



Knowing Who and More

Yanjing Wang
Department of Philosophy
Dec. 6th 2023

Background

Epistemic logic with assignments

Extensions

De Re updates

Conclusions

Background

Is it nonsense?

BEIJING, May 7 2015 – 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 epistemic 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 $K_a K_b H(a)$ 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 fragments 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 $|t : \frac{t_1 \dots t_n}{x_1 \dots x_n}|$ 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 AiML18]

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:

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

$$\varphi ::= (t \approx t) \mid Pt \dots t \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_t\varphi \mid [x := t]\varphi$$

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.

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)
- $R : I \rightarrow 2^{W \times W}$ and R_i is an equivalence relation over W .
- $\rho : \mathbf{P} \times W \rightarrow \bigcup_{n \in \omega} 2^{I^n}$ and ρ assigns each n -ary predicate on each world an n -ary relation on I .
- $\eta : \mathbf{N} \times W \rightarrow I$.

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

Definition (Semantics)

$$\begin{aligned} \mathcal{M}, w, \sigma \models K_t \varphi &\Leftrightarrow \mathcal{M}, v, \sigma \models \varphi \text{ for all } v \text{ s.t. } wR_{\sigma_w(t)}v \\ \mathcal{M}, w, \sigma \models [x := t]\varphi &\Leftrightarrow \mathcal{M}, w, \sigma[x \mapsto \sigma_w(t)] \models \varphi \end{aligned}$$

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

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, σ .

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 .

$$\text{Tr}_w(x) = x \quad \text{Tr}_w(a) = f_a(w)$$

$$\text{Tr}_w(t \approx t') = \text{Tr}_w(t) \approx \text{Tr}_w(t') \quad \text{Tr}_w(P\bar{t}) = Q_P(w, \text{Tr}_w(\bar{t}))$$

$$\text{Tr}_w(\neg\psi) = \neg\text{Tr}_w(\psi) \quad \text{Tr}_w(\varphi \wedge \psi) = \text{Tr}_w(\varphi) \wedge \text{Tr}_w(\psi).$$

$$\text{Tr}_w(K_t\psi) = \forall v(R(w, \text{Tr}_w(t), v) \rightarrow \text{Tr}_v(\psi))$$

$$\text{Tr}_w([x := t]\psi) = \exists x(x \approx \text{Tr}_w(t) \wedge \text{Tr}_w(\psi)) = \forall x(x \approx \text{Tr}_w(t) \rightarrow \text{Tr}_w(\psi))$$

(Given $x \neq t$)

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

What can we express?

- $[x := b]K_a x \approx b$: *a* knows who *b* is. We abbreviate it as $K_a b$.
- $\neg K_a a$: *a* does not know he is called *a* (e.g. the most foolish person may not know that he is the most foolish person).
- $b \approx c \wedge K_a b \wedge \neg K_a c$: *a* knows who *b* is but does not know who *c* is, although they are two names of the same person.
- $[x := a][y := b](K_c M(y, x) \wedge \neg K_c (a \approx x \wedge b \approx y))$: Charles knows who killed whom that night but does not know the names of the murderer and the victim.
- $K_d M(b, a) \wedge \neg K_d a \wedge \neg K_d b$: 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. $K_a K_b H(a)$
- (ii) a knows that the maintenance robot knows that the robot named ' a ', whatever it is, needs help. $[y := b] K_a K_y H(a)$
- (iii) a knows that the maintenance robot knows that it, i.e. the broken robot, needs help. $[x := a] [y := b] K_x K_y H(x)$

Valid and invalid formulas

valid $x \approx y \rightarrow K_t x \approx y, x \not\approx y \rightarrow K_t x \not\approx y.$

invalid $x \approx a \rightarrow K_t x \approx a, x \not\approx a \rightarrow K_t x \not\approx a, x \approx a \rightarrow K_a x \approx a$

valid $K_x \varphi \rightarrow K_x K_x \varphi, \neg K_x \varphi \rightarrow K_x \neg K_x \varphi, K_t \varphi \rightarrow \varphi.$

invalid $K_t \varphi \rightarrow K_t K_t \varphi, \neg K_t \varphi \rightarrow K_t \neg K_t \varphi$

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

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

valid $x \approx a \rightarrow (K_x \varphi \rightarrow K_a \varphi)$

invalid $[x := b] K_a \varphi \rightarrow K_a [x := b] \varphi$

valid $[x := y] K_a \varphi \rightarrow K_a [x := y] \varphi$

System SELAS

TAUT	all axioms of PL	ID	$t \approx t$
DISTK	$K_t(\varphi \rightarrow \psi) \rightarrow (K_t\varphi \rightarrow K_t\psi)$		
Tx	$K_x\varphi \rightarrow \varphi$	SUBP	$\bar{t} \approx \bar{t}' \rightarrow (P\bar{t} \leftrightarrow P\bar{t}') \text{ (} P \text{ can be } \approx \text{)}$
4x	$K_x\varphi \rightarrow K_xK_x\varphi$	SUBK	$t \approx t' \rightarrow (K_t\varphi \leftrightarrow K_{t'}\varphi)$
5x	$\neg K_x\varphi \rightarrow K_x\neg K_x\varphi$	SUBAS	$t \approx t' \rightarrow ([x := t]\varphi \leftrightarrow [x := t']\varphi)$
RIGIDP	$x \approx y \rightarrow K_t x \approx y$	RIGIDN	$x \not\approx y \rightarrow K_t x \not\approx y$
DETAS	$\langle x := t \rangle \varphi \rightarrow [x := t]\varphi$	DAS	$\langle x := t \rangle \top$
KAS	$[x := t](\varphi \rightarrow \psi) \rightarrow ([x := t]\varphi \rightarrow [x := t]\psi)$		
EFAS	$[x := t]x \approx t$		
SUB2AS	$\varphi[y/x] \rightarrow [x := y]\varphi$ ($\varphi[y/x]$ is admissible)		

Rules:

MP	$\frac{\varphi, \varphi \rightarrow \psi}{\psi}$	NECK	$\frac{\vdash \varphi}{\vdash K_t\varphi}$	NECAS	$\frac{\vdash \varphi \rightarrow \psi}{\vdash \varphi \rightarrow [x := t]\psi}$ ($x \notin \text{Fv}(\varphi)$)
-----------	--------------------------------------------------	-------------	--------------------------------------------	--------------	---------------------------------------------------------------------------------------------------------------------

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

Extensions

- Syntax: allowing function symbols in terms to handle **dependent names**, e.g., $K_a([x := f(a)]K_{m(a)}(f(a) \approx x))$. Adam know his mother knows who is **his** father.
- Semantics: allowing **varying domain** models, thus handling the uncertainty of the existence of agents.

In the varying domain model, we have an **outer domain** (I) of all the possible agents, and a **local domain** (I_w) on each world (subset of the outer domain). It may leave space for Sherlock Holmes (non-existing fictional agents to talk about).

Interesting things come up.

Extensions [Wang, Wei, Seligman APAL 22]

Varying domain models with an intuitive condition:

For all $w, v \in W, i \in I, wR_i v$ implies $i \in I_w$ (local domain of w).

It resembles Descartes' *Cogito, ergo sum*: in order to **think** some other world possible, at least the agent should **exist**.

We need to revise **reflexivity**: if $i \in I_w$ then $wR_i w$. Thus $i \in I_w$ iff there is a v such that $wR_i v$. The domain conditions can then be captured by axioms **without explicit quantifiers**:

- (i) $\widehat{K}_x T \rightarrow K_y \widehat{K}_x T$ defines increasing domain,
- (ii) $\widehat{K}_y \widehat{K}_x T \rightarrow \widehat{K}_x T$ defines decreasing domain,
- (iii) $\widehat{K}_x T$ defines strongly constant domain (no gap),
- (iv) $(\widehat{K}_x T \rightarrow K_y \widehat{K}_x T) \wedge (\widehat{K}_y \widehat{K}_x T \rightarrow \widehat{K}_x T)$ for weakly constant.

The T axiom becomes: $\widehat{K}_t T \rightarrow (K_t \varphi \rightarrow \varphi)$. $K_x \widehat{K}_x T$ is valid. See the paper for axiomatizations and (un)decidability of the logic.

Towards a logic of knowing who

A special case can be expressed by $[x := a]K_b(x \approx a)$.

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: $\forall x (K_a \text{come}(x) \vee K_a \neg \text{come}(x))$
 - Weakly exhaustive reading: $\forall x (\text{come}(x) \rightarrow K_a \text{come}(x))$
- 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

- 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?
- **Adding intentional groups**

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

Agents are not just indexes...

De Re updates

De dicto and *de re* knowledge and updates

- Standard epistemic logic mainly focuses on the *de dicto* knowledge expressed by *knowing that* φ
- Epistemic logics of knowing-wh focuses on the *de re* knowledge expressed by *knowing what/how/why* and so on, mostly sharing the general logic form of $\exists xK\varphi(x)$.
- Dynamics of *de dicto* knowledge is well-studied in Dynamic Epistemic Logic (**DEL**).
- What about the dynamics of *de re* knowledge?

De re updates

De dicto updates may change *de re* knowledge as well, but here we care more about genuine *de re* updates.

As an example, consider the *de re* and *de dicto* knowledge and updates about the value of a .

	knowledge	(simple) dynamics
<i>de dicto</i>	knowing that $a = 8$	announcing that $a = 8$
<i>de re</i>	knowing what a is	announcing what a is

“Announcing what a is” roughly amounts to announcing that $a = 0.1$ if $a = 0.1$ and announcing that $a = \pi$ if $a = \pi$ and so on... You may try to use an event model with **infinitely many** non-deterministic announcements to capture this. Can we do better without taming the infinity explicitly?

Existing work and difficulties

- van Eijck, Gattinger, and Wang introduced the $[a]$ operator capturing public announcement of the current value of a , and gave a complete axiomatization without the K operator.
- Baltag introduced new relativized epistemic operators taking care of epistemic dependencies between variables and propositions to pre-encode both the $[\!|\varphi]$ and $[a]$.
- Technically, the *de re* updates such as $[a]$ creates lots of difficulties, e.g., the *no miracle axiom* ($\langle a \rangle K\varphi \rightarrow K[a]\varphi$) fails.

What about more complicated *de re* updates which are not **public events**? For example, telling agent 1 the values of a and b but let agent 2 be ignorant about which is which.

Can we do better in a simpler way without introducing too many technical operators?

Adding public announcements [Cohen, Tang, Wang TARK21]

The language of Public Announcement Logic with Assignments (**PALAS**) is defined by adding the announcement operator:

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

$$\varphi ::= t \approx t \mid P\bar{t} \mid (\varphi \wedge \varphi) \mid \neg\varphi \mid K_i\varphi \mid [x := t]\varphi \mid [!\varphi]\varphi$$

where $x \in \mathbf{X}$, $a \in \mathbf{N}$, $P \in \mathbf{P}$ and $i \in \mathbf{I}$.

$$\mathcal{M}, w, \sigma \vDash [!\psi]\varphi \iff \mathcal{M}, w, \sigma \vDash \psi \text{ implies } \mathcal{M}|_{\psi}^{\sigma}, w, \sigma \vDash \varphi$$

where $\mathcal{M}|_{\psi}^{\sigma}$ is the submodel of \mathcal{M} restricted to the ψ worlds in \mathcal{M} ,

What can we say more?

We write $Kv_i c$ for $[x := c]K_i(x \approx c)$, write $[!a]\varphi$ for $[x := a][!(x \approx a)]\varphi$ when x is not free in φ .

- $Kv_i(\varphi, c) := K_i[!\varphi]Kv_i c$: Agent i would know the value of c given φ ;
- $Kv_i(c, d) := K_i[!c]Kv_i d$: Agent i would know the value of d given the value of c , namely, agent i knows how the value of d functionally depends on the value of c ;
- $Kv_i(c, \varphi) := K_i[!c](K_i\varphi \vee K_i\neg\varphi)$: Agent i would know the truth value of φ given the value of c , i.e., agent i knows how the truth value of φ depends on the value of c ;
- $Kv_i(\psi, \varphi) := K_i([!\psi](K_i\varphi \vee K_i\neg\varphi) \wedge [!\neg\psi](K_i\varphi \vee K_i\neg\varphi))$: Agent i knows how the truth value of φ depends on the truth value of ψ .

On top of the SELAS, we have the following reduction axioms:

Axiom Schemas	
AATOM	$[!\psi]p \leftrightarrow (\psi \rightarrow p)$ (if p is atomic)
ANEG	$[!\psi]\neg\varphi \leftrightarrow (\psi \rightarrow \neg[!\psi]\varphi)$
ACON	$[!\psi](\varphi \wedge \chi) \leftrightarrow ([!\psi]\varphi \wedge [!\psi]\chi)$
AK	$[!\psi]K_i\varphi \leftrightarrow (\psi \rightarrow K_i[!\psi]\varphi)$
ACOM	$[!\psi][!\chi]\varphi \leftrightarrow [!(\psi \wedge [!\psi]\chi)]\varphi$
AASSI	$[!\psi][x := t]\varphi \leftrightarrow [z := x][x := t][!\psi[z/x]]\varphi$ (z does not occur in $[!\psi][x := t]\varphi$)

The expressivity stays the same.

Adding Event Models

The language of Dynamic Epistemic Logic with Assignments (**DELAS**) is defined below:

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

$$\varphi ::= t \approx t \mid P\bar{t} \mid (\varphi \wedge \varphi) \mid \neg\varphi \mid K_i\varphi \mid [x := t]\varphi \mid [\mathcal{E}, e]\varphi$$

where $x \in \mathbf{X}$, $a \in \mathbf{N}$, $P \in \mathbf{P}$, $i \in \mathbf{I}$, and \mathcal{E}, e is a pointed event model w.r.t. **DELAS**.

$$\mathcal{M}, w, \sigma \vDash [\mathcal{E}, e]\varphi \Leftrightarrow \mathcal{M}, w, \sigma \vDash \text{Pre}(e) \Rightarrow \mathcal{M} \otimes \mathcal{E}, (s, e), \sigma \vDash \varphi$$

Examples

Agent 1 is told a password with agent 2 around, but agent 2 is not sure whose password it is: it could be Cindy's (c) or Dave's (d). The following event model captures such an event \mathcal{E} (with precondition specified):

$$\underbrace{e : x \approx c}_{\begin{array}{c} \curvearrowright \\ 1,2 \end{array}} \longleftarrow 2 \longrightarrow \begin{array}{c} \curvearrowright \\ 1,2 \end{array} f : x \approx d$$

Let \mathcal{M}, s be some initial pointed model capturing that 1 and 2 have no idea what the two passwords are. We have

$$\mathcal{M}, s \models [x := c][\mathcal{E}, e](Kv_1c \wedge \neg Kv_1d \wedge \neg Kv_2c \wedge \neg Kv_2d \wedge K_2(Kv_1c \vee Kv_1d)).$$

Examples

Agent 1 and agent 2 are told two numbers (the passwords of c and d) such that agent 1 knows which is which but agent 2 does not know it.

$$\underbrace{e : x \approx c, y \approx d}_{\checkmark^{1,2}} \longleftarrow 2 \longrightarrow \underbrace{f : x \approx d, y \approx c}_{\checkmark^{1,2}}$$

We can verify on the appropriate initial model:

$$\mathcal{M}, s \models [x := c][y := d][\mathcal{E}, e](Kv_1c \wedge Kv_1d \wedge K_2(Kv_1c \wedge Kv_1d)).$$

Axiom Schemas

UATOM	$[\mathcal{E}, e]p \leftrightarrow (Pre(e) \rightarrow p)$ (p is atomic)
UNEG	$[\mathcal{E}, e]\neg\varphi \leftrightarrow (Pre(e) \rightarrow \neg[\mathcal{E}, e]\varphi)$
UCON	$[\mathcal{E}, e](\varphi \wedge \psi) \leftrightarrow ([\mathcal{E}, e]\varphi \wedge [\mathcal{E}, e]\psi)$
UK	$[\mathcal{E}, e]K_i\varphi \leftrightarrow (Pre(e) \rightarrow \bigwedge_{e \rightarrow_i f} [\mathcal{E}, f]\varphi)$
UCOM	$[\mathcal{E}, e][\mathcal{E}', e']\varphi \leftrightarrow [\mathcal{E} \circ \mathcal{E}', (e, e')]\varphi$
UASSI	$[\mathcal{E}, e][x := t]\varphi \leftrightarrow [z := x][x := t][\mathcal{E}', e]\varphi$

Factual changes can be added to the event model (see the paper).

Conclusions

Conclusions

- The assignment operator can turn *de dicto* knowledge and updates into the corresponding *de re* ones.
- The assignment operator admits reduction axioms for announcements and event updates.
- The static logic with the assignment operator is axiomatized completely, so are the dynamified extensions.
- Epistemic logic with assignments can serve as a more powerful logical basis for both *de dicto* and *de re* knowledge and updates.

Repeat what we did for **DEL**.

- Adding the usual *de dicto* common knowledge operator.
- Build a more general framework based on *dynamic logic*.
- Develop the method for non-reductive axiomatizations.
- Try to find the boundary of the decidability given different frame conditions.
- Extend the framework to capture more intuitive *de re* updates.
- Use it as a tool for philosophical discussions.
- ...