Epistemic Logic XIII

Towards a logic of knowing who

Yanjnig Wang
Department of Philosophy

Dec. 16th 2020

Background

Epistemic logic with assignments

Background

Is it nonsense?

BEIJING, May 7 (Xinhua) – Chinese Premier Li Keqiang harshly criticized the country's excessive regulation on Wednesday, ridiculing that a citizen was even asked to prove "your mother is your mother" when obtaining a government permit.

"How ridiculous! The citizen only intended to go travelling abroad and take a vacation," Li was quoted as saying Wednesday at a State Council executive meeting on the cabinet's website.

In social networks, it is serious to prove you are you

How can I know the one who sent me the friend request is indeed Jeremy?

Yanjing: Tell me the name of the bar that we went to last time!

"Jeremy": ....

In fact, there are several "popular" scams in China featuring "smart" use of misleading identities.

Moreover, one can have many names/user names online.

Names can be considered as privacy due to search engines and automatic password crackers.

Identity is not common knowledge

In standard epistmeic logic, agents are identified with their names, and thus their identities are implicit common knowledge. Knowledge operator K_a is indexed by a name (rigid designator), which corresponds to a relation R_a in the models.

One dark and stormy night, Adam was attacked and killed. His assailant, Bob, ran away, but was seen by a passer-by, Charles, who witnessed the crime from start to finish. This led quickly to Bob's arrest. Local news picked up the story, and that is how Dave heard it the next day from the radio, over breakfast.

Now, in one sense we can say that both Charles and Dave know that Bob killed Adam. But clearly there is a difference in what they know about just this fact.

How do we tell the difference in epistemic logic?

An example by Grove (1995)

A broken robot (named a) is calling for help from the maintenance robot (named b) by sending requests. To plan further actions, the broken robot needs to know the maintenance robot knows that it needs help. Does $\mathrm{K}_a\mathrm{K}_b H(a)$ express this, if the names a, b are not commonly known?

(i) a knows that a robot named 'b', whatever it is, knows that a robot named 'a', whatever it is, needs help. (*de dicto*)

(ii) a knows that the maintenance robot knows that a robot named 'a', whatever it is, needs help.

(iii) a, the broken robot, knows that the maintenance robot knows that it, i.e. the broken robot, needs help. (*de re*)

Existing work

- Grove (1995) proposed various semantics for epistemic logic based on model with world-agent pairs.
- Fitting, Thalmann, & Voronkov (2001) proposed *term modal logic* (constants are rigid). Even the propositional part is undecidable (Padmanabha & Ramanujam 2016). Decidable fragements are discussed in (Orlandelli & Corsi 2018) and (Padmanabha & Ramanujam 2018).
- Kooi (2007) allows non-rigid constants and introduces *dynamic term modal logic*. Again, it is highly undecidable.
- Corsi and Orlandelli (2013) proposed a first-order epistemic logic with $|t : \begin{smallmatrix} t_1 \ldots t_n \\ x_1 \ldots x_n \end{smallmatrix}|$ operators based on counterpart semantics.
- Holliday and Perry (2014) use of a version of FOIL with perspective changes to handle the multi-agent case.

Our approach is a minimalistic one: a small fragment of dynamic term modal logic by Kooi suffices, which can be understood by even strangers to those *de dicto* /*de re* discussions.

Epistemic logic with assignments

Epistemic Language with assignments [Wang & Seligman]

Definition (Language ELAS)

Given a set of variables $\mathbf{X}$, a set of names $\mathbf{N}$, and a set of $\mathbf{P}$ of predicate symbols:

$$t ::= x \mid a$$

$$\varphi ::= (t \approx t) \mid Pt \dots t \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \mathsf{K}_t\varphi \mid [x := t]\varphi$$

where $x \in \mathbf{X}$, $a \in \mathbf{N}$, $P \in \mathbf{P}$. We write $\langle x := t\rangle\varphi$ as the abbreviation of $\neg[x := t]\neg\varphi$.

$[x := t]\varphi$: "φ holds after assigning the current value of t to x."

Names are not rigid in the epistemic setting but variables are, according to the semantics to be introduced later.

Model: constant domain FO Kripke structure with equiv. rel

Definition (Epistemic model $\mathcal{M}$ for ELAS)

A tuple $\langle W, I, R, \rho, \eta \rangle$ where:

- W is a non-empty set of possible worlds.
- I is a non-empty set of agents (the constant domain)
- $R : I \to 2^{W \times W}$ and R_i is an equivalence relation over W.
- $\rho : \mathbf{P} \times W \to \bigcup_{n \in \omega} 2^{I^n}$ and ρ assigns each n-ary predicate on each world an n-ary relation on I.
- $\eta : \mathbf{N} \times W \to I$.

$\sigma : \mathbf{X} \to I$ is the assignment for variables. We will consider pointed models with variable assignment: $\mathcal{M}, w, \sigma$.

Semantics

Definition (Semantics)

$$\begin{array}{rcl} \mathcal{M}, w, \sigma \vDash \mathsf{K}_t\varphi & \Leftrightarrow & \mathcal{M}, v, \sigma \vDash \varphi \text{ for all } v \text{ s.t. } w R_{\sigma_w(t)} v \\ \mathcal{M}, w, \sigma \vDash [x := t]\varphi & \Leftrightarrow & \mathcal{M}, w, \sigma[x \mapsto \sigma_w(t)] \vDash \varphi \end{array}$$

where $\sigma_w(t) = \begin{cases} \sigma(t) & t \in \mathbf{X} \\ \eta(t, w) & t \in \mathbf{N} \end{cases}$.

K_t formulas are evaluated based on what t refers to on the current world.

An **ELAS** formula is *valid* (over epistemic models) if it holds on all the (epistemic) models with assignments $\mathcal{M}, s, \sigma$.

Understanding $[x := t]$ by translation

We translate **ELAS** into a (2-sorted) first-order language with a ternary relation symbol R for the accessibility relation, a function symbol f_a for each name a, and an $n+1$-ary relation symbol Q^P for each predicate symbol P.

$$\mathrm{Tr}_w(x) = x \qquad \mathrm{Tr}_w(a) = f_a(w)$$
$$\mathrm{Tr}_w(t \approx t') = \mathrm{Tr}_w(t) \approx \mathrm{Tr}_w(t') \qquad \mathrm{Tr}_w(P\bar{t}) = Q_P(w, \mathrm{Tr}_w(\bar{t}))$$
$$\mathrm{Tr}_w(\neg\psi) = \neg\mathrm{Tr}_w(\psi) \qquad \mathrm{Tr}_w(\varphi \wedge \psi) = \mathrm{Tr}_w(\varphi) \wedge \mathrm{Tr}_w(\psi).$$
$$\mathrm{Tr}_w(\mathrm{K}_t\psi) = \forall v(R(w, \mathrm{Tr}_w(t), v) \to \mathrm{Tr}_v(\psi))$$
$$\mathrm{Tr}_w([x := t]\psi) = \exists x(x \approx \mathrm{Tr}_w(t) \wedge \mathrm{Tr}_w(\psi)) = \forall x(x \approx \mathrm{Tr}_w(t) \to \mathrm{Tr}_w(\psi))$$
(Given $x \neq t$)

We can define free and bound occurrences of variables accordingly to the first-order translation.

What can we express?

- $[x := b]\mathrm{K}_a x \approx b$: *a knows who b* is. We abbreviate it as $\mathrm{K}_a b$.
- $\neg\mathrm{K}_a a$: a does not know he is called a (e.g. the most foolish person may not know that he is the most foolish person).
- $b \approx c \wedge \mathrm{K}_a b \wedge \neg\mathrm{K}_a c$: a knows who b is but does not know who c is, although they are two names of the same person.
- $[x := a][y := b](\mathrm{K}_c M(y, x) \wedge \neg\mathrm{K}_c(a \approx x \wedge b \approx y))$: Charles knows who killed whom that night but does not know the names of the murderer and the victim.
- $\mathrm{K}_d M(b, a) \wedge \neg\mathrm{K}_d a \wedge \neg\mathrm{K}_d b$: Dave knows that a person named Bob killed a person named Adam without knowing who they are.

What can we express?

(i) a, the broken robot, knows that the robot named 'b' knows that the robot named 'a' needs help. $\mathrm{K}_a\mathrm{K}_b H(a)$

(ii) a knows that the maintenance robot knows that the robot named 'a', whatever it is, needs help. $[y := b]\mathrm{K}_a\mathrm{K}_y H(a)$

(iii) a knows that the maintenance robot knows that it, i.e. the broken robot, needs help. $[x := a][y := b]\mathrm{K}_x\mathrm{K}_y H(x)$

Valid and invalid formulas

valid $x \approx y \to \mathsf{K}_t x \approx y, x \not\approx y \to \mathsf{K}_t x \not\approx y.$

invalid $x \approx a \to \mathsf{K}_t x \approx a, x \not\approx a \to \mathsf{K}_t x \not\approx a, x \approx a \to \mathsf{K}_a x \approx a$

valid $\mathsf{K}_x \varphi \to \mathsf{K}_x \mathsf{K}_x \varphi, \neg \mathsf{K}_x \varphi \to \mathsf{K}_x \neg \mathsf{K}_x \varphi, \mathsf{K}_t \varphi \to \varphi.$

invalid $\mathsf{K}_t \varphi \to \mathsf{K}_t \mathsf{K}_t \varphi, \neg \mathsf{K}_t \varphi \to \mathsf{K}_t \neg \mathsf{K}_t \varphi$

valid $[x := y]\varphi \to \varphi[y/x]$ ($\varphi[y/x]$ is admissible)

invalid $[x := a]\varphi \to \varphi[a/x]$

valid $x \approx a \to (\mathsf{K}_x \varphi \to \mathsf{K}_a \varphi)$

invalid $[x := b]\mathsf{K}_a \varphi \to \mathsf{K}_a [x := b]\varphi$

valid $[x := y]\mathsf{K}_a \varphi \to \mathsf{K}_a [x := y]\varphi$

System SELAS

TAUT	all axioms of PL	ID	$t \approx t$
DISTK	$\mathsf{K}_t(\varphi \to \psi) \to (\mathsf{K}_t\varphi \to \mathsf{K}_t\psi)$		
Tx	$\mathsf{K}_x\varphi \to \varphi$	SUBP	$\bar{t} \approx \overline{t'} \to (P\bar{t} \leftrightarrow P\overline{t'})$ (P can be $\approx$)
4x	$\mathsf{K}_x\varphi \to \mathsf{K}_x\mathsf{K}_x\varphi$	SUBK	$t \approx t' \to (\mathsf{K}_t\varphi \leftrightarrow \mathsf{K}_{t'}\varphi)$
5x	$\neg\mathsf{K}_x\varphi \to \mathsf{K}_x\neg\mathsf{K}_x\varphi$	SUBAS	$t \approx t' \to ([x := t]\varphi \leftrightarrow [x := t']\varphi)$
RIGIDP	$x \approx y \to \mathsf{K}_t x \approx y$	RIGIDN	$x \not\approx y \to \mathsf{K}_t x \not\approx y$
DETAS	$\langle x := t\rangle\varphi \to [x := t]\varphi$	DAS	$\langle x := t\rangle\top$
KAS	$[x := t](\varphi \to \psi) \to ([x := t]\varphi \to [x := t]\psi)$		
EFAS	$[x := t]x \approx t$		
SUB2AS	$\varphi[y/x] \to [x := y]\varphi$ ($\varphi[y/x]$ is admissible)		

Rules:

MP $\dfrac{\varphi, \varphi \to \psi}{\psi}$ NECK $\dfrac{\vdash \varphi}{\vdash \mathsf{K}_t\varphi}$ NECAS $\dfrac{\vdash \varphi \to \psi}{\vdash \varphi \to [x := t]\psi}$ ($x \notin \mathrm{Fv}(\varphi)$)

Strategy of the completeness proof

No explicit quantifiers! The Barcan trick does not work.

- Extend the language with countably many new variables.
- Build a pseudo canonical frame using maximal consistent sets for various sublanguages of the extended language, with witnesses for the names.
- Given a maximal consistent set, cut out its generated subframe from the pseudo frame, and build a constant-domain canonical model, by taking certain equivalence classes of variables as the domain.
- Show that the truth lemma holds for the canonical model.
- Take the reflexive symmetric transitive closure of the relations in pseudo model and show that the truth of the formulas in the original language are preserved.
- Extending each consistent set of the original model to a maximal consistent set with witnesses.

Some results

Proposition

$[x := t]$ *cannot be eliminated qua expressivity.*

Theorem

SELAS *is sound and complete w.r.t. S5 models.*

Theorem

SELAS *is decidable over arbitrary models or reflexive models.*

SELAS *is not decidable over S5 models.*

The undecidability is proved by translating Fitting's $\lambda S5=$.

Compare $\lambda x.\Box\langle\langle\lambda y.y \approx x\rangle(c)\rangle(c)$ vs. $[x := c]\Box[y := c]x \approx y$.

Towards a logic of knowing who

A special case can be expressed by $[x := a]\mathsf{K}_b(x \approx a)$.

But there are different interpretations of "knowing who":

- I know who can help Alice. (mention-some)
  - $\exists x K_b(H(x, a))$.
- I know who came to the party yesterday. (mention-all)
  - Strongly exhaustive reading: $\forall x(\mathsf{K}_a come(x) \vee \mathsf{K}_a \neg come(x))$
  - Weakly exhaustive reading: $\forall x(come(x) \to \mathsf{K}_a come(x))$
- You know who gave the talk today, because you know the speaker is:
  - that guy over there (by identification)
  - Yanjing Wang (by name)
  - A guy from Peking University (by affiliation)
  - ...

The meaning of *knowing who* based on *conceptual covers* (Aloni 2001).

Future work

- Model theoretical issues of **ELAS**(with Yu Wei).
- Extension with increasing and varying domain models (with Yu Wei).
- Extension with function symbols (with Yu Wei).
- Extension with a (termed) common knowledge operator.
- Extension with limited quantifications over agents, e.g., all the agents know or some agent knows.
- What if multiple people have the same name?

In general, see what happens if we replace standard epistemic logic with **ELAS** in existing work of epistemic logic.

Agents are not just indexes...