Epistemic Logic XII

A Logic of knowing why

Yanjing Wang
Department of Philosophy, Peking University

Dec. 16th, 2020

Question why

A logic of knowing why

Future work and Conclusions

Question why

We always want to know why

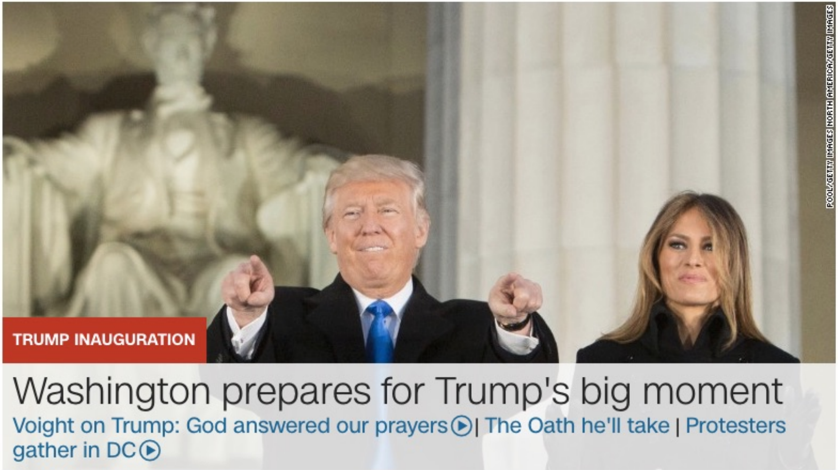
Especially for kids...

I want to know why …

- the window is broken.
- the lump of potassium dissolves in water.
- he stayed in the bookstore all those days.
- cheetahs can run at high speeds.
- blood circulates in the body.

Why-questions drive our civilization forward.

Why?


TRUMP INAUGURATION
Washington prepares for Trump's big moment
Voight on Trump: God answered our prayers | The Oath he'll take | Protesters gather in DC
POOL/GETTY IMAGES NORTH AMERICA/GETTY IMAGES

Knowing why and why-question

Q: why φ? Hintikka's view:

- The *presupposition* of Q is $\mathcal{K}\varphi$;
- The *desideratum* of Q is $\mathcal{K}\varphi$;
- A degenerated case which requires new treatment.
- Hintikka has an interesting result using Craig's Interpolation theorem about scientific explanation.

We share Koura and Schurz's view:

- The *presupposition* of Q is $\mathcal{K}\exists x E(x, \varphi)$ (and $\mathcal{K}\varphi$).
- The *desideratum* of Q is $\exists x \mathcal{K} E(x, \varphi)$.

What exactly does $\exists x$ quantify over?

Why-questions are asking for explanations but there are various kinds of explanations:

I know why φ because... (Bird 98):

- The window broke because the stone was thrown at it.
- The lump of potassium dissolved since as a law of nature potassium reacts with water to form a soluble hydroxide.
- he stayed in the bookstore all those days hoping to see her again.
- Cheetahs can run at high speeds because of the selective advantage this gives them in catching their prey.
- Blood circulates in order to supply the various parts of the body with oxygen and nutrients.

Different types of explanation

- The window broke since a stone was thrown at it. (causal)
- Potassium dissolved since as a law of nature potassium reacts with water to form a soluble hydroxide. (nomic)
- He stayed in the café all day to see her again. (psychological explanation/reason/motive)
- Cheetahs can run at high speeds because of the selective advantage this gives in catching their prey. (“Darwinian”)
- Blood circulates in order to supply the various parts of the body with oxygen and nutrients. (functional)

Scientific explanation

Various models (Schurz 1999):

- Deductive-nomological (Hempel)
- Causal (Salmon)
- Unification (Kitcher)
- ...

Other dimensions:

- probabilistic or non-probabilistic
- explain particular instance or general rule

How to handle explanations formally?

Justification logic (JL):

- To give arithmetic semantics to Intuitionistic logic (InL).
- The Brouwer-Heyting-Kolmogorov (BHK) semantics: a formula is 'true' if it has a proof (actually problematic).
- E.g., a proof of $\varphi \rightarrow \psi$ is a construction transforming proofs of φ into proofs of ψ.
- Gödel thinks S4 is a good provability logic in general.
- Gödel showed that φ is a theorem of Intuitionistic logic iff adding $\Box$ to each subformula of φ results in a theorem of S4.

Related to the BHK semantics: Curry-Howard isomorphism: propositions as types, and proofs as programs.

How to handle explanations formally?

Connect intuitionistic logic with provability (in PA) based on BHK interpretation:

- $InL \to S4 \stackrel{?}{\to} PA$. It is problematic: from T derive $\Box\neg\Box\bot$ contradicting Gödel's 2nd incompleteness theorem.
- We can change the logic to Gödel-löb ($\Box(\Box p \to p) \to \Box p$), but what if we want to keep S4?
- T is OK if we make the proofs explicit.
- $InL \to S4 \to$ *Logic of proofs* (LP, not provability) $\to PA$
- In LP we can express $t{:}\varphi$ meaning t is a proof of φ.
- Justification Logic: Interpreting t as *justification* in epistemology (e.g. Artemov 2008). To bridge the gap between EL and epistemology (van Benthem 91).
- Meaning of knowledge: Hintikka $\forall$ and Plato $\exists$ (Can they meet?)

How to handle various explanations formally?

- Conceptually, justification is not always an explanation:
  - The shadow of a flagpole is xm long can justify that the length of the pole is ym given the time and location.
  - The length of the shadow of a flagpole does not explain why the length of the pole is ym (under causal interpretation).
  - Justification gives reason to *believing* (not necessarily true) but explanation gives reason to *being* (presupposing truth).
- In JL, the technical part does not commit to such interpretation. It is very abstract proof-like thing...

A logic of knowing why

Language and semantics [Xu, Wang, Studer 19]

Given a set of agents I and a set P of basic propositional letters, the language of $\mathcal{L}_Y$ is defined as:

$$\varphi ::= \top \mid p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \mathcal{K}_i\varphi \mid \mathcal{K}y_i\varphi$$

where $p \in \mathsf{P}$ and $i \in \mathsf{I}$.

Since $\vdash \varphi$ may not imply $\vdash \mathcal{K}y_i\varphi$ we need something like *constant specification* in JL to ground reasoning from some self-evident axioms.

A *reasoning ground* is a set Λ of some instances of propositional tautologies.

What can be expressed?

- $\mathcal{K}_i p \wedge \neg \mathcal{K}y_i p$, e.g, I know that Fermat's last theorem is true but I don't know why.
- $\neg \mathcal{K}y_i p \wedge \mathcal{K}_i \mathcal{K}y_j p$, e.g., I do not know why Fermat's last theorem holds but I know that Andrew Wiles knows why.
- $\mathcal{K}_i \mathcal{K}_j p \wedge \neg \mathcal{K}y_i \mathcal{K}_j p$, e.g., I know that you know that the paper has been accepted, but I do not know why you know.
- $\mathcal{K}y_i \mathcal{K}_j p \wedge \mathcal{K}_i \neg \mathcal{K}y_j p$, e.g., I know why you know that the paper has been rejected, but I am sure you do not know why.

Multi-agent Fitting-like Model

A model is a tuple $\mathcal{M} = \langle W, E, \{R_i \mid i \in \mathcal{A}\}, \mathcal{E}, V\rangle$ where:

- W is a non-empty set of possible worlds.
- E is a non-empty set of explanations satisfying following conditions (*not* in the syntax):
  (I) If $s, t \in E$, then $(s \cdot t) \in E$ where $\cdot$ combines two explanations into one
  (II) A special symbol $e \in E$.
- $R_i \subseteq W \times W$ is an equivalence relation over W.
- $\mathcal{E} : E \times \mathcal{L}_y \to 2^W$ is an *admissible explanation function* satisfying:
  (I) $\mathcal{E}(s, \varphi \to \psi) \cap \mathcal{E}(t, \varphi) \subseteq \mathcal{E}(s \cdot t, \psi)$.
  (II) If $\varphi \in \Lambda$, then $\mathcal{E}(e, \varphi) = W$.
- $V : P \to 2^W$ is a valuation function.

We do not have $+$ in JL: $\mathcal{E}(t, \varphi) \cup \mathcal{E}(s, \varphi) \subseteq \mathcal{E}(s + t, \varphi)$

Semantics

Idea: $\mathcal{K}y_i\varphi$ iff $\exists t\, \mathcal{K}_i(t : \varphi) \wedge \mathcal{K}_i\varphi$.

$$
\begin{array}{lcl}
\mathcal{M}, w \Vdash \top & & \text{always} \\
\mathcal{M}, w \Vdash p & \iff & w \in V(p) \\
\mathcal{M}, w \Vdash \neg\varphi & \iff & \mathcal{M}, w \nVdash \varphi \\
\mathcal{M}, w \Vdash \varphi \wedge \psi & \iff & \mathcal{M}, w \Vdash \varphi \text{ and } \ \mathcal{M}, w \Vdash \psi \\
\mathcal{M}, w \Vdash \mathcal{K}_i\varphi & \iff & v \Vdash \varphi \text{ for each } v \text{ such that } wR_iv \\
\mathcal{M}, w \Vdash \mathcal{K}y_i\varphi & \iff & (1)\ \exists t \in E, \text{ for all } v \text{ such that } wR_iv,\ v \in \mathcal{E}(t, \varphi); \\
 & & (2)\ \forall v \in W, wR_iv,\ v \Vdash \varphi.
\end{array}
$$

The *uniform* explanation is important! Now we can see why we don't allow $+$ as a proof term combinator.

$\mathcal{K}_i p \wedge \neg \mathcal{K}y_i p \wedge \mathcal{K}y_j p \wedge \mathcal{K}_i \mathcal{K}y_j p$ holds on the middle world in the following model (reflexive arrows omitted).

$$\begin{array}{c} p \\ r{:}\, p \end{array} \;\overset{}{\text{—}i\text{—}}\; \begin{array}{c} p \\ t{:}\, p \\ s{:}\, p \end{array} \;\text{—}j\text{—}\; \begin{array}{c} p \\ s{:}\, p \end{array}$$

Two properties of models

Definition (Factivity)
A model $\mathcal{M}$ *is factive provided that, whenever* $w \in \mathcal{E}(t, \varphi)$*, then* $\mathcal{M}, w \Vdash \varphi$.

Definition (Introspection Property)
A model $\mathcal{M}$ *is introspective provided that, whenever* $\mathcal{M}, w \Vdash \varphi$ *and* φ *has the form* $\mathcal{K}_i\psi$ *or* $\neg\mathcal{K}_i\psi$ *or* $\mathcal{K}y_i\psi$ *or* $\neg\mathcal{K}y_i\psi$*, then* $\exists t \in E$*, for each* v *such that* wR_iv*,* $v \in \mathcal{E}(t, \varphi)$.

We use $\mathbb{C}$, $\mathbb{C}_F$, $\mathbb{C}_I$, $\mathbb{C}_{FI}$ to denote the model class of all models, factive models, introspective models, and models with both properties.

Factivity does not affect the logic

Theorem

For arbitrary formula set $\Gamma \cup \{\varphi\}$, $\Gamma \Vdash_{\mathbb{C}} \varphi$ *if and only if* $\Gamma \Vdash_{\mathbb{C}_F} \varphi$.

Theorem

For arbitrary formula set $\Gamma \cup \{\varphi\}$, $\Gamma \Vdash_{\mathbb{C}_I} \varphi$ *if and only if* $\Gamma \Vdash_{\mathbb{C}_{FI}} \varphi$.

Hint: $\mathcal{K}y_i\varphi \to \varphi$ is valid.

A proof system SKY

TAUT Classical Propositional Axioms

K $\mathcal{K}_i(\varphi \to \psi) \to (\mathcal{K}_i\varphi \to \mathcal{K}_i\psi)$

KY $\mathcal{K}y_i(\varphi \to \psi) \to (\mathcal{K}y_i\varphi \to \mathcal{K}y_i\psi)$

T $\mathcal{K}_i\varphi \to \varphi$

4 $\mathcal{K}_i\varphi \to \mathcal{K}_i\mathcal{K}_i\varphi$

5 $\neg\mathcal{K}_i\varphi \to \mathcal{K}_i\neg\mathcal{K}_i\varphi$

PRES $\mathcal{K}y_i\varphi \to \mathcal{K}_i\varphi$

4YK $\mathcal{K}y_i\varphi \to \mathcal{K}_i\mathcal{K}y_i\varphi$

5YK $\neg\mathcal{K}y_i\varphi \to \mathcal{K}_i\neg\mathcal{K}y_i\varphi$

MP Modus Ponens

NECK $\vdash \varphi \quad \Rightarrow \ \vdash \mathcal{K}_i\varphi$

NECY If $\varphi \in \Lambda$, then $\vdash \mathcal{K}y_i\varphi$

A even stronger proof system SKYI

TAUT Classical Propositional Axioms

K $\mathcal{K}_i(\varphi \to \psi) \to (\mathcal{K}_i\varphi \to \mathcal{K}_i\psi)$

KY $\mathcal{K}y_i(\varphi \to \psi) \to (\mathcal{K}y_i\varphi \to \mathcal{K}y_i\psi)$

T $\mathcal{K}_i\varphi \to \varphi$

PRES $\mathcal{K}y_i\varphi \to \mathcal{K}_i\varphi$

4KY $\mathcal{K}_i\varphi \to \mathcal{K}y_i\mathcal{K}_i\varphi$

5KY $\neg\mathcal{K}_i\varphi \to \mathcal{K}y_i\neg\mathcal{K}_i\varphi$

4Y $\mathcal{K}y_i\varphi \to \mathcal{K}y_i\mathcal{K}y_i\varphi$

5Y $\neg\mathcal{K}y_i\varphi \to \mathcal{K}y_i\neg\mathcal{K}y_i\varphi$

MP Modus Ponens

NECK $\vdash \varphi \quad \Rightarrow \; \vdash \mathcal{K}_i\varphi$

NECY If $\varphi \in \Lambda$, then $\vdash \mathcal{K}y_i\varphi$

Canonical model

The canonical model $\mathcal{M}^c = (W^c, E^c, \{R_i^c \mid i \in \mathcal{A}\}, \mathcal{E}^c, V^c)$:

- E^c is defined in BNF: $t := e \mid \varphi \mid (t \cdot t)$ where $\varphi \in \mathcal{L}_y$.
- $W^c = \{\langle \Gamma, F, \{f_i \mid i \in \mathcal{A}\}\rangle \mid \langle \Gamma, F\rangle \in \Omega \times \mathcal{P}(E^c \times \mathcal{L}_y)$,
  $f_i : \{\varphi \mid \mathcal{K}y_i\varphi \in \Gamma\} \to E^c$, F satisfies the conditions below$\}$:
  - (I) If $\langle s, \varphi \to \psi\rangle, \langle t, \varphi\rangle \in F$, then $\langle s \cdot t, \psi\rangle \in F$;
  - (II) If $\varphi \in \Lambda$, then $\langle e, \varphi\rangle \in F$;
  - (III) For arbitrary $i \in \mathcal{A}$, if $\mathcal{K}y_i\varphi \in \Gamma$, then $\langle f_i(\varphi), \varphi\rangle \in F$.
- $\langle \Gamma, F, \{f_i \mid i \in \mathcal{A}\}\rangle R_i^c \langle \Delta, G, \{g_i \mid i \in \mathcal{A}\}\rangle$ iff (1) $\Gamma_i^\# \subseteq \Delta$;
  (2) $f_i = g_i$.
- $\mathcal{E}^c : E^c \times \mathcal{L}_y \to 2^{W^c}$; $\mathcal{E}^c(t, \varphi) = \{w \mid \langle t, \varphi\rangle \in F_w\}$
- $V^c(p) = \{w \mid p \in \Gamma_w\}$

f_i mimics the formula $\mathcal{K}_i(t : \varphi)$ by picking a “witness” t.

The crucial steps

- Each MCS Γ can be extended into a $\langle \Gamma, F, \{f_i \mid i \in \mathcal{A}\} \rangle \in W^c$.
- Existence lemma for $\neg \mathcal{K}y_i \varphi$ given $\mathcal{K}_i \varphi$
- Truth lemma for $\mathcal{K}y_i \varphi$.
- $\mathcal{M}^c$ has the introspection property (in the case of $\mathbb{SYI}$).

For the second step, we need to show that for each $\langle t, \varphi \rangle \in F_w$ we can construct a successor v of w such that $\langle t, \varphi \rangle \notin F_v$.

Theorem (Soundness and completeness)
$\mathbb{SKY}$ ($\mathbb{SKYI}$) is sound and strongly complete for $\mathbb{C}_F$ ($\mathbb{C}_{FI}$).

Definition
Given any $\Gamma \in \Omega$, construct F^Γ and f^Γ as follows:

- $F_0^\Gamma = \{\langle \varphi, \varphi \rangle \mid \exists i \in \mathcal{A}, \mathcal{K}y_i \varphi \in \Gamma\} \cup \{\langle e, \varphi \rangle \mid \varphi \in \Lambda\}$
- $F_{n+1}^\Gamma = F_n^\Gamma \cup \{\langle s \cdot t, \psi \rangle \mid \langle s, \varphi \to \psi \rangle \in F_n^\Gamma, \langle t, \varphi \rangle \in F_n^\Gamma$ for some $\varphi\}$
- $F^\Gamma = \bigcup_{n \in \mathbb{N}} F_n^\Gamma$.
- $\forall i \in \mathcal{A}, f_i^\Gamma : \{\varphi \mid \mathcal{K}y_i \varphi \in \Gamma\} \to E^c, f_i^\Gamma(\varphi) = \varphi$.

Lemma ($\mathcal{K}y_i$ Existence Lemma)

For any $\langle \Gamma, F, \{f_i \mid i \in \mathcal{A}\}\rangle \in W^c$, *if* $\mathcal{K}y_i\psi \notin \Gamma$ *then for any* $\langle t, \psi\rangle \in F$, *there exists* $\langle \Delta, G, \{g_i \mid i \in \mathcal{A}\}\rangle \in W^c$ *such that* $\langle t, \psi\rangle \notin G$ *and* $\langle \Gamma, F, \{f_i \mid i \in \mathcal{A}\}\rangle R_i^c \langle \Delta, G, \{g_i \mid i \in \mathcal{A}\}\rangle$.

Suppose $\mathcal{K}y_i\psi \notin \Gamma$, $\langle \Gamma, F, \{f_i \mid i \in \mathcal{A}\}\rangle \in W^c$, and $\langle t, \psi\rangle \in F$. We construct $\langle \Delta, G, \{g_i \mid i \in \mathcal{A}\}\rangle$ as follows:

- $\Delta = \Gamma$
- $\Psi = \{\langle s, \varphi\rangle \mid \langle s, \varphi\rangle \in F$ and $\mathcal{K}y_i\varphi \notin \Gamma\}$
- $\Psi' = \{\langle t \cdot s, \varphi\rangle \mid \langle s, \varphi\rangle \in \Psi\}$
- $G_0 = (F \backslash \Psi) \cup \Psi'$
- $G_{n+1} = G_n \cup \{\langle r \cdot s, \varphi_2\rangle \mid \langle r, \varphi_1 \to \varphi_2\rangle, \langle s, \varphi_1\rangle \in G_n\}$
- $G = \bigcup_{n \in \mathbb{N}} G_n$
- For each $j \in \mathcal{A}$, $g_j : \{\varphi \mid \mathcal{K}y_j\varphi \in \Delta\} \to E^c$ is defined as:

$$g_j(\varphi) = \begin{cases} f_j(\varphi), & \langle f_j(\varphi), \varphi\rangle \notin \Psi \\ t \cdot f_j(\varphi), & \langle f_j(\varphi), \varphi\rangle \in \Psi \end{cases}$$

Connection with Justification Logic

Alternative (more JL friendly) semantics:

$$\mathcal{M}, w \Vdash \mathcal{K}y_i\varphi \iff \begin{array}{l}(1)\ \exists t \in E, w \in \mathcal{E}(t,\varphi);\\ (2)\ \forall v \in W, wR_iv, v \Vdash \varphi.\end{array}$$

Given further requirements (S5 Justification models):

- Monotonicity: $w \in \mathcal{E}(t,\varphi)$ and wRv implies $v \in \mathcal{E}(t,\varphi)$.
- S5 models.

Theorem

$\mathbb{SKY}$ *is sound and strongly complete over S5 monotonic frames w.r.t. the above alternative semantics.*

Comparison with Justification Logic (JT45)

Our language can be viewed as a fragment of first-order justification logic with normal $\mathcal{K}$.

On the semantics, we only accept the (1) and (6) below:

1. $\mathcal{E}(s, \varphi \to \psi) \cap \mathcal{E}(t, \varphi) \subseteq \mathcal{E}(s \cdot t, \psi)$
2. $\mathcal{E}(t, \varphi) \cup \mathcal{E}(s, \varphi) \subseteq \mathcal{E}(s + t, \varphi)$
3. $\mathcal{E}(t, \varphi) \subseteq \mathcal{E}(!t, t{:}\varphi)$
4. $\overline{\mathcal{E}(t, \varphi)} \subseteq \mathcal{E}(?t, \neg(t{:}\varphi))$
5. Monotonicity: $w \in \mathcal{E}(t, \varphi)$ and wRv implies $v \in \mathcal{E}(t, \varphi)$.
6. R is equivalence relation.
7. all the axioms in the system are grounded.

As for the constant specification:

We include only tautologies (but not all axioms) in our tautology ground Λ.

For example, if we had $(\mathcal{K}_i\varphi \to \varphi) \in \Lambda$, then we could derive $\mathcal{K}y_i(\mathcal{K}_i\varphi \to \varphi)$ by NECY, which would imply $\mathcal{K}y_i\mathcal{K}_i\varphi \to \mathcal{K}y_i\varphi$ by DISTY.

This would be a strange consequence: e.g., I know why I know that the window is broken implies that I know why it is broken.

Comparison with Justification Logic

Justification Logic	Our logics
$t{:}(\varphi \to \psi) \to s{:}\varphi \to (t \cdot s){:}\psi$	$\mathcal{K}y_i(\varphi \to \psi) \to (\mathcal{K}y_i\varphi \to \mathcal{K}y_i\psi)$
$t{:}\varphi \to (s + t){:}\varphi$	$\mathcal{K}y_i\varphi \to \mathcal{K}y_i\varphi$
$t{:}\varphi \to \varphi$	$\mathcal{K}y_i\varphi \to \varphi$
$t{:}\varphi \to !t{:}(t{:}\varphi)$	$\mathcal{K}y_i\varphi \to \mathcal{K}y_i\mathcal{K}y_i\varphi$
$\neg t{:}\varphi \to ?t{:}(\neg t{:}\varphi)$	$\neg\mathcal{K}y_i\varphi \to \mathcal{K}y_i\neg\mathcal{K}y_i\varphi$
$t{:}\varphi \to \Box\varphi$	$\mathcal{K}y_i\varphi \to \mathcal{K}_i\varphi$
$t{:}\varphi \to \Box t{:}\varphi$	$\mathcal{K}y_i\varphi \to \mathcal{K}_i\mathcal{K}y_i\varphi$
$\neg t{:}\varphi \to \Box\neg t{:}\varphi$	$\neg\mathcal{K}y_i\varphi \to \mathcal{K}_i\neg\mathcal{K}y_i\varphi$

Future work and Conclusions

Future work

Model theoretical, proof theoretical questions...

Pragmatics w.r.t. contrast:

- Why did the chicken cross the road? To get to the other side! Why did you rob the bank? 'Coz there is money.
- I know why John went to Beijing on Monday.
  - I know why John, not Mary, went to Beijing on Monday.
  - I know why John went to Beijing, not Shanghai, on Monday.
  - I know why John went to Beijing on Monday (not Tuesday).
- Some part of it can be formalized as $\mathcal{K}y_i(\varphi \wedge \neg\psi)$.

Understanding Why (Yu Wei's coming talk)

A true *semantic* semantics!

Commonly knowing why? Various possible definitions.

Finer structure of non-deductive explanations?

Richard Feynman's poem

I wonder why.
I wonder why.
I wonder why I wonder
I wonder why I wonder why
I wonder why I wonder!