Temporal epistemic logic for reasoning with delay in awareness

Elena Popova

HSE, MI RAS

SCAN 2023

Logical Omniscience in Standard Epistemic Logic

Standard epistemic logic (S5) fails to represent the real notion of reasoning.

Logical Omniscience

If an agent i knows $\varphi_1, ..., \varphi_n$ and ψ follows semantically from $\varphi_1, ..., \varphi_n$, then the agent i also knows ψ.

Non-ideal vs. ideal agents

Figure: Bogdanov-Belsky "Mental calculation"

Non-ideal agents are limited in resources:

- memory
- attention
- **time**
- etc.

Way to Avoid Logical Omniscience

Implicit and Explicit Knowledge (Fagin, Halpern, 1988)

- Implicit knowledge = knowledge of a perfect reasoner
- Explicit knowledge = implicit knowledge + awareness

Dynamic of Awareness

- External dynamics of awareness (van Benthem, Velázquez-Quesada, 2010), DEL-based
- Internal dynamics of awareness, temporal epistemic logic

(van Benthem, Velázquez-Quesada 2010)

- $[+\varphi]\psi$ – "after the agent considers φ, ψ is the case"
- $[-\varphi]\psi$ – "after the agent drops φ, ψ is the case"
- $[!\varphi]\psi$ – "after the agent implicit observes φ, ψ is the case".

What is the trigger for obtaining explicit knowledge?

Awareness Dynamics: changes in implicit knowledge + factual changes

Moment 1. Agent doesn't know (implicitly) that φ

Moment 2. Agent knows (implicitly) that φ

Moment 3. Agent knows (explicitly) that φ took place at the previous moment

Temporal Epistemic Logic for Awareness with Minimum Delay

Syntax TELAMD

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_i\varphi \mid A_i\varphi \mid X\varphi \mid Y\varphi$$

where $p \in Var$, $i \in Ag$.

- $K_i\varphi$ — agent i knows that φ
- $A_i\varphi$ — agent i is aware that φ
- $\Box\varphi := A_i\varphi \wedge K_i\varphi$ – agent i knows explicitly that φ
- $X\varphi$ — in all variants of the next step, it will be that φ is true
- $Y\varphi$ — in all variants of the previous step, it was true that φ
- $\hat{X}\varphi := \neg X\neg\varphi$, $\hat{Y}\varphi := \neg Y\neg\varphi$, $X^n := \underbrace{X\ldots X}_{n \text{ times}}\varphi$, $Y^n := \underbrace{Y\ldots Y}_{n \text{ times}}\varphi$.
- $K_i^?\varphi := K_i\varphi \vee K_i\neg\varphi$

TELAMD: Model

TELAMD Model

$\mathcal{M} = (W, (\sim_i)_{i \in Ag}, V, \rightsquigarrow, (A_i)_{i \in Ag})$, where

- $(W, (\sim_i)_{i \in Ag}, V)$ — Kripke model ($W \neq \varnothing$, $\sim_i \subseteq W \times W$, $V : Prop \longrightarrow \mathcal{P}(W)$)
- $\rightsquigarrow \subseteq W \times W$ — temporal relation
  $x \rightsquigarrow y$: y is the next temporal stage of x evolution
- $A_i : W \longrightarrow \mathcal{P}(\mathcal{L})$ — awareness function

Properties of TELAMD Model

Restrictions on $\rightsquigarrow$

1. $\forall x \forall y \forall z((y \rightsquigarrow x \wedge z \rightsquigarrow x) \rightarrow y = z)$
   reverse functionality
2. $\forall x \forall y \forall z((x \rightsquigarrow y \wedge y \sim_i z) \rightarrow \exists w(w \rightsquigarrow z \wedge x \sim_i w))$
   perfect recall
3. $\forall x \forall y((x \sim_i y \wedge \exists z(x \rightsquigarrow z)) \rightarrow \exists w(y \rightsquigarrow w))$
   knowledge about the absence of the forward dead end
4. $\forall x \forall y((x \sim_i y \wedge \neg \exists z(x \rightsquigarrow z)) \rightarrow \neg \exists w(y \rightsquigarrow w))$
   knowledge about the forward dead end
5. $\forall x \forall y((x \sim_i y \wedge \exists z(z \rightsquigarrow x)) \rightarrow \exists w(w \rightsquigarrow y))$
   knowledge about the absence of the backward dead end
6. $\forall x \forall y((x \sim_i y \wedge \neg \exists z(z \rightsquigarrow x)) \rightarrow \neg \exists w(w \rightsquigarrow y))$
   knowledge about the backward dead end.

Properties of TELAMD Model

Restrictions on $\rightsquigarrow$ and A_i

1. $x \rightsquigarrow y \Rightarrow A_i(x) \subseteq A_i(y)$
   maintenance of awareness
2. $(\varphi \in A_i(x) \wedge x \sim_i y) \Rightarrow \varphi \in A_i(y)$
   knowledge about awareness
3. $\forall x([\forall y(x \sim_i y \rightarrow y \models \varphi) \wedge \exists x' \exists x''(x' \rightsquigarrow x \wedge x' \sim_i x'' \wedge x'' \not\models \varphi)] \Rightarrow$
   $\forall z(x \rightsquigarrow z \rightarrow \hat{Y}\varphi \in A_i(z)))$
   trigger for awareness

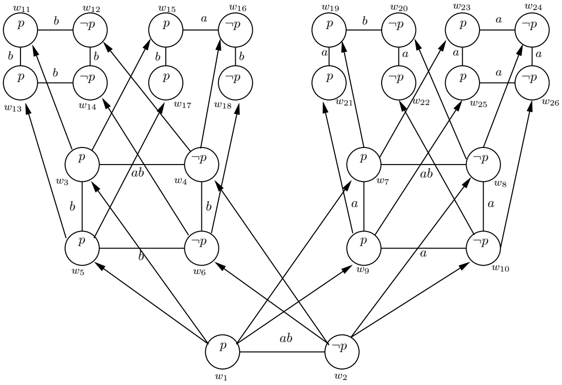
Semantics of TELAMD

Semantics

- $\mathcal{M}, w \vDash p \iff w \in V(p)$
- $\mathcal{M}, w \vDash \neg\varphi \iff \mathcal{M}, w \nvDash \varphi$
- $\mathcal{M}, w \vDash \varphi \wedge \psi \iff \mathcal{M}, w \vDash \varphi$ and $\mathcal{M}, w \vDash \psi$
- $\mathcal{M}, w \vDash K_i\varphi \iff \forall w'(w \sim_i w' \rightarrow \mathcal{M}, w' \vDash \varphi)$
- $\mathcal{M}, w \vDash X\varphi \iff \forall w'(w \rightsquigarrow w' \rightarrow \mathcal{M}, w' \vDash \varphi)$
- $\mathcal{M}, w \vDash Y\varphi \iff \forall w'(w' \rightsquigarrow w \rightarrow \mathcal{M}, w' \vDash \varphi)$
- $\mathcal{M}, w \models A_i\varphi \iff \varphi \in A_i(w)$

Example: Awareness Function is Trivial, No Factual Changes

There is an envelope on the table. Someone will leave the room, then the one who stays can look into the envelope or do nothing. Then, Alice or Brian will open the envelope.

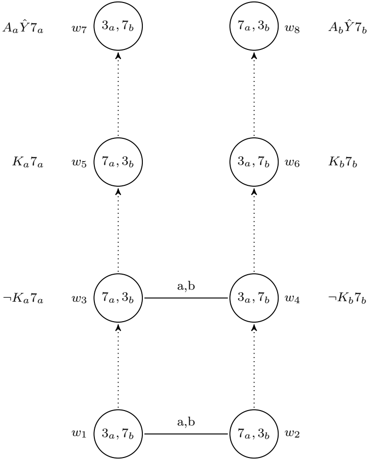


- $\mathcal{M}, w_1 \models \neg K^?_a p \land K_a \hat{X} K^?_a p$
- $\mathcal{M}, w_1 \models X \hat{X} K^?_a p$
- $\mathcal{M}, w_1 \models \hat{X}\hat{X}(K_b p \land \hat{Y} \neg K_b p)$

Example: Awareness is not Trivial + Factual Changes

At first, the players do not know what cards they have. Then, they swap cards. Next, the dealer reveals the card of one of the players. And the players swap cards again .

- $\mathcal{M}, w_5 \models K_a 7_a \wedge \neg \Box_a 7_a$
- $\mathcal{M}, w_5 \not\models A_a 7_a$
- $\mathcal{M}, w_5 \models X A_a \hat{Y} 7_a$
- $\mathcal{M}, w_7 \models A_a \hat{Y} 7_a$
- $\mathcal{M}, w_7 \models \neg 7_a \wedge \Box_a \hat{Y} 7_a$



Axiom system of *TELAMD*

Axioms:

1. CPL tautologies
2. $S5$ axioms for K_i
3. $X(\varphi \to \psi) \to (X\varphi \to X\psi)$
4. $Y(\varphi \to \psi) \to (Y\varphi \to Y\psi)$
5. $\varphi \to X\hat{Y}\varphi$
6. $\varphi \to Y\hat{X}\varphi$
7. $\hat{Y}\varphi \to Y\varphi$
8. $K_iX\varphi \to XK_i\varphi$
9. $X\bot \to K_iX\bot$
10. $\neg X\bot \to K_i\neg X\bot$
11. $Y\bot \to K_iY\bot$
12. $A_i\varphi \to XA_i\varphi$
13. $A_i\varphi \to K_iA_i\varphi$
14. $(\hat{Y}\neg K_i\varphi \wedge K_i\varphi) \to XA_i\hat{Y}\varphi$

Rules:

1. modus ponens
2. necessitation rules for X, Y, K_i

Some theorems of TELAMD

$\vdash \neg Y\bot \to K_i\neg Y\bot$

1. $X\hat{Y}\top$
2. $K_iX\hat{Y}\top$
3. $K_iX\neg Y\bot$
4. $\neg K_iX\neg Y\bot \to \bot$
5. $Y\neg K_iX\neg Y\bot \to Y\bot$
6. $\neg Y\bot \to \neg Y\neg K_iX\neg Y\bot$
7. $\neg Y\bot \to \hat{Y}K_iX\neg Y\bot$
8. $K_iX(\neg Y\bot) \to XK_i(\neg Y\bot)$
9. $\hat{Y}K_iX\neg Y\bot \to \hat{Y}XK_i\neg Y\bot$
10. $\hat{Y}X(K_i\neg Y\bot) \to (K_i\neg Y\bot)$
11. $\neg Y\bot \to K_i\neg Y\bot$

Some theorems of TELAMD

$\vdash K_i\varphi \to XK_i\hat{Y}\varphi$

1. $\varphi \to X\hat{Y}\varphi$
2. $K_i\varphi \to K_iX\hat{Y}\varphi$
3. $K_iX(\hat{Y}\varphi) \to XK_i(\hat{Y}\varphi)$
4. $K_i\varphi \to XK_i\hat{Y}\varphi$

Some theorems of TELAMD

$\vdash \hat{Y} K_i \varphi \to K_i \hat{Y} \varphi$

1. $\varphi \to X \hat{Y} \varphi$
2. $K_i \varphi \to K_i X \hat{Y} \varphi$
3. $K_i \varphi \to X K_i \hat{Y} \varphi$
4. $\hat{Y} K_i \varphi \to \hat{Y} X K_i \hat{Y} \varphi$
5. $\hat{Y} K_i \varphi \to K_i \hat{Y} \varphi$ by theorem $\hat{Y} X \varphi \to \varphi$

Soundness and Strong Completeness of TELAMD

Theorem

TELAMD is sound and strongly complete w.r.t. to TELAMD models.

$$\Gamma \models_{TELAMD} \varphi \iff \Gamma \vdash_{TELAMD} \varphi$$

Decidability of TELAMD

Theorem

TELAMD has the finite model property.

Theorem

TELAMD is decidable.

Awareness in Time and Logical Omniscience

- $(A_\circ)$ $A_i(\varphi \circ \psi) \to (A_i\varphi \wedge A_i\psi)$, where $\circ \in \{\wedge, \vee, \to, \equiv\}$.
- $(XA_\circ)$ $A_i(\varphi \circ \psi) \to (XA_i\hat{Y}\varphi \wedge XA_i\hat{Y}\psi)$
- (FC) $(p \to Xp) \wedge (\neg p \to X\neg p)$
- $TELAMD \nvdash \Box_i(\varphi \to \psi) \to (\Box_i\varphi \to \Box_i\psi)$
- $TELAMD \nvdash \Box_i(\varphi \to \psi) \to (\Box_i\varphi \to X\Box_i\psi)$
- $TELAMD \nvdash \Box_i(\varphi \to \psi) \to (\Box_i\varphi \to X\Box_i\hat{Y}\psi)$
- $TELAMD \oplus (A_\circ) \vdash \Box_i(\varphi \to \psi) \to (\Box_i\varphi \to \Box_i\psi)$
- $TELAMD \oplus (XA_\circ) \vdash \Box_i(\varphi \to \psi) \to (\Box_i\varphi \to X\Box_i\hat{Y}\psi)$
- $TELAMD \oplus (XA_\circ) \oplus (FC) \vdash \Box_i(\varphi^0 \to \psi^0) \to (\Box_i\varphi^0 \to X\Box_i\psi^0)$, where φ^0 means that the formula does not contain modal operators.

References

1. van Benthem J., Velázquez-Quesada F.R. The dynamics of awareness // Synthese. 2010. Vol. 177. P. 5–27.
2. Fagin R., Halpern J.Y., Moses Y., Vardi M.Y. Reasoning About Knowledge. Cambridge: The MIT Press, 1995.
3. Fagin R., Halpern J. Y. Belief, Awareness, and Limited Reasoning // Artificial Intelligence. 1988. Vol. 34. No 1. P. 39–76.
4. Levesque H.J. A logic of implicit and explicit belief // AAAI'84: Proceedings of the Fourth AAAI Conference on Artificial Intelligence. Austin: AAAI Press, 1984. P. 198–202.