracking evidence

Justification Logic can analyize different justifications for the same fact. Consider a well-known example from (Russell 1912).

> *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(which was true back in 1912). But if he believes that Mr. Balfour was the late Prime Minister (which was false in 1912), he will still believe that the late Prime Minister's last name began with a 'B,' yet this belief, though true, would not be thought to constitute knowledge.*

Here 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:

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

Tracking evidence

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

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

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

However, this derivation utilizes the fact that r is a factive justification for B to conclude $w{:}B \to B$, which constitutes 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 truth-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 $\mathcal{S}$. This is easy to establish by choosing an appropriate model.

Moore sentences

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*

Self-referential justifications

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

1. $(R \wedge \neg[c \cdot x]{:}R) \rightarrow R$, *logical axiom*
2. $c{:}((R \wedge \neg[c \cdot x]{:}R) \rightarrow R)$, *Constant Specification*
3. $x{:}(R \wedge \neg[c \cdot x]{:}R) \rightarrow [c \cdot x]{:}R$, *from 2, by Application*
4. $x{:}(R \wedge \neg[c \cdot x]{:}R) \rightarrow (R \wedge \neg[c \cdot x]{:}R)$, *Factivity*
5. $x{:}(R \wedge \neg[c \cdot x]{:}R) \rightarrow \neg[c \cdot x]{:}R$, *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, i.e., contains a justification assertion $c{:}A(c)$.

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. The question of whether or not modal logics can be realized without using self-referential justifications was a major open question in this area.

Self-referentiality in general

The principal result by Kuznets states that self-referentiality of justifications is unavoidable in realization of S4 in LP. The current state of things is given by the following theorem due to Kuznets:

> *Self-referentiality can be avoided in realizations of modal logics* K *and* D. *Self-referentiality cannot be avoided in realizations of modal logics* T, K4, D4 *and* S4.

This theorem establishes that a system of justification terms for S4 will necessarily be self-referential. This creates a serious, though not directly visible, constraint on provability semantics. In the Logic of Proofs LP it was dealt with by a non-trivial fixed-point construction.

Logical Omniscience Test

The logical omniscience feature assumes that an epistemic agent knows all logical consequences of his assumptions. Justification Logic offers a general theoretical framework that views logical omniscience as a computational complexity problem. Artemov & Kuznets suggested the following approach: we assume that the knowledge of an agent is represented by an epistemic logical system E; we call such an agent *not logically omniscient* if for any valid knowledge assertion $\mathcal{A}$ of type F *is known*, a proof of F in E can be found in polynomial time in the size of $\mathcal{A}$. A.& K. showed that agents represented by major modal logics of knowledge and belief are logically omniscient whereas agents represented by justification logic systems are not logically omniscient with respect to t *is a justification for* F.

Justified Common Knowledge

Consider n agents with a commonly trusted evidence system. Its forgetful projection defines *justified common knowledge* modality $\mathbf{J}$ stronger than common knowledge: $\mathbf{J}X$ states that agents share sufficient evidence for X.

The common knowledge modality is represented by the condition

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

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

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

Justified common knowledge has the same modal principles as McCarthy's common knowledge. David Lewis' version of common knowledge is also more close to justified common knowledge. A public announcement of X after which X holds at all states, not only at reachable states, yields $\mathbf{J}X$, not $\mathbf{C}X$.

The axiomatic description of $\mathbf{J}$ is significantly simpler than that of $\mathbf{C}$. Moreover, in the standard epistemic scenarios, $\mathbf{J}$ is conservative with respect to $\mathbf{C}$ and hence provides a lighter alternative to the latter.

Quantification and LP

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.

> *The first-order logic of proofs is not recursively enumerable* (Artemov & Yavorskaya, 2001. *The logic of proofs with quantifiers over proofs is not recursively enumerable* (Yavorsky 2001).

Earlier this year, Artemov & Yavorskaya found the first-order logic of proofs FOLP capable of realizing first-order modal logic FOS4 and, therefore, the first-order intuitionistic logic HPC. Two kinds of proof semantics for FOLP have been offered: *parametric semantics*, in which proof objects are interpreted as derivations with parameters, and *generic semantics* with proof terms interpreted as provably computable functions from parameters to formal derivations. Both provide semantics of proofs for first-order S4 and a first-order Brouwer-Heyting-Kolmogorov-style semantics for HPC.

FOS4 may be viewed as a general purpose first-order justification logic; it opens the door to a general theory of first-order justification.

First-order LP: format

In the language FOLP, the proof predicate is represented by formulas of the form

$$t{:}_X A$$

where X is the set of individual variables that are considered global parameters. Variables from X and only them are free in $t{:}_X A$. 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 $t{:}_X A$.

Proofs are represented by proof terms which do not contain individual variables. An arithmetical interpretation $*$, commutes with the Boolean connectives and quantifiers and

$$(t{:}_X F)^* = \mathit{Prof}\,(t^*(\underline{X}), F^*(\underline{X})),$$

i.e., $(t{:}_X F)^*$ is evaluated by the natural arithmetical formula asserting that t is a proof of F with global variables X.

First-order LP: axioms

FOLP is axiomatized by the following schemas. Here A, B are formulas, s, t are terms, X is a set of individual variables, and y is an individual variable.

A1 classical axioms of first-order logic

A2 $t{:}_{Xy}A \rightarrow t{:}_X A$, $\quad y$ is not free in A

A3 $t{:}_X A \rightarrow t{:}_{Xy}A$

B1 $t{:}_X A \rightarrow A$

B2 $s{:}_X(A \rightarrow B) \wedge t{:}_X A \rightarrow (s \cdot t){:}_X B$

B3 $t{:}_X A \rightarrow (t+s){:}_X A$, $\quad s{:}_X A \rightarrow (t+s){:}_X A$

B4 $t{:}_X A \rightarrow\, !t{:}_X t{:}_X A$

B5 $t{:}_X A \rightarrow \Box_x(t){:}_X \forall x A$, $x \notin X$

FOLP has the following inference rules:

R1 $\vdash A, A \rightarrow B \;\rightarrow\; \vdash B$ *Modus Ponens*

R2 $\vdash A \;\rightarrow\; \vdash \forall x A$ *generalization*

R3 $\vdash c{:}A$, where A is an axiom, c is a proof constant *axiom necessitation.*

First-order LP: realization

We derive an explicit version of the converse Barcan Formula $\Box\forall x A \to \forall x \Box A$.

1. $\forall x A \to A$ - logical axiom;
2. $c{:}(\forall x A \to A)$ - axiom necessitation;
3. $c{:}_{\{x\}}(\forall x A \to A)$ - from 2, by axiom A3;
4. $c{:}_{\{x\}}(\forall x A \to A) \to (u{:}_{\{x\}}\forall x A \to (c\cdot u){:}_{\{x\}}A)$ - axiom B2;
5. $u{:}_{\{x\}}\forall x A \to (c\cdot u){:}_{\{x\}}A$ - from 3, 4, by Modus Ponens;
6. $u{:}\forall x A \to u{:}_{\{x\}}\forall x A$ - by axiom A3;
7. $u{:}\forall x A \to (c\cdot u){:}_{\{x\}}A$ - from 5, 6;
8. $\forall x[u{:}\forall x A \to (c\cdot u){:}_{\{x\}}A]$ - from 7, by generalization;
9. $u{:}\forall x A \to \forall x(c\cdot u){:}_{\{x\}}A$ - since x is not free in the antecedent of 8.

Realization Theorem. *If* $\mathsf{FOS4} \vdash A$*, then there is a normal realization* A^r *such that* $\mathsf{FOLP} \vdash A^r$.

Since HPC can be faithfully embedded in $\mathsf{FOS4}$ via Gödel's transation, this also provides HPC with a BHK-style semantics of proofs.

Conclusions

Justification Logic extends the logic of knowledge by a formal theory of justification. It is capable of formalizing a significant portion of reasoning about justifications, e.g., Kripke, Russell, and Gettier examples. This formalization has been used for the resolution of paradoxes, verification, hidden assumption analysis, and eliminating redundancies.

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

- intended provability semantics for Gödel's provability logic $\mathsf{S4}$ and $\mathsf{FOS4}$, a formalization of Brouwer-Heyting-Kolmogorov semantics;
- a rigorous definition of the Logical Omniscience property and a demonstration that evidence assertions in Justification Logic are not logically omniscient;
- a more general approach to common knowledge;
- evidence tracking framework (in progress).

Thank You!