WORKSHOP ON RECENT TRENDS IN PROOF THEORY, Bern, July 9-11, 2008

On the unusual effectiveness of proof theory in epistemology

Sergei Artemov

The City University of New York

Bern, July 11, 2008

This is a report about the extending the ideas and methods of Proof Theory to a new, promising area. It seems that this is a kind of development in which Proof Theory might be interested.

This is a report about the extending the ideas and methods of Proof Theory to a new, promising area. It seems that this is a kind of development in which Proof Theory might be interested.

Similar stories about what the Logic of Proofs brings to foundations, constructive semantics, combinatory logic and lambda-calculi, theory of verification, cryptography, etc., lie mostly outside the scope of this talk.

Mainstream Epistemology:
tripartite approach to knowledge (usually attributed to Plato)

Knowledge $\sim$ *Justified True Belief.*

A core topic in Epistemology, especially in the wake of papers by Russell, Gettier, and others: questioned, criticized, revised; now is generally regarded as a necessary condition for knowledge.

Logic of Knowledge: the model-theoretic approach (Kripke, Hintikka, ...) has dominated modal logic and formal epistemology since the 1960s.

$\Box F \quad \sim \quad F$ *holds at all possible worlds (situations).*

Logic of Knowledge: the model-theoretic approach (Kripke, Hintikka, ...) has dominated modal logic and formal epistemology since the 1960s.

$$\Box F \quad \sim \quad F \text{ holds at all possible worlds (situations).}$$

Easy, visual, useful in many cases, but misses the mark considerably: *What if F holds at all possible worlds, e.g., a mathematical truth, say $P \neq NP$, but the agent is simply not aware of the fact due to lack of evidence, proof, justification, etc.?*

Logic of Knowledge: the model-theoretic approach (Kripke, Hintikka, ...) has dominated modal logic and formal epistemology since the 1960s.

$$\Box F \quad \sim \quad F \text{ holds at all possible worlds (situations).}$$

Easy, visual, useful in many cases, but misses the mark considerably: *What if F holds at all possible worlds, e.g., a mathematical truth, say $P \neq NP$, but the agent is simply not aware of the fact due to lack of evidence, proof, justification, etc.?*

Speaking informally: modal logic offers a limited formalization

$$\textit{Knowledge} \quad \sim \quad \textit{True Belief.}$$

There were no justifications in the modal logic of knowledge, hence a principal gap between mainstream and formal epistemology.

Obvious defect: Logical Omniscience

A basic principle of modal logic (of knowledge, belief, etc.):

$$\Box(F \rightarrow G) \rightarrow (\Box F \rightarrow \Box G).$$

At each world, the agent is supposed to "know" all logical consequences of his/her assumptions.

Obvious defect: Logical Omniscience

A basic principle of modal logic (of knowledge, belief, etc.):

$$\Box(F \rightarrow G) \rightarrow (\Box F \rightarrow \Box G).$$

At each world, the agent is supposed to "know" all logical consequences of his/her assumptions.

"Each agent who knows the rules of Chess should know whether there is a winning strategy for White."

Obvious defect: Logical Omniscience

A basic principle of modal logic (of knowledge, belief, etc.):

$$\Box(F \to G) \to (\Box F \to \Box G).$$

At each world, the agent is supposed to "know" all logical consequences of his/her assumptions.

"Each agent who knows the rules of Chess should know whether there is a winning strategy for White."

"Suppose one knows a product of two (very large) primes. In what sense does he/she know each of the primes, given that factorization may take billions of years of computation?"

Awareness models (Fagin & Halpern)

$\Box F$ $\sim$ F *holds at all possible worlds and the agent is aware of* F.

Awareness models (Fagin & Halpern)

$\Box F$ $\sim$ *F holds at all possible worlds and the agent is aware of F.*

Without closure conditions, the awareness function exhibits weird behavior, e.g., the agent could be aware of $F \wedge F$ and not aware of F. Straightforward closure conditions bring Logical Omniscience back to the model.

Adding a proof-theoretical component:

Proofs	$\Rightarrow$	*Justifications*
Proofs of assumptions	$\Rightarrow$	*Constants*
Proofs of hypotheses	$\Rightarrow$	*Variables*
Proofs without hypotheses	$\Rightarrow$	*Ground justifications*
Proofs with hypotheses	$\Rightarrow$	*Justifications with variables*
Rules	$\Rightarrow$	*Operations on justifications*
Cut elimination	$\Rightarrow$	*Recovery of explicit knowledge from modal derivations*

What does this approach bring to Epistemology?

- mathematical theory of justification where justification logic systems enjoy normal closure properties and clean epistemic semantics, soundness and completeness theorems

- richer language for dealing with Knowledge, Belief, Evidence

- old/new evidence-based semantics for modal epistemic logic

- natural handling of Logical Omniscience

- new mathematics and appealing connections to other fields

- etc.

The basics of Justification Logic

Brouwer-Heyting-Kolmogorov (*BHK*) *semantics*:

- a proof of $A \wedge B$ consists of a proof of A and a proof of B,

- a proof of $A \vee B$ is given by presenting either a proof of A or a proof of B,

- a proof of $A \rightarrow B$ is a construction transforming proofs of A into proofs of B

Gödel's modal logic of provability

Gödel (1933) introduced the modal logic S4 as the system for axiomatizing classical provability:

Axioms and rules of classical propositional logic

$\Box(F \rightarrow G) \rightarrow (\Box F \rightarrow \Box G)$ *Normality*

$\Box F \rightarrow F$ *Reflexivity*

$\Box F \rightarrow \Box\Box F$ *Transitivity*

Necessitation Rule: $\dfrac{\vdash F}{\vdash \Box F}$.

Based on Brouwer's understanding of logical truth as provability, Gödel defined a translation $tr(F)$ of the propositional formula F in the intuitionistic language into the language of classical modal logic:

tr(F) = *prefix every subformula of* F *with the provability modality* $\Box$.

Informally speaking, when the usual procedure of determining the classical truth of a formula is applied to $tr(F)$, 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 $tr(F)$ provides a proper embedding of the intuitionistic logic IPC into S4, i.e., an embedding of IPC into classical logic extended by the provability operator.

Theorem [Gödel, McKinsey & Tarski, 1933-1948]:

$$\text{IPC } \textit{proves } F \quad \Leftrightarrow \quad \text{S4 } \textit{proves } tr(F).$$

Still, Gödel's original goal of defining IPC in terms of classical provability was not reached because the connection of S4 to the usual mathematical notion of provability was not established.

Gödel considered the (straightforward) interpretation of $\Box F$ to be

F is provable in Peano Arithmetic PA

and noticed that this semantics is inconsistent with S4. Indeed, S4 $\vdash \Box(\Box F \rightarrow F)$, and when $\Box$ = *Provable* = formal provability predicate in Peano Arithmetic PA, and F as $\bot$, this formula becomes false:

Provable (Consis PA) .

The situation following Gödel's paper of 1933 can be described by the scheme below in which '$\hookrightarrow$' denotes a 'provability' embedding:

$$\text{IPC} \hookrightarrow \text{S4} \hookrightarrow \ ? \ \hookrightarrow \text{PA} .$$

Alternative Gödel's format for provability

In his Vienna lecture of 1938, Gödel mentioned the possibility of building an explicit version of S4 with the proposition

$$t \text{ is a proof of } F$$

interpreted via **the proof predicate in** PA:

$$\mathit{Proof}(t, F).$$

This lecture remained unpublished until 1995. By that time, the full Logic of Proofs LP had already been discovered by the author.

Principal observation:
For each specific derivation p, PA $\vdash$ *Proof* $(p, F) \rightarrow F$.

Indeed,
• If *Proof* (p, F) holds, then F is evidently provable in PA, and so is the formula *Proof* $(p, F) \rightarrow F$.

• If $\neg$*Proof* (p, F) holds, then it is provable in PA (since *Proof* (x, y) is decidable) and *Proof* $(p, F) \rightarrow F$ is again provable.

Logic of Proofs LP: the language

Proof polynomials are terms built from *proof variables* $x, y, z, \ldots$ and *proof constants* $a, b, c, \ldots$ by means of two binary operations: *application* '$\cdot$' and *choice* '$+$', and one unary *proof checker* '!'.

Using t to denote any proof polynomial and S any sentence letter, the **formulas of the Logic of Proofs** are defined by the grammar

$$A = S \mid A \rightarrow A \mid A \wedge A \mid A \vee A \mid \neg A \mid t{:}A \ .$$

Logic of Proofs LP

The standard axioms and rules of classical propositional logic

$t{:}(F \to G) \ \to \ (s{:}F \to (t \cdot s){:}G)$ *Application*

$t{:}F \ \to\ !t{:}(t{:}F)$ *Proof Checker*

$s{:}F \to (s+t){:}F, \quad t{:}F \to (s+t){:}F$ *Choice*

$t{:}F \to F$ *Reflexivity*

$\vdash c{:}A$, where A is an axiom
and c is a proof constant *Constant Specification Rule*

LP$_{CS}$ is LP with the Constant Specification Rule replaced by a set of axioms

$$CS \subseteq \{c{:}A \mid A \text{ is an axiom and } c \text{ is a proof constant}\}$$

One of the basic properties of LP is its capability of internalizing its own derivations:

if $$A_1, \ldots, A_n \vdash F,$$

then for some proof polynomial $t(x_1, \ldots, x_n)$,

$$x_1{:}A_1, \ldots, x_n{:}A_n \vdash t(x_1, \ldots, x_n){:}F.$$

One of the basic properties of LP is its capability of internalizing its own derivations:

if $$A_1, \ldots, A_n \vdash F,$$

then for some proof polynomial $t(x_1, \ldots, x_n)$,

$$x_1{:}A_1, \ldots, x_n{:}A_n \vdash t(x_1, \ldots, x_n){:}F.$$

Note that the Curry-Howard isomorphism covers only a simple instance of the internalization property where all of $A_1, \ldots, A_n, B$ are purely propositional formulas containing no proof terms.

Internalization, the general form:

if $$\Gamma, \vec{y}{:}\Delta \vdash F,$$

then for some proof polynomial $t(\vec{x}, \vec{y})$,

$$\vec{x}{:}\Gamma, \vec{y}{:}\Delta \vdash t(\vec{x}, \vec{y}){:}F.$$

Realization of S4 in the Logic of Proofs LP

S4 is the forgetful projection of LP, i.e.,

1. *The forgetful projection of LP is S4-compliant.*

2. *For each theorem F of S4, one can recover a witness (proof polynomial) for each occurrence of $\Box$ in F in such a way that the resulting formula F^r is derivable in LP.*

Realization gives a semantics of proofs for S4.

$$\mathsf{S4} \vdash F \quad \Leftrightarrow \quad \exists r\ \mathsf{LP} \vdash F^r$$

Part (1) of the Realization theorem is straightforward. Part (2) is not so easy. Let us try the 'naive' approach: induction on a given derivation in S4. Realization of S4 axioms is trivial. Necessitation is covered by Internalization. Modus Ponens?

$$\frac{A \rightarrow B \qquad A}{B}$$

By I.H., the premises are realizable. Therefore, in LP,

$$\frac{A^r \rightarrow B^r \qquad A^r}{B^r}\ .$$

What is wrong with this "proof?"

Part (1) of the Realization theorem is straightforward. Part (2) is not so easy. Let us try the 'naive' approach: induction on a given derivation in S4. Realization of S4 axioms is trivial. Necessitation is covered by Internalization. Modus Ponens?

$$\frac{A \rightarrow B \qquad A}{B}$$

By I.H., the premises are realizable. Therefore, in LP,

$$\frac{A^r \rightarrow B^r \qquad A^r}{B^r}\ .$$

What is wrong with this "proof?"

These r's depend on a derivation, hence the A^r's may be different. How to reconcile/unify them?

Cut-elimination is the answer!

The Realization Algorithm (all versions known so far) work with cut-free proofs in S4.

We start with a cut-free S4-proof.

$$\dfrac{\dfrac{\dfrac{\dots}{\Box A \Rightarrow \Box A \lor B}}{\Box A \Rightarrow \Box(\Box A \lor B)} \qquad \dfrac{\dfrac{\dfrac{\dots}{B \Rightarrow \Box A \lor B}}{\Box B \Rightarrow \Box A \lor B}}{\Box B \Rightarrow \Box(\Box A \lor B)}}{\Box A \lor \Box B \Rightarrow \Box(\Box A \lor B)}$$

Then we fill in negative $\Box$'s with distinct variables.

$$\dfrac{\dfrac{\dfrac{\cdots}{x{:}A \Rightarrow \Box A \vee B}}{x{:}A \Rightarrow \Box(\Box A \vee B)} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\cdots}{B \Rightarrow \Box A \vee B}}{y{:}B \Rightarrow \Box A \vee B}}{y{:}B \Rightarrow \Box(\Box A \vee B)}}{}}{x{:}A \vee y{:}B \Rightarrow \Box(\Box A \vee B)}$$

Some positive $\Box$'s follow up easily.

$$
\dfrac{
\dfrac{\dfrac{\ldots}{x{:}A \Rightarrow x{:}A \vee B}}{x{:}A \Rightarrow \Box(x{:}A \vee B)}
\qquad
\dfrac{\dfrac{\dfrac{\ldots}{B \Rightarrow x{:}A \vee B}}{y{:}B \Rightarrow x{:}A \vee B}}{y{:}B \Rightarrow \Box(x{:}A \vee B)}
}{x{:}A \vee y{:}B \Rightarrow \Box(x{:}A \vee B)}
$$

All remaining $\Box$'s are special cases of Internalization.

$$\dfrac{\dfrac{\dfrac{\ldots}{x{:}A \Rightarrow x{:}A\vee B}}{x{:}A \Rightarrow \Box(x{:}A\vee B)} \qquad \dfrac{\dfrac{\dfrac{\ldots}{B \Rightarrow x{:}A\vee B}}{y{:}B \Rightarrow x{:}A\vee B}}{y{:}B \Rightarrow \Box(x{:}A\vee B)}}{x{:}A\vee y{:}B \Rightarrow \Box(x{:}A\vee B)}$$

We have to reconcile two different proof terms, hence $u_1 + u_2$.

$$\dfrac{\dfrac{\dfrac{\ldots}{x{:}A \Rightarrow x{:}A \vee B}}{x{:}A \Rightarrow [u_1 + u_2]{:}(x{:}A \vee B)} \qquad \dfrac{\dfrac{\dfrac{\ldots}{B \Rightarrow x{:}A \vee B}}{y{:}B \Rightarrow x{:}A \vee B}}{y{:}B \Rightarrow [u_1 + u_2]{:}(x{:}A \vee B)}}{x{:}A \vee y{:}B \Rightarrow [u_1 + u_2]{:}(x{:}A \vee B)}$$

It remains for us to specify u_1 and u_2, each at its node.

$$\dfrac{\dfrac{\dfrac{\cdots}{x{:}A \Rightarrow x{:}A \vee B}}{x{:}A \Rightarrow [u_1+u_2]{:}(x{:}A \vee B)} \qquad \dfrac{\dfrac{\dfrac{\cdots}{B \Rightarrow x{:}A \vee B}}{y{:}B \Rightarrow x{:}A \vee B}}{y{:}B \Rightarrow [u_1+u_2]{:}(x{:}A \vee B)}}{x{:}A \vee y{:}B \Rightarrow [u_1+u_2]{:}(x{:}A \vee B)}$$

First we find $u_1 := s(x,y)$ such that $\vdash x{:}A \to s(x,y){:}(x{:}A \vee B)$.

$$
\dfrac{\dfrac{\dfrac{\dots}{x{:}A \Rightarrow x{:}A \vee B}}{x{:}A \Rightarrow [u_1 + u_2]{:}(x{:}A \vee B)} \qquad \dfrac{\dfrac{\dfrac{\dots}{B \Rightarrow x{:}A \vee B}}{y{:}B \Rightarrow x{:}A \vee B}}{y{:}B \Rightarrow [u_1 + u_2]{:}(x{:}A \vee B)}}{x{:}A \vee y{:}B \Rightarrow [u_1 + u_2]{:}(x{:}A \vee B)}
$$

Finding such $s(x,y)$:

1. $x{:}A \rightarrow (x{:}A \vee B)$ - propositional axiom
2. $a{:}[x{:}A \rightarrow (x{:}A \vee B)]$ - specifying constant a
3. $x{:}A \rightarrow\, !x{:}x{:}A$ - proof checking axiom
4. $!x{:}x{:}A \rightarrow [a\cdot !x]{:}(x{:}A \vee B)$ - from 2, by application
5. $x{:}A \rightarrow [a\cdot !x]{:}(x{:}A \vee B)$ - from 3,4.

Hence $s(x,y) = a\cdot !x$ where $a{:}[x{:}A \rightarrow (x{:}A \vee B)]$.

Substitute u_1 for $a{\cdot}!x$:

$$\dfrac{\dfrac{\dfrac{\ldots}{x{:}A \Rightarrow x{:}A\vee B}}{x{:}A \Rightarrow [a{\cdot}!x+u_2]{:}(x{:}A\vee B)} \qquad \dfrac{\dfrac{\dfrac{\ldots}{B \Rightarrow x{:}A\vee B}}{y{:}B \Rightarrow x{:}A\vee B}}{y{:}B \Rightarrow [a{\cdot}!x+u_2]{:}(x{:}A\vee B)}}{x{:}A\vee y{:}B \Rightarrow [a{\cdot}!x+u_2]{:}(x{:}A\vee B)}$$

Now we look for $u_2 := t(x,y)$ such that $\vdash y{:}B \to t(x,y){:}(x{:}A \vee B)$.

$$
\dfrac{
\dfrac{\dfrac{\dots}{x{:}A \Rightarrow x{:}A \vee B}}{x{:}A \Rightarrow [a\cdot !x + u_2]{:}(x{:}A \vee B)}
\qquad
\dfrac{\dfrac{\dfrac{\dots}{B \Rightarrow x{:}A \vee B}}{y{:}B \Rightarrow x{:}A \vee B}}{y{:}B \Rightarrow [a\cdot !x + u_2]{:}(x{:}A \vee B)}
}{x{:}A \vee y{:}B \Rightarrow [a\cdot !x + u_2]{:}(x{:}A \vee B)}
$$

Finding such $t(x, y)$:

1. $B \rightarrow (x{:}A \vee B)$ - propositional axiom
2. $b{:}[B \rightarrow (x{:}A \vee B)]$ - specifying constant b
3. $y{:}B \rightarrow [b{\cdot}y]{:}(x{:}A \vee B)]$ - by application, from 2

So, $t(x, y)$ is $b{\cdot}y$ where $b{:}[B \rightarrow (x{:}A \vee B)]$.

The final step: a derivation in LP:

$$
\dfrac{
\dfrac{\dfrac{\ldots}{x{:}A \Rightarrow x{:}A\vee B}}{x{:}A \Rightarrow [a\cdot !x + b\cdot y]{:}(x{:}A\vee B)}
\qquad
\dfrac{\dfrac{\dfrac{\ldots}{B \Rightarrow x{:}A\vee B}}{y{:}B \Rightarrow x{:}A\vee B}}{y{:}B \Rightarrow [a\cdot !x + b\cdot y]{:}(x{:}A\vee B)}
}{x{:}A\vee y{:}B \Rightarrow [a\cdot !x + b\cdot y]{:}(x{:}A\vee B)}
$$

where $a{:}[x{:}A \rightarrow (x{:}A\vee B)]$ and $b{:}[B \rightarrow (x{:}A \vee B)]$.

The original algorithm was exponential.

A polynomial realization algorithm was suggested by Brezhnev & Kuznets. It produces proof polynomials of at most quadratic size in the length of the given cut-free derivation in S4.

Provability semantics of LP

Interpretations respect Boolean connectives and

$$(p{:}F)^* \quad = \quad \mathit{Proof}\,(p^*, F^*).$$

Completeness theorem (A):

LP *derives all valid logical principles in its language.*

Some foundational consequences

Completing Gödel's effort concerning the logic of provability and BHK semantics:

$$\mathsf{IPC} \hookrightarrow \mathsf{S4} \hookrightarrow \mathsf{LP} \hookrightarrow \mathsf{PA}\ .$$

Foundational consequences: existential semantics of modality.

Models		Proofs
$\Gamma \models F$	$\Leftrightarrow$	$\Gamma \vdash F$
$\forall$ models ...		$\exists$ a proof ...
$\forall$-semantics		$\exists$-semantics
Kripke semantics		???
$\forall$ possible worlds ...		???

There has been a model-theoretic Kripke ($\forall$) semantics of $\Box$, but there were no proof-theoretic ($\exists$) readings of $\Box$ (cf. Gödel's provability interpretation).

Foundational consequences: existential semantics of modality.

Models		Proofs
$\Gamma \models F$	$\Leftrightarrow$	$\Gamma \vdash F$
$\forall$ models ...		$\exists$ a proof ...
$\forall$-semantics		$\exists$-semantics
Kripke semantics		Realizability
$\forall$ possible worlds ...		$\exists$ realization ...

$\forall$-semantics fits to modeling computational processes.
$\exists$-semantics is natural for epistemic/provability situations, which is what Epistemology seems to need...

From Proofs to Justifications

Explicit analogues (LP-style) of all major epistemic modal logics: K, K4, KD4, T, S5, etc.

Basic Justification Logic J = the explicit counterpart of K:

The standard axioms and rules of classical propositional logic

$t{:}(F \rightarrow G) \ \rightarrow \ (s{:}F \rightarrow (t\cdot s){:}G)$ *Application*

$s{:}F \rightarrow (s+t){:}F, \quad t{:}F \rightarrow (s+t){:}F$ *Choice*

Constant Specification (ranging from $\emptyset$ to the total CS).

Justifications are not necessarily factive, positive introspection/proof checker is not assumed.

Epistemic models for J (Fitting-style)

Model is $(W, R, \mathcal{E}, \Vdash)$, where

• (W, R) is a K-frame;

• $\mathcal{E}$ is an evidence function: for each term t and formula F, $\mathcal{E}(t, F)$ is a set of $u \in W$ where t is a possible evidence for F. An evidence function has natural closure properties that agree with operations of J, i.e.

$$\mathcal{E}(s, F) \subseteq \mathcal{E}(s+t, F) \cap \mathcal{E}(t+s, F)$$
$$\mathcal{E}(s, F \to G) \cap \mathcal{E}(t, F) \subseteq \mathcal{E}(s \cdot t, G);$$

• $u \Vdash t{:}F$ iff $u \in \mathcal{E}(t, F)$ and $v \Vdash F$ whenever uRv.

Justification Logic J is capable of formalizing paradigmatic epistemic examples involving justifications: Gettier, Russell's prime minister example, Kripke's red barn example, etc.

Gettier example

Smith has applied for a job, but has a justified belief that 'Jones will get the job.' He also has a justified belief that 'Jones has 10 coins in his pocket.' Smith therefore (justifiably) concludes ... that 'the man who will get the job has 10 coins in his pocket.'

In fact, Jones does not get the job. Instead, Smith does. However, as it happens, Smith also has 10 coins in his pocket. So his belief that 'the man who will get the job has 10 coins in his pocket' was justified and true. But it does not appear to be knowledge.

Formalizing the data

JJ = *Jones gets the job*

SJ = *Smith gets the job*

JC = *Jones has 10 coins in his pocket*

SC = *Smith has 10 coins in his pocket*

x = *whatever evidence Smith had about* JJ

y = *whatever evidence Smith had about* JC

Explicitly made assumptions:

1. $x{:}JJ$ (x is a justification of 'Jones gets the job')
2. $y{:}JC$ (y is a justification of 'Jones has 10 coins in his pocket')
3. $\neg JJ$ (Jones does not get the job)
4. SJ (Smith gets the job)
5. SC (Smith has 10 coins in his pocket)

Justification Logic methods show that these assumptions are not sufficient to derive Gettiers conclusion *Smith is justified in believing that 'the man who will get the job has 10 coins in his pocket.'*

In this setting, the sentence 'the man who will get the job has 10 coins in his pocket' can be represented by the formula

$$(JJ \rightarrow JC) \wedge (SJ \rightarrow SC).$$

No justified knowledge assertion for this formula, i.e.,

$$t{:}[(JJ \rightarrow JC) \wedge (SJ \rightarrow SC)],$$

is derivable from the assumptions $x{:}JJ$, $y{:}JC$, $\neg JJ$, SJ, SC.

Countermodel for Gettier's claim

$W = \{1,2\}$, $R = \{(1,2)\}$, $\mathcal{E}$ is total, i.e., $\mathcal{E}(t,F) = W$ for each t,F.

$$\begin{array}{rcl} \text{'belief world' } 2 & \bullet & JJ,\ JC,\ SJ,\ \neg SC \\ & \uparrow & \\ \text{'real world' } 1 & \bullet & \neg JJ,\ JC,\ SJ,\ SC \end{array}$$

All assumptions hold at 1,2.

Furthermore, $2 \not\Vdash (JJ \rightarrow JC) \wedge (SJ \rightarrow SC)$, hence

$$1 \not\Vdash t{:}[(JJ \rightarrow JC) \wedge (SJ \rightarrow SC)]$$

for each t.

Augmented set of assumptions

It is now easy to spot a missing assumption:

Jones and Smith cannot both have this job

(which is, of course, a default here). When it is added explicitly, the formal reasoning goes smoothly.

6. $z{:}(JJ \rightarrow \neg SJ)$ (z is a justification of 'Jones and Smith cannot both have the job')

Derivation of Gettier's claim

7.$(z \cdot x){:}(\neg SJ)$, from 1,6, by Application

8. $p{:}[\neg SJ \rightarrow (SJ \rightarrow SC)]$, Internalization of a tautology

9. $(z \cdot x){:}(\neg SJ) \rightarrow (p \cdot (z \cdot x)){:}(SJ \rightarrow SC)$, by Application

10. $(p \cdot (z \cdot x)){:}(SJ \rightarrow SC)$, from 7,9, by Modus Ponens

11. $c{:}[JC \rightarrow (JJ \rightarrow JC)]$, by Internalization

12 $y{:}JC \rightarrow (c \cdot y){:}(JJ \rightarrow JC)$, by Application

13. $(c \cdot y){:}(JJ \rightarrow JC)$, from 2,12, by Modus Ponens

14. $t{:}[(JJ \rightarrow JC) \wedge (SJ \rightarrow SC)]$, for an appropriate t, from 10 and 13

Metatheory of the Gettier example

Missing assumption analysis has just been performed.

We can also eliminate redundancies: no coins/pockets are needed...

Red Barn Example (Goldman – Kripke)

There are a number of fake barns or facades of barns in a certain locality. In the midst of these fake barns is one real barn, which is painted red and no fake barns are painted red. Jones is driving along the highway, looks around and happens to see the real barn, and so forms the belief *I see a barn*. Though Jones has gotten lucky, he could just as easily have been deceived and not known it. So this is not knowledge.

An alternate example: Jones looks around and forms the belief *I see a red barn*. This is knowledge, since Jones couldn't have been wrong, as fake barns cannot be painted red. This is a troubling account however, because it seems that the first statement, *I see a barn*, can be inferred from *I see a red barn.*

Formalization of *RBE* in epistemic modal logic

B - 'the object which the agent sees is a barn'

R - 'the object which the agent sees is red'

$\Box$ is the belief modality of Jones.

1. $\Box B$
2. $\Box(B \wedge R)$
3. $\Box(B \wedge R) \rightarrow (B \wedge R)$

Case (2) is knowledge, whereas (1) is not knowledge, by the problem's description. On the other hand, (1) logically follows from (2) in any epistemic modal logic:

$(B \wedge R) \rightarrow B$, logical axiom

$\Box[(B \wedge R) \rightarrow B]$, by Necessitation

$\Box(B \wedge R) \rightarrow \Box B$, by the Normality axiom.

This is a paradox, which is faithfully reproduced in modal logic.

Justification Logic provides a clean resolution of this paradox

Let us use the language of explicit justifications here.

Assumptions:

1. $u{:}B$.
2. $v{:}(B \wedge R)$
3. $v{:}(B \wedge R) \rightarrow (B \wedge R)$.

Reasoning:

4. $(B \wedge R) \rightarrow B$, logical axiom
5. $a{:}[(B \wedge R) \rightarrow B]$, Constant Specification
6. $v{:}(B \wedge R) \rightarrow (a \cdot v){:}B$, by Application.

Justification Logic provides a clean resolution of this paradox

Let us use the language of explicit justifications here.

Assumptions:

1. $u{:}B$.
2. $v{:}(B \wedge R)$
3. $v{:}(B \wedge R) \rightarrow (B \wedge R)$.

Reasoning:

4. $(B \wedge R) \rightarrow B$, logical axiom
5. $a{:}[(B \wedge R) \rightarrow B]$, Constant Specification
6. $v{:}(B \wedge R) \rightarrow (a \cdot v){:}B$, by Application.

'Paradox' disappears! Instead of deriving (1) from (2), we have obtained the correct conclusion that $(a \cdot v){:}B$, i.e., Jones knows B for reason $a \cdot v$, NOT u. Hence, after observing a red facade, Jones indeed knows B but this knowledge has nothing to do with (1), which remains a case of belief rather then knowledge.

Justification Logic provides a clean resolution of this paradox

Let us use the language of explicit justifications here.

Assumptions:

1. $u{:}B$.
2. $v{:}(B \wedge R)$
3. $v{:}(B \wedge R) \rightarrow (B \wedge R)$.

Model where (2) and (3) hold but not (1).

$W = \{a\}$, $R = \emptyset$, $\mathcal{E}$ at a is the minimal admissible evidence function containing total constant specification and $(v, B \wedge R)$. B and R are both true at a.

‘real world’ a $\bullet$ B, R

Apparently, $\mathcal{E}(u, B)$ does not hold and $a \Vdash$ (2), (3), but not (1).

Towards ‘truth tracking’

This example shows that we can perform some ‘truth tracking’ using explicit justifications and other machinery of Justification Logic.

Triple action of Proof Theory in Epistemology

1. Proofs as justifications.

2. Cut-elimination for realization theorems.

3. Truth tracking.