Recent Improvements to Action Language mA*

Tran Cao Son

New Mexico State University
MSC CS, PO Box 30001
Las Cruces, New Mexico 88003

Knowledge Representation and Multiagent Systems Conventicle
University of New South Wales
13 – 14 May 2024

Acknowledgments

- Collaborators and students:

and David Buckingham

- Funding agency: various grants from NSF

Outline

1. Background

2. Dealing with False Beliefs

3. Dealing with Second Order False Beliefs

4. Dealing with Untruthful Announcements

5. Summary

Outline

1. Background

2. Dealing with False Beliefs

3. Dealing with Second Order False Beliefs

4. Dealing with Untruthful Announcements

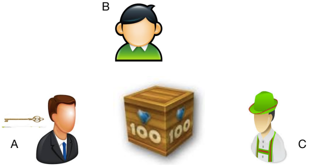
5. Summary

A high-level action specification language for RAC in multi-agent domains that

- allows for the representation of different types of actions which are typically found in multi-agent domains
- allows for the representation of action observability
- has a transition function based semantics

From Baral et al. (2022):

- Nobody knows whether the coin lies heads or tails up;
- The box is locked; needs key to open; only *A* has the key of the box;



- Peeking into an open box will learn whether the coin lies heads or tails up;
- Observing another agent peeking into the box allows the observing agent to know that the action executor knows the status of the coin but does not change his knowledge about the status of the coin;
- Distracting an agent causes the distracted agent to not look at the box;
- Signaling an agent to look at the box causes this agent to look at the box;
- Announcing that the coin lies heads up will make this a common knowledge among the agents that are listening.

A question of interest:

Can *A* know the status of the coin,
let *B* know that he knows it, and does not allow *C* to be aware of it?

Different types of actions:

- Ontic action: opening a box
- Sensing action: peeking into the box

Special for multi-agent environment

- Announcement action: announcing that the coin lies heads (or tails) up
- Manipulating observability: distracting another agents from watching self (or signaling another agents to watch self)
- Manipulating beliefs: peeking while other agents are looking

Defined over a set of agents $\mathcal{A}$ and a set of fluents P.

Ontic action: opening a box

$open(X)$ **causes** $opened$ and

$open(X)$ **executable** $has\_key(X)$

Sensing action: peeking into the box

$peek(X)$ **determines** $tail$ and

$peek(X)$ **executable** $opened, looking(X)$

Announcement action: announcing that the coin lies head (or tail) up

$shout\_tail(X)$ **announces** $tail$

Defined over a set of agents $\mathcal{A}$ and a set of fluents P.

Ontic action: opening a box

$open(X)$ **causes** $opened$ and
$open(X)$ **executable** $has\_key(X)$

Sensing action: peeking into the box

$peek(X)$ **determines** $tail$ and
$peek(X)$ **executable** $opened, looking(X)$

Announcement action: announcing that the coin lies head (or tail) up

$shout\_tail(X)$ **announces** $tail$

How about?

Manipulating observability: distracting/signaling another agents from watching self
Manipulating beliefs: peeking while other agents are looking

Defined over a set of agents $\mathcal{A}$ and a set of fluents P.

Ontic action: opening a box

$open(X)$ **causes** $opened$ and
$open(X)$ **executable** $has\_key(X)$

Sensing action: peeking into the box

$peek(X)$ **determines** $tail$ and
$peek(X)$ **executable** $opened, looking(X)$

Announcement action: announcing that the coin lies head (or tail) up

$shout\_tail(X)$ **announces** $tail$

How about?

Manipulating observability: distracting/signaling another agents from watching self this is ontic action!
Manipulating beliefs: peeking while other agents are looking sensing!
Need: specification of observability

Classification of observability

- Full observers: those who observe the action occurrence and fully aware of its effects
- Partial observers: those who observe the action occurrence but do not know of its effects
- Oblivious: those who are not aware of the action occurrence.

Possible classification of observability

action type	full observers	partial observers	oblivious
ontic actions	$\surd$		$\surd$
sensing actions	$\surd$	$\surd$	$\surd$
announcement actions	$\surd$	$\surd$	$\surd$

Three Agents and Coin Box

$X, Y \in \{A, B, C\}, X \neq Y$:

X **observes** $open(X)$	X full observer
X **observes** $peek(X)$	$--$
Y **observes** $open(X)$ **if** $looking(Y)$	$--$
Y **aware_of** $peek(X)$ **if** $looking(Y)$	Y partially observer
Y **observes** $shout\_tail(X)$	X full observer
$\{X, Y\}$ **observes** $distract(X, Y)$	X, Y full observer
$\{X, Y\}$ **observes** $signal(X, Y)$	$--$

- Epistemic logic language
- Kripke structure and epistemic state
- Satisfaction of formula in Kripke model and epistemic state
- Event model, update model, update template, and the $\otimes$ operator

$\mathcal{A}$: set of agents; P: set of propositions.

Multi-agent epistemic logic language $L(P, \mathcal{A})$

$$\varphi \overset{def}{=} \top \mid \bot \mid p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathbf{B}_i\varphi \mid C_X\varphi$$

where $p \in P$, $i \in \mathcal{A}$, $X \subseteq \mathcal{A}$.
$\mathbf{B}_i\varphi$: "agent i knows/believes φ" and
$C_X\varphi$: "φ is a common knowledge between the agents in X"

Kripke structure

$\mathcal{M} = (W, R, \pi)$, where *(i)* W is the domain, a finite set of worlds ($\mathcal{M}[S]$); *(ii)* $R : \mathcal{A} \rightarrow 2^{W \times W}$ assigns an accessibility relation R_i to each agent $i \in \mathcal{A}$ ($\mathcal{M}[i]$). *(iii)* $\pi : P \rightarrow 2^W$: valuation of that variable.
A pointed Kripke structure is a pair $(\mathcal{M}, w)$ where $\mathcal{M} = (W, R, \pi)$ and $w \in W$.

Given: $(\mathcal{M}, w)$ with $\mathcal{M} = (W, R, \pi)$ and a formula φ, $(\mathcal{M}, w) \models \varphi$ is defined as follows:

- $(\mathcal{M}, w) \models \top$ always;
- $(\mathcal{M}, w) \models \bot$ never;
- $(\mathcal{M}, w) \models p$ iff $w \in \pi(p)$;
- $(\mathcal{M}, w) \models \neg\varphi$ iff $(\mathcal{M}, w) \not\models \varphi$;
- $(\mathcal{M}, w) \models \varphi_1 \wedge \varphi_2$ iff $(\mathcal{M}, w) \models \varphi_1$; and $(\mathcal{M}, w) \models \varphi_2$;
- $(\mathcal{M}, w) \models K_i\varphi$ if for all $v \in W$, if wR_iv then $(\mathcal{M}, v) \models \varphi$; and
- $(\mathcal{M}, w) \models C_X\varphi$ if for all $v \in W$, if $w(\bigcup_{j \in X} R_j)^*v$ then $(\mathcal{M}, v) \models \varphi$

where $(\bigcup_{j \in X} R_j)^*$ is the transitive closure of $\bigcup_{j \in X} R_j$.

$\mathcal{M} \models \varphi$ if $(\mathcal{M}, w) \models \varphi$ for each $w \in W$.

$\mathcal{M} = (S, R, \pi)$: a Kripke model

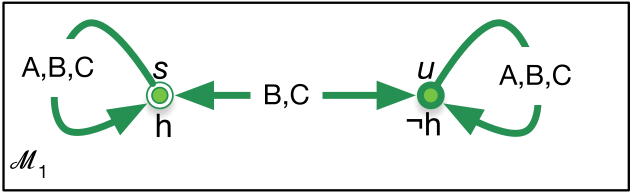
- **K** $\stackrel{def}{=} \forall i \in \mathcal{A}, \varphi, \psi \in L(P, \mathcal{A}).[\mathcal{M} \models (K_i\varphi \wedge K_i(\varphi \Rightarrow \psi)) \Rightarrow K_i\psi]$;
- **T** $\stackrel{def}{=} \forall i \in \mathcal{A}, \psi \in L(P, \mathcal{A}).[\mathcal{M} \models K_i\psi \Rightarrow \psi]$;
- **4** $\stackrel{def}{=} \forall i \in \mathcal{A}, \psi \in L(P, \mathcal{A}).[\mathcal{M} \models K_i\psi \Rightarrow K_iK_i\psi]$;
- **5** $\stackrel{def}{=} \forall i \in \mathcal{A}, \psi \in L(P, \mathcal{A}).[\mathcal{M} \models \neg K_i\psi \Rightarrow K_i\neg K_i\psi]$; and
- **D** $\stackrel{def}{=} \forall i \in \mathcal{A}, \psi \in L(P, \mathcal{A}).[\mathcal{M} \models \neg K_i \perp]$.

- $\mathcal{M}$ is **T**- (**4**-, **K**-, **5**-, **D**-, respectively) model if it satisfies property **T** (**4**, **K**, **5**, **D**, respectively).
- $\mathcal{M}$ is a **S5** model if it satisfies the properties **K**, **T**, **4**, and **5**.
- $\mathcal{M}$ is a **KD45** model if it satisfies the properties **K**, **D**, **4**, and **5**.

An epistemic state is a pair $(\mathcal{M}, W_d)$ where $\mathcal{M} = (W, R, \pi)$ is a Kripke structure and $W_D \subseteq W$. A truth value of a formula φ with respect to an epistemic state $(\mathcal{M}, W_d)$ is defined by

$$(\mathcal{M}, W_d) \models \varphi \qquad \text{iff} \qquad \forall w \in W_d.[(\mathcal{M}, w) \models \varphi]$$

$(\mathcal{M}_1, s)$: an epistemic state of the three agents and a coin in the box



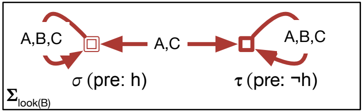
A knows that h (assuming that s is the true state)
B and C do not know whether h

An update model for $L(P, \mathcal{A})$ is a quadruple $\mathcal{E} = (E, Q, pre, sub)$ where:

- E is a finite non-empty set of events;
- $Q : \mathcal{A} \to 2^{E \times E}$ assigns an accessibility relation to each agent $i \in \mathcal{A}$;
- $pre : E \to L(P, \mathcal{A})$ assigns to each event a precondition; and
- $sub : E \to SUB(P, \mathcal{A})$ assigns to each event a substitution where each substitution is a set $\{p \leftarrow \varphi \mid p \in P\}$.

A pair $(\mathcal{E}, E_d)$ consisting of an update model $\mathcal{E} = (E, Q, pre, sub)$ and a non-empty set of designated events $E_d \subseteq E$ is called an epistemic action.

Graphical Representation of an Epistemic Action

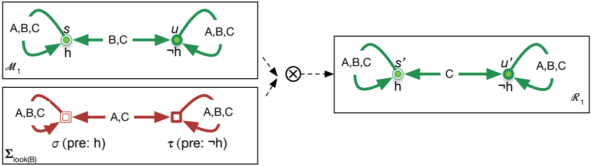


Given an epistemic action $(\mathcal{E}, E_d)$ and an epistemic state $(\mathcal{M}, W_d)$

- $(\mathcal{E}, E_d)$ is executable in $(\mathcal{M}, W_d)$ if for each $w \in W_d$ there exists at least one $e \in E_d$ such that $(\mathcal{M}, w) \models pre(e)$.
- $(\mathcal{M}, W_d) \otimes (\mathcal{E}, E_d) = ((W', R', \pi'), W'_d)$ where
  - $W' = \{(w, e) \in W \times E \mid (\mathcal{M}, w) \models pre(e)\}$
  - $R'_i = \{((w, e), (v, f)) \in W' \times W' \mid wR_iv \text{ and } eQ_if\}$
  - $\pi'(w, e)(p) = \top$ iff $(\mathcal{M}, w) \models sub(e)(p)$
  - $W'_d = \{(w, e) \in W' \mid w \in W_d \text{ and } e \in E_d\}$

  if $(\mathcal{E}, E_d)$ is *executable* in $(\mathcal{M}, W_d)$.

Example: Everyone watches while B is looking at the coin



$$s' = (s, \sigma) \text{ and } u' = (u, \tau)$$

Domain

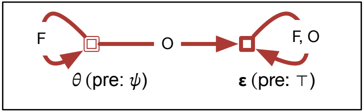
A set of statements about action effects and observability over the pair $(P, \mathcal{A})$ is a multi-agent domain.

A multi-agent domain specifies a collection of epistemic actions defined as follows. Given a Kripke model $(\mathcal{M}, s)$ and an action occurrence a,

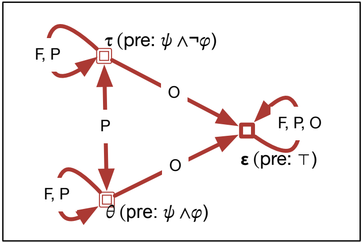
Frame of reference $\rho = (F, P, O)$

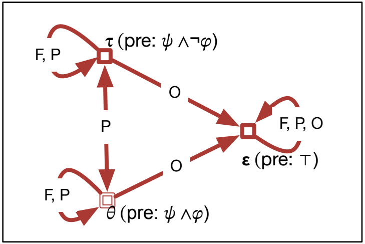
- F - the set of agents who are fully observer of the occurrence. $F = \{i \mid i\ \textbf{observes}\ \ a\ \textbf{if}\ \ \varphi, (\mathcal{M}, s) \models \varphi\}$
- P - the set of agents who are partially observer of the occurrence. $P = \{i \mid i\ \textbf{aware\_of}\ \ a\ \textbf{if}\ \ \varphi, (\mathcal{M}, s) \models \varphi\}$
- O - the set of agents who are oblivious of the occurrence. $O = \mathcal{A} \setminus (F \cup P)$

$(\mathcal{M}, s)$, a with executability condition ψ, frame of reference $\rho = (F, P, O)$



a **causes** ℓ





a **determines** φ

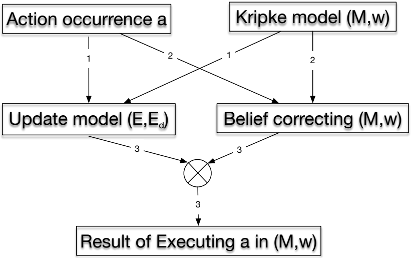
a **announces** φ

Automatically generated from the domain specification and $(\mathcal{M}, s)$

Given a domain D, an action a, and a pointed Kripke model (M, s), Φ_D defines a set of pointed Kripke models $\Phi_D(\mathrm{a}, (M, s))$ in three steps:

- Compute the update template (E, E_d) of the action occurrence a in (M, s)
- Correct false beliefs of full observers of a: this creates a new pointed Kripke model (M', s')
- Let $\Phi_D(\mathrm{a}, (M, s)) = (M', s') \otimes (E, E_d)$

Illustration



Dealing with False Beliefs

For a pointed Kripke model (M, s), an agent $i \in \mathcal{AG}$, and a formula φ, we say that i has false belief about φ in (M, s) if

$$(M, s) \models \varphi \text{ and } (M, s) \models \mathbf{B}_i \neg\varphi.$$

For a set of agents S, a pointed Kripke model (M, s), and a formula φ, such that $(M, s) \models \varphi$, let $M[S, \varphi]$ be obtained from M by replacing $M[i]$ with $M[S, \varphi][i]$ where

- $M[S, \varphi][i] = (M[i] \setminus M[i]^s) \cup \{(s, s)\}$ for $i \in S$ and $(M, s) \models \mathbf{B}_i \neg\varphi$ where $M[i]^s = \{(s, u) \mid (s, u) \in M[i]\}$; and
- $M[S, \varphi][i] = M[i]$ for other agents, i.e., $i \in \mathcal{AG} \setminus S$ or $i \in S$ and $(M, s) \not\models \mathbf{B}_i \neg\varphi$.

- If a is not executable in (M, s) then $\Phi_D(\mathrm{a}, (M, s)) = \emptyset$
- If a is executable in (M, s) and $(\mathcal{E}, E_d)$ is the representation of the occurrence of a in (M, s) then
  - $\Phi_D(\mathrm{a}, (M, s)) = (M, s) \otimes (\mathcal{E}, E_d)$ if a is a ontic action instance;
  - $\Phi_D(\mathrm{a}, (M, s)) = M_1[F_D(\mathrm{a}, M_1, s), \varphi] \otimes (\mathcal{E}, E_d)$ where $M_1 = M[P_D(\mathrm{a}, M, s), \psi]$ if a is a sensing action instance that senses φ and $(M, s) \models \varphi$;
  - $\Phi_D(\mathrm{a}, (M, s)) = M_1[F_D(\mathrm{a}, M_1, s), \neg\varphi] \otimes (\mathcal{E}, E_d)$ where $M_1 = M[P_D(\mathrm{a}, M, s), \psi]$ if a is a sensing action instance that senses φ and $(M, s) \models \neg\varphi$; and
  - $\Phi_D(\mathrm{a}, (M, s)) = M_1[F_D(\mathrm{a}, M_1, s), \varphi] \otimes (\mathcal{E}, E_d)$ where $M_1 = M[P_D(\mathrm{a}, M, s), \psi]$ if a is an announcement action instance that announces φ and $(M, s) \models \varphi$.

Finally, for a set of pointed Kripke models $\mathcal{M}$,

- if a is not executable in some $(M, s) \in \mathcal{M}$ then $\Phi_D(\mathrm{a}, \mathcal{M}) = \emptyset$;
- if a is executable in every $(M, s) \in \mathcal{M}$ then

$$\Phi_D(\mathrm{a}, \mathcal{M}) = \bigcup_{(M,s) \in \mathcal{M}} \Phi_D(\mathrm{a}, (M, s)).$$

Properties of Φ_D

Given the occurrence of an action

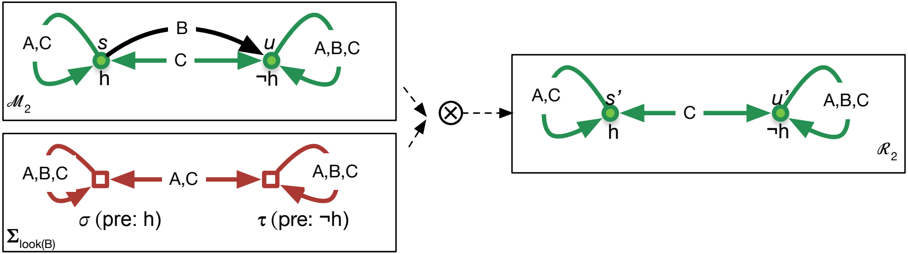
- full observers of the action occurrence know the true values of fluents that are affected (or determined) by the action occurrence;
- partial observers of the action occurrence do not know the true values of fluents but know that full observers know; and
- beliefs of oblivious agents do not change.

Outline

1. Background
2. Dealing with False Beliefs
3. Dealing with Second Order False Beliefs
4. Dealing with Untruthful Announcements
5. Summary

This problem is not unique to $m\mathcal{A}^*$. It is a problem for all approaches using the product update model operator $\otimes$.

Product Update ($\otimes$) Sensitive to False Beliefs



B looks at the coin and becomes ignorant — undesirable

Why is it undesirable for an agent to become ignorant?

- Counter intuitive: if an agent executes a sensing action then it should learn the true value of the sensed formula and correct its false belief instead of becoming ignorant!
- False belief is a nature part of multi-agent domain:
- Loss of **KD45** property because loss of seriality, which implies
  - reasoning about both knowledge and beliefs of agents is impossible if only one modality (belief) is used, i.e., both modalities are needed for reasoning about knowledge and beliefs of agents
  - it is desirable to reason about both knowledge and beliefs

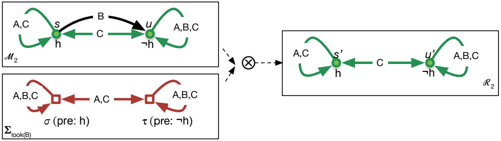
Why is it undesirable for an agent to become ignorant?

- Counter intuitive: if an agent executes a sensing action then it should learn the true value of the sensed formula and correct its false belief instead of becoming ignorant!
- False belief is a nature part of multi-agent domain:
- Loss of **KD45** property because loss of seriality, which implies
  - reasoning about both knowledge and beliefs of agents is impossible if only one modality (belief) is used, i.e., both modalities are needed for reasoning about knowledge and beliefs of agents
  - it is desirable to reason about both knowledge and beliefs

First Improvement: new product update operator

- allow agents to correct false beliefs by executing actions
- maintains **KD45** property of a **KD45** state if the update model is also **KD45**

Original $\otimes$: If an agent a has a false belief about f then sensing f makes a ignorant.



Assume that s is the true state in $\mathcal{M}_2$ ($\mathbf{B}_B\neg h$, $\mathbf{B}_A h$, $\neg(\mathbf{B}_C h \wedge \mathbf{B}_C\neg h)$, $\ldots$)
B has false belief about h

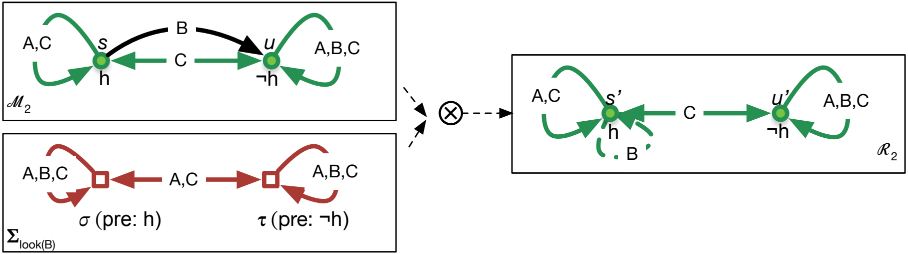
Event: *look*(B) "A, C watch while B executes the action *senses h*"

Expectation: $\mathbf{B}_B h$, $\mathbf{B}_A h$, $\neg(\mathbf{B}_C h \wedge \mathbf{B}_C\neg h)$, $\mathbf{B}_C(\mathbf{B}_B h \vee \mathbf{B}_B\neg h)$,

$\Sigma_{look(B)}$ is the update model for the **event** *look*(B)

$\mathcal{R}_2 = (\mathcal{M}_2, s) \otimes \Sigma_{look(B)}$

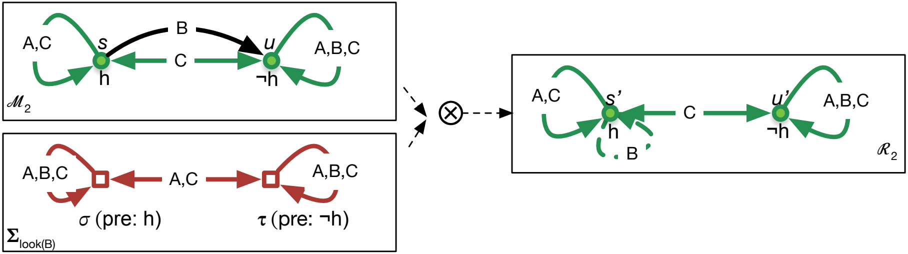
- B becomes ignorant in $\mathcal{R}_2$ because the link labeled B from s to u is not transferred to $\mathcal{R}_2$ (because $\otimes$)
  this is rightly so!
- B did not learn the value of h!
  there should be a loop labeled B at $s' = (s, \sigma)$ — $\otimes$ never adds links



Change # 1

if (x, σ) is a new world and $(\sigma, \sigma) \in R_i$ and for every u such that $(x, u) \in \mathcal{M}[i]$ there exists no $\tau \in \Sigma$ such that $(\sigma, \tau) \in R_i$ and $(\mathcal{M}, u) \models pre(\tau)$

then $((x, \sigma), (x, \sigma)) \in \mathcal{M}'[i]$



Change # 1

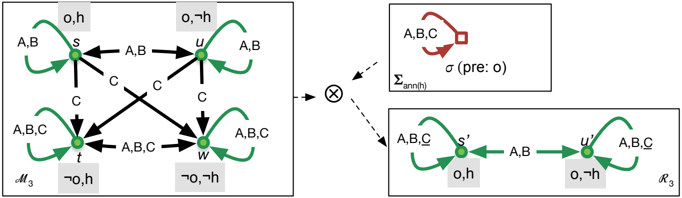
if (x, σ) is a new world and $(\sigma, \sigma) \in R_i$ (i considers that σ is possible) and for every u such that $(x, u) \in \mathcal{M}[i]$ there exists no $\tau \in \Sigma$ such that $(\sigma, \tau) \in R_i$ and $(\mathcal{M}, u) \models pre(\tau)$

(no other event is compatible to σ at world x)

then $((x, \sigma), (x, \sigma)) \in \mathcal{M}'[i]$

(for i, x must be the true world if σ occurs)

Change #1 helps but also introduces undesirable accessibility.



$\mathcal{M}_3$ and s is the true state of the world $(\neg(\mathbf{B}_C h \vee \mathbf{B}_C \neg h), \mathbf{B}_C \neg o)$

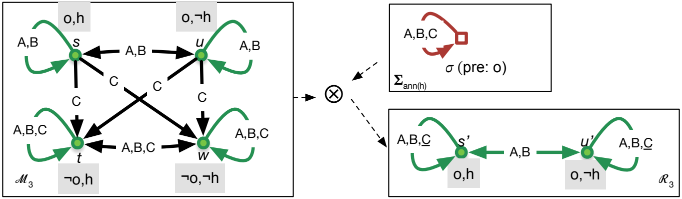
A announces o — C miraculously knows h (with **Change #1**)

Reason: **Change #1** introduces loop labeled C around $s' = (s, \sigma)$ and $u' = (u, \sigma)$

s and u are connected through un-directional links labeled C

This connection must be maintained somehow.

Change #1 helps but also introduces undesirable accessibility.



$\mathcal{M}_3$ and s is the true state of the world $(\neg(\mathbf{B}_C h \vee \mathbf{B}_C \neg h), \mathbf{B}_C \neg o)$

A announces o — C miraculously knows h (with **Change #1**)

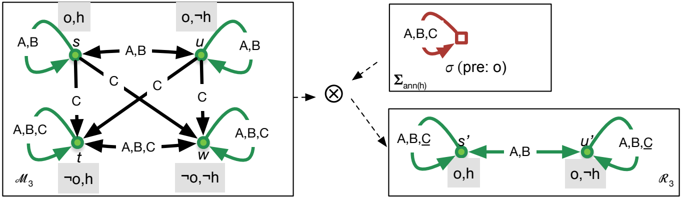
Reason: **Change #1** introduces loop labeled C around $s' = (s, \sigma)$ and $u' = (u, \sigma)$

s and u are connected through un-directional links labeled C

This connection must be maintained somehow.

maintaining $\neg(\mathbf{B}_C h \vee \mathbf{B}_C \neg h)$ requires links labeled between s' and u'

Change #1 helps but also introduces undesirable accessibility.



$\mathcal{M}_3$ and s is the true state of the world ($\neg(\mathbf{B}_C h \vee \mathbf{B}_C \neg h)$, $\mathbf{B}_C \neg o$)
A announces o — C miraculously knows h (with **Change #1**)
Reason: **Change #1** introduces loop labeled C around s' and u'

Change #2: Assume $\mathcal{M}' = \mathcal{M} \otimes \Sigma$

if $x' = (x, \sigma)$ and $y' = (y, \tau)$ are new worlds and
(x', x') and (y', y') in $\mathcal{M}'$ because of **Change #1**,
$(x, y) \not\in \mathcal{M}[i]$, $(y, x) \not\in \mathcal{M}[i]$, and x and y are connected by i
then $((x, \sigma), (y, \tau)) \in \mathcal{M}'[i]$.

Kripke structure: $\mathcal{M}$

Uupdate model: $\mathbf{\Sigma} = \langle \Sigma, R_1, \ldots, R_n, pre, sub \rangle$

$\mathcal{M}' = \mathcal{M} \otimes \mathbf{\Sigma}$

(i) $\mathcal{M}'[S] = \{(s, \tau) \mid s \in \mathcal{M}[S], \tau \in \Sigma, (\mathcal{M}, s) \models pre(\tau)\}$;

(ii) For (s, τ) and (s', τ') in $\mathcal{M}'[S]$, $((s, \tau), (s', \tau')) \in \mathcal{M}'[i]$ iff

(a) $(s, s') \in \mathbf{B}_i$ and $(\tau, \tau') \in R_i$; or

(b) $(s, \tau) = (s', \tau')$, $(\tau, \tau) \in R_i$, and $\mathbf{C_i}(s, \tau)$ is true;

(c) $(s, \tau) \neq (s', \tau')$, $\mathbf{C_i}(s, \tau)$ and $\mathbf{C_i}(s', \tau')$ are true, $(\tau, \tau), (\tau, \tau')$, $(\tau', \tau') \in R_i$, $(s, s'), (s', s) \notin \mathcal{M}[i]$, and s and s' are connected by i.

(iii) For all $(s, \tau) \in M'[S]$ and $f \in F$, $M'[\pi]((s, \tau)) \models f$ if $f \rightarrow \varphi \in sub(\tau)$ and $(M, s) \models \varphi$.

$\mathbf{C_i}(\mathbf{x}, \tau)$ encodes "for every u such that $(x, u) \in \mathcal{M}[i]$ there exists no $\tau' \in \Sigma$ such that $(\tau, \tau') \in R_i$ and $(\mathcal{M}, u) \models pre(\tau')$."

For update models defined by the action language $m\mathcal{A}^*$, pointed Kripke structure $(\mathcal{M}, s)$, if i is a full observer of an action a, $(\mathcal{M}', s') = (\mathcal{M}, s) \otimes \Sigma_a$,

- a is a sensing action (a **determines** φ), $(\mathcal{M}, s) \models \varphi$, and $(\mathcal{M}, s) \models \mathbf{B}_i \neg\varphi$ then $(\mathcal{M}', s') \models \mathbf{B}_i \varphi$
- a is an announcement action (a **announces** φ), $(\mathcal{M}, s) \models \varphi$, and $(\mathcal{M}, s) \models \mathbf{B}_i \neg\varphi$ then $(\mathcal{M}', s') \models \mathbf{B}_i \varphi$
- a is an ontic action with precondition ψ, $(\mathcal{M}, s) \models \psi$, and $(\mathcal{M}, s) \models \mathbf{B}_i \neg\psi$ then $(\mathcal{M}', s') \models \mathbf{B}_i \psi$

New - Thanks to Michael and Mehdi for asking the difficult question!

For a full observer i, and a sensing/announcement action a such that
$(\mathcal{M}, s) \models \varphi \wedge \mathbf{B}_i \neg\varphi$
then for every fluent formula φ',
if $(\mathcal{M}', s') \models \mathbf{B}_i \varphi'$ then $(\mathcal{M}', s') \models \varphi'$
if $(\mathcal{M}, s) \models \neg\varphi' \wedge \mathbf{B}_i \varphi'$ then $(\mathcal{M}', s') \not\models \mathbf{B}_i \varphi'$

New - Thanks to Michael and Mehdi for asking the difficult question!

For a full observer i, and a sensing/announcement action a such that
$(\mathcal{M}, s) \models \varphi \wedge \mathbf{B}_i \neg\varphi$
then for every fluent formula φ',
if $(\mathcal{M}', s') \models \mathbf{B}_i \varphi'$ then $(\mathcal{M}', s') \models \varphi'$
New belief is not false!
if $(\mathcal{M}, s) \models \neg\varphi' \wedge \mathbf{B}_i \varphi'$ then $(\mathcal{M}', s') \not\models \mathbf{B}_i \varphi'$
False belief is removed!

New - Thanks to Michael and Mehdi for asking the difficult question!

For a full observer i, and a sensing/announcement action a such that $(\mathcal{M}, s) \models \varphi \wedge \mathbf{B}_i \neg\varphi$

then for every fluent formula φ',

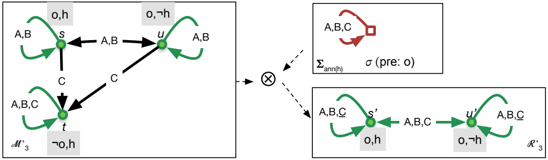
if $(\mathcal{M}', s') \models \mathbf{B}_i \varphi'$ then $(\mathcal{M}', s') \models \varphi'$

New belief is not false!

if $(\mathcal{M}, s) \models \neg\varphi' \wedge \mathbf{B}_i \varphi'$ then $(\mathcal{M}', s') \not\models \mathbf{B}_i \varphi'$

False belief is removed!

It is not all good yet! $(\mathcal{M}_3', s) \models \mathbf{B}_C h$ but $(\mathcal{R}_3', s') \not\models \mathbf{B}_C h$)

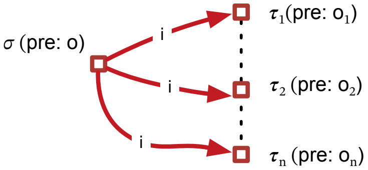


It seems that Change #2 must be applied more discretely!

Let $\mathbf{\Sigma} = \langle \Sigma, R_1, \ldots, R_n, pre, sub\rangle$ be a **KD45** update model an update model. $\mathbf{\Sigma}$ is said to be **KD45** well-defined with respect to *i* if for every $\sigma \in \Sigma$:

- $(\sigma, \sigma) \in R_i$; or
- $S_\sigma = \{pre(\tau) \mid (\sigma, \tau) \in R_i\}$ is complete

complete = for any possible world s and interpretation π over $\mathcal{P}$, there exists some $\varphi \in S_\sigma$ such that $\pi[s] \models \varphi$



Intuition: if σ is not an event for *i* then at any possible world, at least one of $\tau_1, \ldots, \tau_n$ is compatible with σ for *i*

Let $\mathbf{\Sigma} = \langle \Sigma, R_1, \ldots, R_n, pre, sub \rangle$ be a **KD45** update model an update model. $\mathbf{\Sigma}$ is said to be **KD45** well-defined with respect to i if for every $\sigma \in \Sigma$:

- $(\sigma, \sigma) \in R_i$; or
- $S_\sigma = \{pre(\tau) \mid (\sigma, \tau) \in R_i\}$ is complete

$\mathbf{\Sigma}$ is well-defined if it is well-defined with respect to all agents $1, \ldots, n$.

- If $\mathcal{M}$ is **KD45** Kripke structure and Σ is **KD45** well-defined then $\mathcal{M} \otimes^{new} \Sigma$ is **KD45**
- all update models defined for the action language $m\mathcal{A}^*$ are **KD45** well-defined

- Majority of work on dealing with false beliefs rely on belief revision, not aware of one that deals with update models
- Correcting false beliefs in the presence of update models: $m\mathcal{A}^*$ paper by Baral et al. (2022) employs an ad-hoc method to correct false beliefs prior to the execution of an action
  This approach suffers from the problem that leads to Change #2

- Herzig et al. (2005): only ontic and sensing actions, assumes that actions are always executable
  can be verified that this paper only considers well-defined update models
- Son et al. (2015): primitive update models as sufficient condition for maintaining **KD45** property, update model of ontic action in $m\mathcal{A}^*$ is not primitive
  can be verified that primitive update models are well-defined
- Aucher (2008): semantical condition on the initial Kripke structure that guarantees that the result of its update by a serial update model is serial
  our condition is applied on the update model
- Baltag and Renne (2016): the language of serial Public Announcement Logic (sPAL) that maintains the **KD45** of Kripke models after the execution of a truthful public announcement
  sPAL does not employ update models and requires that no agent has false belief about the announced formula before the action is executed

Conclusions

Introducing a new product update operator $\otimes$ that

- allows an agent to correct its false beliefs after the execution of an action
- maintains **KD45** property of a **KD45** state if the update model is also **KD45**

Outline

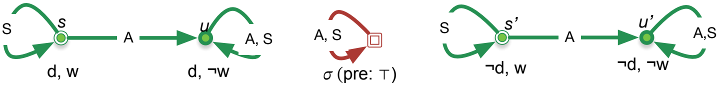
1. Background
2. Dealing with False Beliefs
3. Dealing with Second Order False Beliefs
4. Dealing with Untruthful Announcements
5. Summary

Sally and Anne are in a room containing a box and a basket. Sally places a marble in a basket and leaves the room, but secretly watches the room without Anne knowing. Anne then takes the marble from the basket and places it in a box.

When Sally returns, a child is asked "where does Anne expect her to look for the marble?"

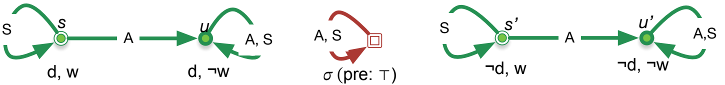
Because Sally observed Anne moving the marble, we know that Sally knows that the marble is in the box.

$m\mathcal{A}^*$ outcome:



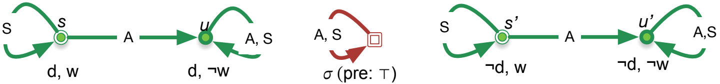
Sally and Anne are in a room containing a box and a basket. Sally places a marble in a basket and leaves the room, but secretly watches the room without Anne knowing. Anne then takes the marble from the basket and places it in a box.
When Sally returns, a child is asked "where does Anne expect her to look for the marble?" in the basket is the correct answer! $\mathbf{B}_{Anne}\mathbf{B}_{Sally} d$
Because Sally observed Anne moving the marble, we know that Sally knows that the marble is in the box.

$m\mathcal{A}^*$ outcome: $\mathbf{B}_{Anne}\mathbf{B}_{Sally}\neg d$



Correct answer: $\mathbf{B}_{Anne}\mathbf{B}_{Sally}d$

$m\mathcal{A}^*$ outcome: $\mathbf{B}_{Anne}\mathbf{B}_{Sally}\neg d$



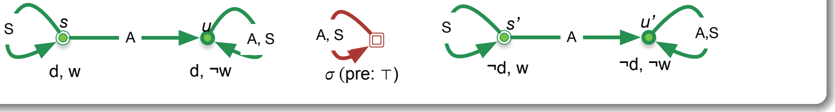
Sally **observes** Anne's action ($\neg d$) **if** Sally is watching (w)

Reasons:

- Observability is globally computed in $m\mathcal{A}^*$
- Observability should be considered **locally**

Correct answer: $\mathbf{B}_{Anne}\mathbf{B}_{Sally}d$

$m\mathcal{A}^*$ outcome: $\mathbf{B}_{Anne}\mathbf{B}_{Sally}\neg d$

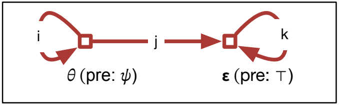


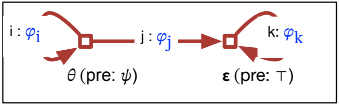
Sally **observes** Anne's action ($\neg d$) **if** Sally is watching (w)

Reasons:

- Observability is globally computed in $m\mathcal{A}^*$
  Both agents are full observers at s (correct) and this is common knowledge (incorrect)
- Observability should be considered **locally**
  To Anne, Sally is oblivious at u.

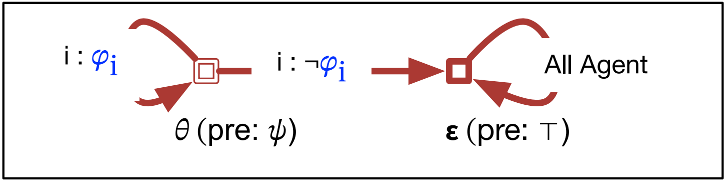
Proposed by Bolander (2018)





- Introduction of a condition for accessibility relation between two events $(\sigma, i : \varphi_i, \sigma')$ (e.g., $(\theta, i : \varphi_i, \theta)$)
- Change in construction of $\mathcal{M}' = \mathcal{M} \otimes \Sigma$ $((u, \sigma), (v, \sigma')) \in \mathcal{M}'[i]$ if $(u, v) \in \mathcal{M}[i]$, $(\sigma, i : \varphi_i, \sigma')$, **and** $(\mathcal{M}, u) \models \varphi_i$

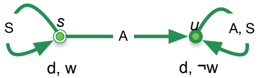
Ontic Actions



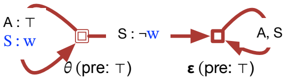
φ_i is the condition for *i* to be fully observer of the action

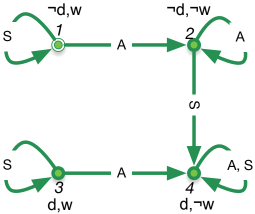
- accessibility relations in the resulting state are dynamically generated

Reference. Pham et al. (2022a)



A observes p(b) if ⊤
S observes p(b) if w

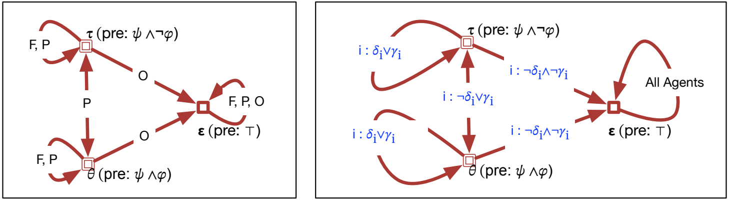




$p(b)$ **causes** $\neg d$ (p(b) Anne put the marble into the box)

- $1 = (s, \theta)$, $2 = (u, \theta)$, $3 = (s, \epsilon)$, $4 = (u, \epsilon)$
- $(1, 1) \in \mathcal{M}'[S]$: $(s, s) \in \mathcal{M}[S]$, $(\theta, w : S, \theta)$, $(\mathcal{M}, s) \models w$
- $(2, 2) \not\in \mathcal{M}'[S]$: $(u, u) \in \mathcal{M}[S]$, $(\theta, w : S, \theta)$, **but** $(\mathcal{M}, u) \not\models w$
- $(2, 4)t \in \mathcal{M}'[S]$: $(u, u) \in \mathcal{M}[S]$, $(\theta, \neg w : S, \epsilon)$, **and** $(\mathcal{M}, u) \models \neg w$
- $(1, 1) \not\in \mathcal{M}'[A]$: $(s, s) \not\in \mathcal{M}[A]$,
- ...

Sensing action: sensed formula φ, executability condition ψ



(left: original $m\mathcal{A}^*$, right: $m\mathcal{A}^*$ with edge-conditioned update model)

(Truthful) announcement: same structure, only differs in the designated events (only θ)

Let $\mathcal{T}$ be a $m\mathcal{A}^*$ action domain, a an action, and $\langle M, s\rangle$ a state. Assume that (**a**) a is executable in $\langle M, s\rangle$ and $\langle M', s'\rangle$ is the result of the execution of a in $\langle M, s\rangle$ (**b**) i, j are full observers (**c**) i believes that j is oblivious. Then, i will have a second order false belief about j belief's about the effects of the action, e.g.,

1. if a is an ontic action that makes l true and $\langle M, s\rangle \models \mathbf{B}_i\mathbf{B}_j\neg l$ then $\langle M', s'\rangle \models \mathbf{B}_i l \wedge \mathbf{B}_j l \wedge \mathbf{B}_i\mathbf{B}_j\neg l$;
2. if a is a sensing action that senses φ, $\langle M, s\rangle \models \varphi^*$ ($\varphi^* \in \{\varphi, \neg\varphi\}$) $\langle M, s\rangle \models \mathbf{B}_i\mathbf{B}_j\neg\varphi^*$ then $\langle M', s'\rangle \models \mathbf{B}_i\varphi^* \wedge \mathbf{B}_j\varphi^* \wedge \mathbf{B}_i\mathbf{B}_j\neg\varphi^*$
3. similar for announcement actions

Second approach: **employing the original update models**

When an ontic action occurs in $(\mathcal{M}, s)$

- θ: the event for full observers
- ϵ: the event for oblivious agents

They are created based on the evaluation of observability statements at s

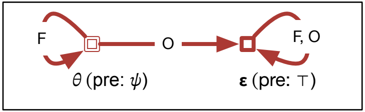
Second approach: **employing the original update models**

When an ontic action occurs in $(\mathcal{M}, s)$

- θ: the event for full observers
- ϵ: the event for oblivious agents

They are created based on the evaluation of observability statements at s

Generic update model for ontic action:



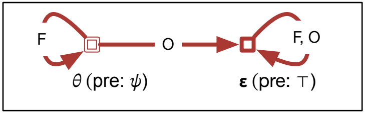
Second approach: **employing the original update models**

When an ontic action occurs in $(\mathcal{M}, s)$

- θ: the event for full observers
- ϵ: the event for oblivious agents

They are created based on the evaluation of observability statements at s

Generic update model for ontic action:



Similar events can be created for agents at different worlds: θ_u, θ_v, ϵ_u, ϵ_v, ...

- How to characterize θ_u? What are the events of the update model?
- How should they connect to each other?

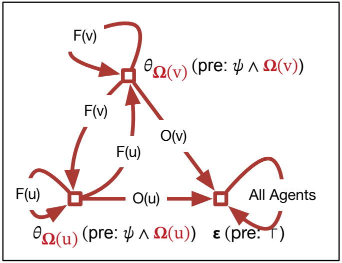
For a Kripke structure $\mathcal{M}$ and a world $u \in \mathcal{M}[S]$, the frame of reference $(F(u), P(u), O(u))$ can be characterized by

$$\boldsymbol{\Omega}(\mathbf{u}) = \Omega(F(u), P(u), O(u)) = \Big(\bigwedge_{i\in F(u)} \bigvee_{[i \textbf{ observes } a \textbf{ if } \varphi]\in D} \varphi\Big) \wedge \Big(\bigwedge_{i\in P(u)} \bigvee_{[i \textbf{ aware\_of } a \textbf{ if } \varphi]\in D} \varphi\Big) \wedge \Big(\bigwedge_{i\in O(u)} \bigvee_{[i \textbf{ observes } a \textbf{ if } \varphi]\in D \cup [i \textbf{ observes } a \textbf{ if } \varphi]\in D} \neg\varphi\Big)$$

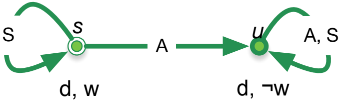
Ontic Actions

Events are $\{\theta_{\Omega(u)} \mid u \in \mathcal{M}[S]\} \cup \{\epsilon\}$

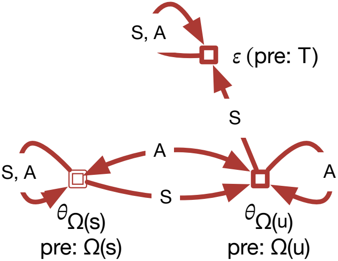
- $\theta_{\Omega(u)}$ and $\theta_{\Omega(v)}$ should be connected by full observers
- $\theta_{\Omega(u)}$ and $\epsilon_{\Omega(u)}$ should be connected by oblivious agents
- ϵ represents $\epsilon_{\Omega(u)}$ for every u

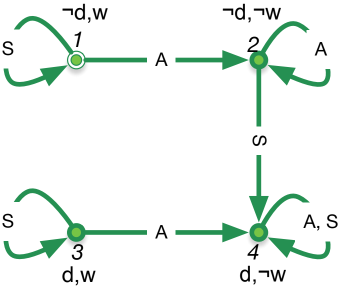

F(v)
$\theta_{\mathbf{\Omega}(\mathrm{v})}$ (pre: $\psi \wedge \mathbf{\Omega}(\mathrm{v})$)
F(v)
F(u)
O(v)
F(u)
O(u)
All Agents
$\theta_{\mathbf{\Omega}(\mathrm{u})}$ (pre: $\psi \wedge \mathbf{\Omega}(\mathrm{u})$)
$\boldsymbol{\varepsilon}$ (pre: $\top$)

Update model with 2 worlds u and v



F(s) = {S, A} and O(s) = {}
F(u) = {A} and O(u) ={A}





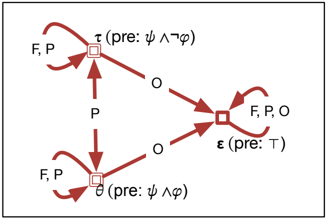
$1 = (s, \theta_{\Omega(s)})$, $2 = (u, \theta_{\Omega(u)})$,
$3 = (s, \epsilon)$, and $4 = (u, \epsilon)$.

$\mathbf{B}_{Anne}\neg d$ and $\mathbf{B}_{Sally}\neg d$

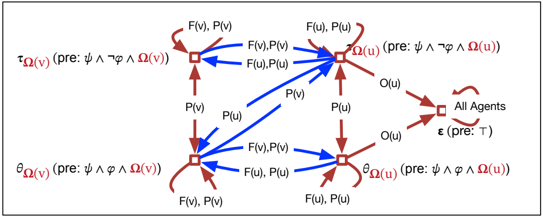
$\mathbf{B}_{Anne}\mathbf{B}_{Sally} d$, $\mathbf{B}_{Sally}\mathbf{B}_{Anne}\mathbf{B}_{Sally} d$

Sensing Actions (Announcement actions: similar)

Events: $\{\theta_{\Omega(u)} \mid u \in \mathcal{M}[S]\} \cup \{\tau_{\Omega(u)} \mid u \in \mathcal{M}[S]\} \cup \{\epsilon\}$



$m\mathcal{A}^*$



with local observability

Part of the update model for sensing action with 2 worlds u and v (Missing links between $\tau_{\Omega(v)}$ and $\theta_{\Omega(u)}$ and links from $\tau_{\Omega(v)}$ and $\theta_{\Omega(v)}$ to ϵ)

Let $\mathcal{T}$ be a $m\mathcal{A}^*$ action domain, a an action, and $\langle M, s\rangle$ a state. Assume that (**a**) a is executable in $\langle M, s\rangle$ and $\langle M', s'\rangle$ is the result of the execution of a in $\langle M, s\rangle$ (**b**) i, j are full observers (**c**) i believes that j is oblivious. Then, i will have a second order false belief about j belief's about the effects of the action, e.g.,

1. if a is an ontic action that makes l true and $\langle M, s\rangle \models \mathbf{B}_i\mathbf{B}_j\neg l$ then $\langle M', s'\rangle \models \mathbf{B}_i l \wedge \mathbf{B}_j l \wedge \mathbf{B}_i\mathbf{B}_j\neg l$;
2. if a is a sensing action that senses φ, $\langle M, s\rangle \models \varphi^*$ $(\varphi^* \in \{\varphi, \neg\varphi\}$
   $\langle M, s\rangle \models \mathbf{B}_i\mathbf{B}_j\neg\varphi^*$ then $\langle M', s'\rangle \models \mathbf{B}_i\varphi^* \wedge \mathbf{B}_j\varphi^* \wedge \mathbf{B}_i\mathbf{B}_j\neg\varphi^*$
   $\langle M, s\rangle \models \mathbf{B}_i(\neg(\mathbf{B}_j\varphi^* \vee \mathbf{B}_j\neg\varphi^*))$ then
   $\langle M', s'\rangle \models \mathbf{B}_i\varphi^* \wedge \mathbf{B}_j\varphi^* \wedge \mathbf{B}_i(\neg(\mathbf{B}_j\varphi^* \vee \mathbf{B}_j\neg\varphi^*))$ for $\varphi^* \in \{\varphi, \neg\varphi\}$
   (not sure if this is correct for edge-conditioned update models)
3. similar for announcement actions

Edge-conditioned vs. Local observability

- Edge-conditioned: number of events smaller (2 vs. many more, entailment checking is not required when creating the model)
- Local observability: computing the result is simpler (entailment checking is not required when computing the result)

Bolander (2018): two criteria for formalism based to deal with the second order false-belief tasks

- Robustness: applicable for generic theories ($m\mathcal{A}^*$ theories are sufficiently generic)
- Faithfulness, easy to understand (faithfulness). (models can be automatically generated from $m\mathcal{A}^*$ theories)

Context: multi-agent epistemic planning or action languages

- Engesser et al. (2024): *repetition-free epistemic-doxastic* (REDA), for reasoning about actions with knowledge and belief
  focuses on formulas of modal depth at most two (REDA$^{\leq 2}$)
  belief and knowledge operators
  assumes that the belief operator is serial, transitive, and euclidean
  maintains **KD45** property
- Buckingham et al. (2020): knowledge and belief operator, maintains **KD45** property
- Rajaratnam and Thielscher (2021): DER for representing and reasoning with event models for epistemic planning
  expressiveness of update models
  cannot deal with second order false beliefs
- Pham et al. (2022a): discussed in previous slide

Outline

1. Background
2. Dealing with False Beliefs
3. Dealing with Second Order False Beliefs
4. Dealing with Untruthful Announcements
5. Summary

$m\mathcal{A}^*$ and many other work: only truthful announcements

There exists a large body of research about untruthful announcements but mostly in philosophy or logics research.

$m\mathcal{A}^*$ and many other work: only truthful announcements

There exists a large body of research about untruthful announcements but mostly in philosophy or logics research.

The different facets of untruthful announcements

In the literature, often associated with

- Lying: Son tells everyone that he is a billionaire (everyone has a good laugh and nobody believes him!)
- Misleading: Son tells everyone that he is not attending (or attending) IJCAI-24 even though he is unsure whether he will be attending it (some might believe, some might not believe him!)

$m\mathcal{A}^*$ and many other work: only truthful announcements

There exists a large body of research about untruthful announcements but mostly in philosophy or logics research.

The different facets of untruthful announcements

In the literature, often associated with

- Lying: Son tells everyone that he is a billionaire (everyone has a good laugh and nobody believes him!)
- Misleading: Son tells everyone that he is not attending (or attending) IJCAI-24 even though he is unsure whether he will be attending it (some might believe, some might not believe him!)

The act of Son saying that he is a billionaire or telling everyone that he is attending IJCAI-24 is just an announcement action. Syntactically, it can be specified

Son **announces** Son-is-a-Billionare

Claim: human makes untruthful announcements pretty often :-)

So, understanding when and why such an announcement is made is

- interesting: this will help us to understand the nature of untruthful announcements (besides being a difficult academic exercise!)
- necessary: if we were to build a computer system that works with human, it is necessary for the system to understand the intention of the human and react accordingly

Claim: human makes untruthful announcements pretty often :-)

So, understanding when and why such an announcement is made is

- interesting: this will help us to understand the nature of untruthful announcements (besides being a difficult academic exercise!)
- necessary: if we were to build a computer system that works with human, it is necessary for the system to understand the intention of the human and react accordingly

Syntactically, the act of Son saying that he is a billionaire can be specified by the $m\mathcal{A}^*$ statement

Son **announces** Son-is-a-Billionare

What is its update model?

When φ is announced, how does an agent behave if this is an untruthful announcement?

- unaware of the announcement: oblivious, nothing happens to this group of agents
- full observer
  - the agent can derive that the announcer(s) is untruthful
  - the agent happens to believe that $\neg\varphi$
  - the agent does not believe whether φ nor does it believe whether $\neg\varphi$
- partial observer - does not know what is announced but is aware that there are interactions among full observers

When φ is announced, how does an agent behave if this is an untruthful announcement?

- unaware of the announcement: oblivious, nothing happens to this group of agents
- full observer
  - the agent can derive that the announcer(s) is untruthful
  - the agent happens to believe that $\neg\varphi$
    should he change? After all, his belief might be wrong!
  - the agent does not believe whether φ nor does it believe whether $\neg\varphi$
    this might depend on the attitude of the agent towards the agent(s) who makes the announcement
- partial observer - does not know what is announced but is aware that there are interactions among full observers

When φ is announced, how does an agent behave if this is an untruthful announcement?

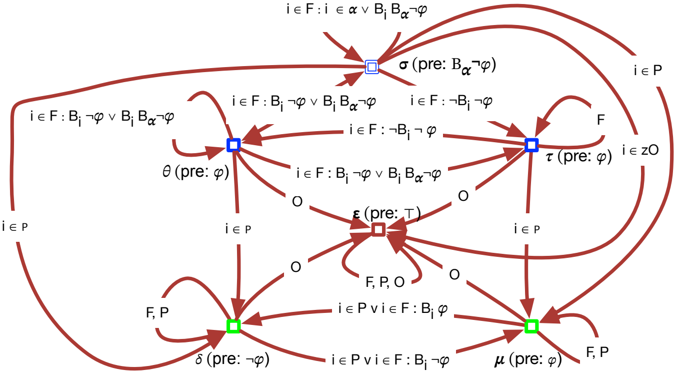
- unaware of the announcement: oblivious, nothing happens to this group of agents
- full observer
  - the agent can derive that the announcer(s) is untruthful
  - the agent happens to believe that $\neg\varphi$
    should he change? After all, his belief might be wrong!
  - the agent does not believe whether φ nor does it believe whether $\neg\varphi$
    this might depend on the attitude of the agent towards the agent(s) who makes the announcement
- partial observer - does not know what is announced but is aware that there are interactions among full observers

Several possibilities for each scenario!

- (**A1: agents are opinionated**) if an agent is certain about the truth of a formula, even if it is incorrect in the actual world, or if she realizes that the announcement is untruthful (i.e. she knows that the announcers make a false statement) then she will not change her belief about the formula, regardless of what the announcers say;
- (**A2: agents are eager to remove their uncertainties**) if an agent is uncertain about the truth of a formula and cannot reason that the announcers are untruthful then she will believe what the announcers say.

An announcement *a* by α of φ is made in $(\mathcal{M}, s)$

- *i* is a full observer
  1. $(M, s) \models \mathbf{B}_i\mathbf{B}_\alpha\neg\varphi$:
     i realizes that it is a lie so *i* will not change its belief about φ
  2. $(M, s) \models \mathbf{B}_i\neg\varphi$ or $(M, s) \models \mathbf{B}_i\varphi$:
     by (**A1**), the belief of *i* should not be changed;
  3. $(M, s) \models \neg(\mathbf{B}_i\varphi \vee \mathbf{B}_i\neg\varphi) \wedge \neg\mathbf{B}_i\mathbf{B}_\alpha\neg\varphi$:
     by (**A2**), the belief of *i* about φ should be changed and $\mathbf{B}_i\varphi$ is true after the announcement.
- *i* is a partial observer
  i is unaware of the (truth value of the) announced formula
  i's belief about φ does not change
  i's belief about fully observers changes
- *i* is not aware of the execution of action a: nothing changes for *i*.



Σ_a - update model for lying about φ in (M, s)

σ, θ, τ: events for full observers

α or $\mathbf{B}_i\mathbf{B}_\alpha\neg\varphi$: no belief change

δ, μ: partial observers

ϵ: oblivious agents

a **announces** φ

a occurs in (M, s) with $(M, s) \models \mathbf{B}_\alpha \neg\varphi$

Assume that $(M', s') = (M, s) \otimes \Sigma_a$

1. $(M', s') \models \mathbf{C}_{F_t} \neg\varphi$ where $F_t = \{i \in F \mid (M, s) \models \mathbf{B}_i \neg\varphi\}$;
2. $(M', s') \models \mathbf{C}_{F_{uf}} \varphi$ where $F_{uf} = \{i \in F \mid (M, s) \models \neg\mathbf{B}_i \neg\varphi\}$;
3. $(M', s') \models \mathbf{C}_P(\mathbf{C}_{F_u} \varphi \vee \mathbf{C}_{F_u} \neg\varphi)$ where $F_u = \{i \in F \mid (M, s) \models \neg(\mathbf{B}_i \varphi \vee \mathbf{B}_i \neg\varphi)\}$;
4. $(M', s') \models \mathbf{C}_F(\mathbf{C}_P(\mathbf{C}_{F_u} \varphi \vee \mathbf{C}_{F_u} \neg\varphi))$;
5. $(M', s') \models \mathbf{B}_j \eta$ iff $(M, s) \models \mathbf{B}_j \eta$ for a formula η and $j \in O$; and
6. $(M', s') \models \mathbf{B}_i \mathbf{B}_j \eta$ iff $(M, s) \models \mathbf{B}_i \mathbf{B}_j \eta$ for a formula η, $i \in F \cup P$ and $j \in O$.

***a* announces φ**

a occurs in (M, s) with $(M, s) \models \mathbf{B}_\alpha \neg\varphi$

Assume that $(M', s') = (M, s) \otimes \Sigma_a$

1. $(M', s') \models \mathbf{C}_{F_t} \neg\varphi$ where $F_t = \{i \in F \mid (M, s) \models \mathbf{B}_i \neg\varphi\}$;
2. $(M', s') \models \mathbf{C}_{F_{uf}} \varphi$ where $F_{uf} = \{i \in F \mid (M, s) \models \neg\mathbf{B}_i \neg\varphi\}$;
3. $(M', s') \models \mathbf{C}_P(\mathbf{C}_{F_u}\varphi \vee \mathbf{C}_{F_u}\neg\varphi)$ where $F_u = \{i \in F \mid (M, s) \models \neg(\mathbf{B}_i\varphi \vee \mathbf{B}_i\neg\varphi)\}$;
4. $(M', s') \models \mathbf{C}_F(\mathbf{C}_P(\mathbf{C}_{F_u}\varphi \vee \mathbf{C}_{F_u}\neg\varphi))$;
5. $(M', s') \models \mathbf{B}_j\eta$ iff $(M, s) \models \mathbf{B}_j\eta$ for a formula η and $j \in O$; and
6. $(M', s') \models \mathbf{B}_i\mathbf{B}_j\eta$ iff $(M, s) \models \mathbf{B}_i\mathbf{B}_j\eta$ for a formula η, $i \in F \cup P$ and $j \in O$.

- preliminary implementation in an epistemic planner Pham et al. (2023)
- update models for misleading announcements Pham et al. (2022b)
- This results might need to be revisited with local observability

Outline

1. Background
2. Dealing with False Beliefs
3. Dealing with Second Order False Beliefs
4. Dealing with Untruthful Announcements
5. Summary

Summary

$m\mathcal{A}^*$ with

- false beliefs
- second order false beliefs
- untruthful announcements
- not in this talk: other extensions such as non-deterministic actions, uncertainty in observability, etc.

References

Aucher, G. (2008). Consistency preservation and crazy formulas in BMS. In Hölldobler, S., Lutz, C., and Wansing, H., editors, *Logics in Artificial Intelligence, 11th European Conference, JELIA 2008, Dresden, Germany, September 28 - October 1, 2008. Proceedings*, volume 5293 of *Lecture Notes in Computer Science*, pages 21–33. Springer.

Baltag, A. and Renne, B. (Winter 2016). Dynamic epistemic logic. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy (Winter 2016 Edition)*.

Baral, C., Gelfond, G., Pontelli, E., and Son, T. C. (2022). An action language for multi-agent domains. *Artif. Intell.*, 302:103601.

Bolander, T. (2018). *Seeing Is Believing: Formalising False-Belief Tasks in Dynamic Epistemic Logic*, pages 207–236. Springer International Publishing, Cham.

Buckingham, D., Kasenberg, D., and Scheutz, M. (2020). Simultaneous representation of knowledge and belief for epistemic planning with belief revision. pages 172–181.

Engesser, T., Herzig, A., and Perrotin, E. (2024). Towards epistemic-doxastic planning with observation and revision. In *AAAI*.

Herzig, A., Lang, J., and Marquis, P. (2005). Action Progression and Revision in Multiagent Belief Structures. In *Sixth Workshop on Nonmonotonic Reasoning, Action, and Change (NRAC)*.

References (cont.)

Pham, L., Izmirlioglu, Y., Son, T. C., and Pontelli, E. (2022a). A new semantics for action language mA^*. In Aydogan, R., Criado, N., Lang, J., Sánchez-Anguix, V., and Serramia, M., editors, *PRIMA 2022: Principles and Practice of Multi-Agent Systems - 24th International Conference, Valencia, Spain, November 16-18, 2022, Proceedings*, volume 13753 of *Lecture Notes in Computer Science*, pages 553–562. Springer.

Pham, L., Son, T. C., and Pontelli, E. (2022b). Update models for lying and misleading announcements. In Hong, J., Bures, M., Park, J. W., and Cerný, T., editors, *SAC '22: The 37th ACM/SIGAPP Symposium on Applied Computing, Virtual Event, April 25 - 29, 2022*, pages 911–916. ACM.

Pham, L., Son, T. C., and Pontelli, E. (2023). Planning in multi-agent domains with untruthful announcements. In Koenig, S., Stern, R., and Vallati, M., editors, *Proceedings of the Thirty-Third International Conference on Automated Planning and Scheduling, July 8-13, 2023, Prague, Czech Republic*, pages 334–342. AAAI Press.

Rajaratnam, D. and Thielscher, M. (2021). Representing and reasoning with event models for epistemic planning. In *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning*, volume 18, pages 519–528.

Son, T. C., Pontelli, E., Baral, C., and Gelfond, G. (2015). Exploring the KD45n Property of a Kripke Model after the Execution of an Action Sequence. In *AAAI*.