Proceedings of the 21st International Conference on Principles of Knowledge Representation and Reasoning
Main Track

Action Language m$\mathcal{A}^*$ with Higher-Order Action Observability

David Buckingham[1] , **Matthias Scheutz**[1] , **Tran Cao Son**[2] , **Francesco Fabiano**[2]
[1]Computer Science, Tufts University, Medford, MA 02155
[2]Computer Science, New Mexico State University, Las Cruces, NM 88003
dbucki01@tufts.edu, Matthias.Scheutz@tufts.edu, stran@nmsu.edu, ffabiano@nmsu.edu

Abstract

This paper presents a novel semantics for the m$\mathcal{A}^*$ epistemic action language that takes into consideration dynamic per-agent observability of events. Different from the original m$\mathcal{A}^*$ semantics, the observability of events is defined locally at the level of possible worlds, giving a new method for compiling event models. Locally defined observability represents agents' uncertainty and false-beliefs about each others' ability to observe events. This allows for modeling second-order false-belief tasks where one agent does not know the truth about another agent's observations and resultant beliefs. The paper presents detailed constructions of event models for ontic, sensing, and truthful announcement action occurrences and proves various properties relating to agents' beliefs after the execution of an action. It also shows that the proposed approach can model second order false-belief tasks and satisfies the robustness and faithfulness criteria discussed by (Bolander 2018).

1 Introduction

Epistemic Planning (EP) addresses planning problems involving the beliefs, uncertainties, and knowledge of multiple agents. A critical question for EP is how to represent actions in such planning domains. The event models of *Dynamic Epistemic Logic* (DEL) (Baltag, Moss, and Solecki 1998; Baltag and Moss 2004; Van Ditmarsch, van Der Hoek, and Kooi 2007) offer a powerful paradigm, but face several challenges when incorporated into an epistemic planning paradigm, most notably, their construction. Indeed, many approaches to epistemic planning with event models simply present example event models without discussing how they should come to be (Bolander and Birkegaard Andersen 2011; Andersen, Bolander, and Jensen 2012; Engesser et al. 2017). The so-called syntactic approaches to EP sidestep these issues by avoiding formal models, but pay the cost of restricted expressivity, omitting, for example, disjunctive formulas (Muise et al. 2022) or common knowledge (Wan, Fang, and Liu 2021).

The epistemic action language m$\mathcal{A}^*$ (Baral et al. 2022) addresses the aforementioned issues by providing a natural-language-like interface to event models. This language can handle the beliefs and uncertainties about actions among multiple agents, and models action observability dynamically according to values that can be changed by the occurrences of actions such as whether an agent is distracted. It has been shown to be sufficiently expressive for modeling several domains used by epistemic planning systems (*e.g.*, (Muise et al. 2022; Wan, Fang, and Liu 2021; Le et al. 2018)). It has also been demonstrated that with proper extension, the language can be used for systems dealing with untruthful announcements (*e.g.*, (Pham, Son, and Pontelli 2023a)). As stated in (Baral et al. 2022) and discussed in (Rajaratnam and Thielscher 2021), m$\mathcal{A}^*$ does not have the expressiveness of full event models because it does not model beliefs and uncertainties *about* action observability. Therefore, it cannot represent situations where one agent wrongly believes that (or is uncertain about whether) another agent observes an action. This includes the *second-order false-belief* problem where one agent wrongly believes that another agent does not observe some event, and therefore develops a false belief about the second agent's beliefs. This is illustrated in the following example.

Example 1 (From (Braüner, Blackburn, and Polyanskaya 2016))**.** *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.

However, because Anne incorrectly believes that Sally did not observe her moving the marble, Anne now has the second-order false belief that Sally believes that the marble is in the basket. Thus, the child should answer that Anne expects Sally to look in the basket.

This is not the result supplied by m$\mathcal{A}^*$. The following figure details this issue[1]*:*

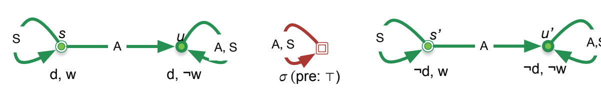

Figure 1: m$\mathcal{A}^*$ outcome

The beliefs of Sally and Anne after Sally places the marble in the basket, leaves the room, and watches Anne are shown

[1]Formal representation will be given in the later section.

in the Kripke structure on the left of Figure 1[2] *where* d *and* w *denote "marble in the basket" and "Sally is watching", respectively. The event model*[3] *encoding that Anne transfers the marble from the basket to the box, following* $\mathrm{m}\mathcal{A}^*$*, is shown in the middle of Figure 1. The result of Anne's action is shown on the right, which indicates that Anne believes that Sally believes that the marble is in the box* ($\neg\mathrm{d}$)*! This is, of course, counterintuitive.*

The above example shows that $\mathrm{m}\mathcal{A}^*$ cannot model the second-order false-belief task. We will see later that this is also true for other languages that were developed with a similar goal as $\mathrm{m}\mathcal{A}^*$ such as the language called *Dynamic Epistemic Representation* (DER) by Rajaratnam and Thielscher (2021). On the other hand, modeling second-order false-belief tasks has been an intensive research topic in several areas (*e.g*., philosophy, cognitive science, psychology, game development, or logics) and has several practical uses as summarized in the KRR-2023 invited talk[4] by Verbrugge (2023). Furthermore, understanding how to model this task will enable the development of computational tools for explaining behaviors in false-belief tasks and, potentially, detecting deceptive behaviors.

The main contribution of this paper is a novel semantics for $\mathrm{m}\mathcal{A}^*$ that takes into consideration local observability (Section 3) and solves the second-order false-belief task (Section 4). We argue that the new approach satisfies the two criteria, *robustness* and *faithfulness*, which are proposed by (Bolander 2018) for any formalism dealing with second-order false-belief tasks (Section 4). We relate $\mathrm{m}\mathcal{A}^*$ under the new semantics to other formalisms such as (Bolander 2018; Pham et al. 2022; Engesser, Herzig, and Perrotin 2024; Rajaratnam and Thielscher 2021) or some other extensions of $\mathrm{m}\mathcal{A}^*$ (Section 5) and conclude in Section 6.

2 Preliminaries

2.1 Dynamic Epistemic Logic

Dynamic Epistemic Logic defines transitions between states represented as Kripke structures by means of event models (also called "update models"). Kripke structures represent both the material condition of the task environment and agents' beliefs, including their uncertainties and false-beliefs (Fagin et al. 1995). Similarly, event models capture beliefs, knowledge, and uncertainty about events that occur, and model the interactions between agents' mental states when events occur. An accessible introduction to using DEL for EP is provided in (Bolander 2017). Here we review the elements of DEL necessary for the present work, adopting much of the notation of (Baral et al. 2022).

Assume a finite set of atomic propositional variables $\mathcal{P}$ called *propositions*. A *propositional literal* is either p or $\neg p$, where $p \in \mathcal{P}$. $\mathcal{L}^{\mathcal{P}}$ is the language of logical formulas over $\mathcal{P}$ built with the usual Boolean connectives, defined with the BNF:

$$\varphi \coloneqq p \mid \neg\varphi \mid (\varphi \wedge \varphi)$$

where $p \in \mathcal{P}$. A *propositional formula* is a formula in $\mathcal{L}^{\mathcal{P}}$.

Given a finite set of names $\mathcal{G}$ called *agents*, the language $\mathcal{L}^{\mathcal{P}}_{\mathcal{G}}$ augments $\mathcal{L}^{\mathcal{P}}$ with a modal operator $\mathbf{B}_i$ for each agent $i \in \mathcal{G}$, and is defined with the BNF:

$$\varphi \coloneqq p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \mathbf{B}_i\varphi \mid \mathbf{C}_g\varphi$$

where $p \in \mathcal{P}$, $i \in \mathcal{G}$, and $g \subseteq \mathcal{G}$. The modal operator $\mathbf{B}_i$ indicates i's belief, *e.g*., $\mathbf{B}_i\varphi$ reads as "i believes that φ". The modal operator $\mathbf{C}_g$ indicates common belief among agents g. We will say that "i knows φ" if $\varphi \wedge \mathbf{B}_i\varphi$. As usual, let $\varphi \vee \psi \coloneqq \neg(\neg\varphi \wedge \neg\psi)$, $\varphi \rightarrow \psi \coloneqq \neg\varphi \vee \psi$, $\bot \coloneqq p \wedge \neg p$, and $\top \coloneqq \neg\bot$.

A Kripke structure is a tuple $\langle W, V, R_1, \ldots, R_n\rangle$, where W is a set of *worlds*, $V : W \rightarrow 2^{\mathcal{P}}$ assigns a *valuation* over $\mathcal{P}$ to each world, and for $1 \leq i \leq n, R_i \subseteq W \times W$ is a binary relation over W. A Kripke structure is *serial* if for every $u \in W$ and $i \in \mathcal{G}$, there exists some $\langle u, v\rangle$ in R_i. A *state* is a pair $\langle\langle W, V, R_1, \ldots, R_n\rangle, d\rangle$ where $d \in W$ is the *designated world*, whose valuation gives the "actual" value of each proposition.

Given a state $\langle M, s\rangle$, where $M = \langle W, V, R_1, \ldots R_n\rangle$ and $s \in W$, entailment of formulas in $\mathcal{L}^{\mathcal{P}}$ or $\mathcal{L}^{\mathcal{P}}_{\mathcal{G}}$ are defined as follows:

$\langle M, u\rangle \models p$	iff $p \in V(u)$
$\langle M, u\rangle \models \neg\varphi$	iff $\langle M, u\rangle \not\models \varphi$
$\langle M, u\rangle \models \varphi \wedge \psi$	iff $\langle M, u\rangle \models \varphi$ and $\langle M, u\rangle \models \psi$
$\langle M, u\rangle \models \mathbf{B}_i\varphi$	iff for all $v \in W$, if $\langle u, v\rangle \in R_i$ then $\langle M, v\rangle \models \varphi$
$\langle M, u\rangle \models \mathbf{C}_g\varphi$	iff for all $v \in W$, if $u(\cup_{i\in g} R_i)^* v$ then $\langle M, v\rangle \models \varphi$

where $p \in \mathcal{P}$, $i \in \mathcal{G}$, $g \subseteq \mathcal{G}$, $\varphi, \psi \in \mathcal{L}^{\mathcal{P}}_{\mathcal{G}}$, and $(R)^*$ is the transitive closure of R. If $\varphi \in \mathcal{L}^{\mathcal{P}}$ then we may also write $u \models \varphi$ iff $\langle M, u\rangle \models \varphi$ for some Kripke structure M (since the semantics involves only the valuation at u).

An $\mathcal{L}^{\mathcal{P}}_{\mathcal{G}}$*-substitution* is a set $\{p_1 \leftarrow \varphi_1, \ldots, p_k \leftarrow \varphi_k\}$, where each p_i is a distinct proposition in $\mathcal{P}$ and each φ_i is in $\mathcal{L}^{\mathcal{P}}_{\mathcal{G}}$. In what follows, we will often write $\emptyset$ to denote the substition $\{p \leftarrow p \mid p \in \mathcal{P}\}$. $SUB_{\mathcal{L}^{\mathcal{P}}_{\mathcal{G}}}$ denotes the set of all $\mathcal{L}^{\mathcal{P}}_{\mathcal{G}}$-substitutions.

An *event model* is a tuple $\Sigma = \langle E, R^{\Sigma}_1, \ldots, R^{\Sigma}_n, \; pre, sub\rangle$, where E is a set of *events*, for $1 \leq i \leq n, R^{\Sigma}_i \subseteq E \times E$ is a binary relation over E, $pre : E \rightarrow \mathcal{L}^{\mathcal{P}}_{\mathcal{G}}$ is a function mapping each event to a formula (the event's preconditions), and $sub : E \rightarrow SUB_{\mathcal{L}^{\mathcal{P}}_{\mathcal{G}}}$ is a function mapping each event to a substitution (ontic effects).

Given Kripke structure $M = \langle W, V, R_1, \ldots, R_n\rangle$ and event model $\Sigma = \langle E, R^{\Sigma}_1, \ldots, R^{\Sigma}_n, pre, sub\rangle$, the *product update* induced by Σ in M defines a new Kripke structure $M \times \Sigma \coloneqq \langle W', V', R'_1, \ldots, R'_n\rangle$ where

$$W' = \{\langle u, e\rangle \mid u \in W, e \in E, \langle M, u\rangle \models pre(e)\}, \quad (1)$$

[2]Labeled circles and labeled links represent the worlds and accessibility relations of the agents, respectively and double circle encodes the true state of the world. Interpretations of the worlds are given below them.

[3]Event models are drawn similar to Kripke structures with the key difference: squares represent events.

[4]https://kr.org/KR2023/InvitedTalkSlides/RinekeVerbrugge.pdf

$$\langle\langle u,e\rangle,\langle v,f\rangle\rangle \in R'_i \text{ iff } \langle u,e\rangle,\langle v,f\rangle \in W',$$
$$\langle u,v\rangle \in R_i, \text{ and } \langle e,f\rangle \in R_i^\Sigma, \quad (2)$$

and for all $\langle u,e\rangle \in W'$ and $p \in \mathcal{P}, p \in V'(\langle u,e\rangle)$ iff

$$(p \in V(u) \text{ and for every } p' \leftarrow \varphi \in sub(e), p \neq p') \text{ or}$$
$$(p \leftarrow \varphi \in sub(e) \text{ and } \langle M,u\rangle \models \varphi). \quad (3)$$

An *event template* is a pair $\langle\Sigma,\Gamma\rangle$, where $\Sigma = \langle E, R_1^\Sigma, \ldots, R_n^\Sigma, pre, sub\rangle$ is an event model and $\Gamma \subseteq E$ are the *designated events*. An event template $\langle\Sigma,\Gamma\rangle$ applied in a state $\langle M,d\rangle$ causes a *state update* resulting in a set of new states, $\langle M,d\rangle \times \langle\Sigma,\Gamma\rangle := \{(M \times \Sigma, (d,e)) \mid e \in \Gamma, \langle M,d\rangle \models pre(e)\}$. Intuitively, $\langle M,d\rangle \times \langle\Sigma,\Gamma\rangle$ is the result of executing (Σ,Γ) in $\langle M,d\rangle$. For deterministic actions, this set of new states will be a singleton because there will only be one event in Γ whose precondition is satisfied at d.

2.2 m$\mathcal{A}^*$ Syntax

This section gives an overview of m$\mathcal{A}^*$ syntax. Each action defined by m$\mathcal{A}^*$ is either an ontic action, a sensing action, or an announcement action. A strength of m$\mathcal{A}^*$ is that it defines agents' observability of actions dynamically, *i.e.*, as conditioned by formulas. Three tiers of observers are supported: full observers who know an action occurs, oblivious agents, who know nothing about the action occurrence, and an intermediate class, partial observers, applicable only for sensing and announcement actions. Partial observers know *that* full observers learn the values of sensed and announced propositions, but do not themselves learn those values.

Given a finite set of actions A and a finite set of agents $\mathcal{G}$, the m$\mathcal{A}^*$ language consists of statements of the following forms:

1. "executable a if ψ"
2. "a causes l if ψ"
3. "a determines φ"
4. "a announces φ"
5. "i observes a if φ"
6. "i aware_of a if φ"

where $a \in A$, $\psi \in \mathcal{L}^\mathcal{P}$, $\varphi \in \mathcal{L}^\mathcal{P}$, l is a propositional literal, and $i \in \mathcal{G}$. A *theory* $\mathcal{T}$ is a finite collection of such statements that defines the actions of a planning domain. The first form means that a can occur only if ψ is true. We refer to ψ as the *precondition* of a. When $\psi = \top$ the statement will be omitted. The second form means that if ψ is true, then a causes fluent p to become true if $l = p$, or false if $l = \neg p$. When $\psi = \top$ the statement is written without the trivial condition. The third form means that a is a sensing action, causing full observers to learn the value of φ. The fourth form means that a is an announcement action, causing full observers to learn that φ (as in m$\mathcal{A}^*$, false announcements such as lies are not allowed). We refer to φ in statements of the third and fourth form as *sensing* or *announcement* formula, respectively. Statements of the fifth form specify that i is a full observer of a if φ is true. Statements of the sixth form specify that i is a partial observer of a if φ is true. An agent that is neither a full observer nor a partial observer is oblivious.

It is assumed that an action theory $\mathcal{T}$ is consistent with respect to every Kripke structure $M = \langle W, V, R_1, \ldots, R_n\rangle$, *i.e.*, contains no two statements specifying contradictory effects or observability. Thus, for every world $u \in W$ and every pair of statements "a causes p if ψ" and "a causes $\neg p$ if ψ'" in $\mathcal{T}$, $\langle M,u\rangle \not\models \psi \wedge \psi'$, and for every pair of statements "i observes a if φ" and "i aware_of a if φ'" in $\mathcal{T}$, $\langle M,u\rangle \not\models \varphi \wedge \varphi'$. It is further assumed that each action is precisely either an ontic action, a sensing action, or an announcement action. Thus, for any a, $\mathcal{T}$ contains exclusively either "a causes l if ψ" (a is an ontic action) or "a determines φ" (a is a sensing action) or "a announces φ" (a is an announcement action). It is assumed that every action is associated with at most one statement of type 1, "executable a if ψ", and we will say that the *precondition* of a is ψ (or $\top$ if there is no such statement).

An event-model-based semantics takes an m$\mathcal{A}^*$ action theory, an action name, and a state and builds an event template which, when applied in that state, produces a new state expressing the effects of the action. We direct the reader to (Baral et al. 2022) for a full discussion of m$\mathcal{A}^*$, including the original semantics for compiling action definitions into event models.

The *move-marble* action in Example 1, whereby Anne moves the marble to the box, is defined in the following action theory:

- *move-marble* causes $\neg d$
- A observes *move-marble*
- S observes *move-marble* if w

The action of Anne moving the marble to the box occurs when both Anne and Sally know that the marble is in the basket (d), Sally is watching Anne (w), Anne does not know that Sally is watching, and Sally knows that Anne does not know that Sally is watching.

3 m$\mathcal{A}^*$ Semantics with Local Observability

In this section, we present a novel event model semantics for m$\mathcal{A}^*$ that solves the second-order false-belief task. Our approach is similar to that developed in the original paper of m$\mathcal{A}^*$ by (Baral et al. 2022). We assume a consistent theory $\mathcal{T}$. Let $\langle M,s\rangle$ be a state where $M = \langle W, V, R_1, \ldots, R_n\rangle$ and a be an action. We will develop an event model that characterizes the occurence of a in $\langle M,s\rangle$. We start with a discussion of the intuition of this approach.

3.1 Intuition

Let us assume that a is an ontic action (*e.g.*, the *move-marble* by Anne). In general, at the first level of observability, the execution of an ontic action creates two groups of agents. The first group includes agents who observe the action occurrence and the second group consists of agents who do not observe the ation occurrence[5]. m$\mathcal{A}^*$ creates an event

[5]There can be other types of agents, e.g., agents who are uncertain whether the action occurs. We leave the consideration of this type of agents for the future as m$\mathcal{A}^*$ did not consider it.

model $\Sigma = \langle \{\theta, \epsilon\}, R_1^\Sigma, \ldots, R_n^\Sigma, pre, sub\rangle$ as shown in Figure 2 where F and O denote the set of full observers and oblivious agents, respectively. This event model is first instantiated at $\langle M, s\rangle$, *i.e.*, the set F and O are computed by evaluating the observability statements in $\mathcal{T}$ at the world s; and then used to compute the product update induced by Σ and M, the result of the execution of a in $\langle M, s\rangle$.

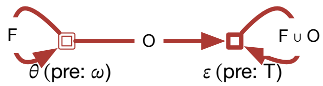

Figure 2: Generic event model for ontic actions (sub is omitted) in m$\mathcal{A}^*$

Intuitively, to take into consideration the uncertainty of agents' observability of other agents, we can instantiate this generic model at each world of M. For each $u \in W$, there might be two events θ_u and ϵ_u associated with it, representing the view of full observers (θ_u) and the view of oblivious agents (ϵ_u), respectively. The questions that we need to answer are

- What are the events of the final event model?
- What is the precondition of each event?
- What are the links between the events?
- What is the substitution of each event?

To answer these questions, let us observe that the instantiation of ϵ_u at any world u should be identical for all worlds in W, *i.e.*, for each $i \in \mathcal{G}$ and $u \in W$, there is a loop labeled i at ϵ_u. Furthermore, the precondition for ϵ_u is $\top$ as it represents the event indicating that *nothing happened* for oblivious agents. As such, we could combine all of these events into a single event, say ϵ. The precondition and substitution of this event are obviously $\top$ and $\emptyset$, respectively.

We next will focus on the collection of events $\{\theta_u \mid u \in W\}$. First, we notice that each θ_u can be characterized by the set of full observers; given u, this set can be characterized by the formula, say ω_u from statements of the form *"i observes a if φ"*. The precondition of an event must be then the executability condition of the action together with this formula. For an agent i, there is a link between θ_u and θ_v whenever i is a full observer in both u and v. Furthermore, if i is a full observer in u and oblivious in v, then there is a link labeled i from θ_u to ϵ.

Similar considerations should be made when an announcement or sensing action occurs. This intuition gives rise to the following constructions of the event models for the occurrence of a in $\langle M, s\rangle$. As usual, we will need to distinguish between ontic and epistemic actions.

3.2 Formal Definition

Consider action a defined by theory $\mathcal{T}$ applied in state $\langle M, s\rangle$ where $M = \langle W, V, R_1, \ldots, R_n\rangle$. For each $u \in W$, let

$$F(u) := \{i \mid \exists \text{"}i \text{ observes } a \text{ if } \varphi\text{"} \in \mathcal{T}.u \models \varphi\}$$
$$P(u) := \{i \mid \exists \text{"}i \text{ aware\_of } a \text{ if } \varphi\text{"} \in \mathcal{T}.u \models \varphi\}$$
$$O(u) := \mathcal{G} \setminus F(u) \cup P(u) \quad (4)$$

We refer to agents in $F(u)$, $P(u)$, and $O(u)$ as fully observers, partially observers, and oblivious agents, respectively. Under the assumptions in m$\mathcal{A}^*$, the sets $F(u)$, $P(u)$, and $O(u)$ are pairwise disjoint. Furthermore, the set of partial observers $P(u)$ is empty if a is an ontic action; and, if a is a sensing or announcement action then partial observers know *that* something has been sensed or announced, but not *what* was sensed or announced. To characterize the triple $\langle F(u), P(u), O(u)\rangle$, we define

$$\Omega(F(u), P(u), O(u)) :=$$
$$\Big(\bigwedge_{i \in F(u)} \bigvee_{\text{"}i \text{ observes } a \text{ if } \varphi\text{"} \in \mathcal{T}} \varphi\Big) \wedge$$
$$\Big(\bigwedge_{i \in P(u)} \bigvee_{\text{"}i \text{ aware\_of } a \text{ if } \varphi\text{"} \in \mathcal{T}} \varphi\Big) \wedge$$
$$\Big(\bigwedge_{i \in O(u)} \bigvee_{\text{"}i \text{ observes } a \text{ if } \varphi\text{"} \in \mathcal{T} \cup \text{"}i \text{ aware\_of } a \text{ if } \varphi\text{"} \in \mathcal{T}} \neg\varphi\Big)$$

It is easy to see that for each $u \in W$, $\Omega(F(u), P(u), O(u))$ is consistent if $\mathcal{T}$ is consistent. From now on, we will use $\Omega(u)$ and $\Omega(F(u), P(u), O(u))$ interchangeably if no confusion is possible.

For later use, given a state $\langle M, u\rangle$ and an agent i, we define

$$F^i(M, u) := \{j \mid \exists \text{"}j \text{ observes } a \text{ if } \varphi\text{"} \in \mathcal{T}. \langle M, u\rangle \models \mathbf{B}_i\varphi\}$$
$$P^i(M, u) := \{j \mid \exists \text{"}j \text{ aware\_of } a \text{ if } \varphi\text{"} \in \mathcal{T}. \langle M, u\rangle \models \mathbf{B}_i\varphi\}$$
$$O^i(M, u) := \{j \in \mathcal{G} \mid \forall \text{"}j \text{ observes } a \text{ if } \varphi\text{"} \in \mathcal{T}. \langle M, u\rangle \models \mathbf{B}_i\neg\varphi, \forall \text{"}j \text{ aware\_of } a \text{ if } \varphi\text{"} \in \mathcal{T}. \langle M, u\rangle \models \mathbf{B}_i\neg\varphi\}$$

Intuitively, $F^i(M, u)$, $P^i(M, u)$, and $O^i(M, u)$ are the sets of agents that agent i believes to be full observers, partial observers, and oblivious agents, respectively, in state $\langle M, u\rangle$. Note that some agents might not be in any of these categories: $F^i(M, u) \cup P^i(M, u) \cup O^i(M, u) \subseteq \mathcal{G}$. However, if u is serial, *i.e.*, $\exists v \in W. \langle u, v\rangle \in R_i$, then because of the requirement against contradictory observability statements in $\mathcal{T}$, these sets are non-intersecting: $F^i(M, u) \cap P^i(M, u) = F^i(M, u) \cap O^i(M, u) = P^i(M, u) \cap O^i(M, u) = \varnothing$.

Event Models for Ontic Actions Let a be an ontic action with precondition ψ. The execution of a in state $\langle M, s\rangle$ where $M = \langle W, V, R_1, \ldots, R_n\rangle$, induces an event template $\langle \Sigma, \Gamma\rangle$ where $\Sigma = \langle E, R_1^\Sigma, \ldots, R_n^\Sigma, pre, sub\rangle$ that is defined as follows. The set of events is

$$E = \{\theta_{\Omega(u)} \mid u \in W\} \cup \{\epsilon\},$$

and the designated event is $\Gamma = \{\theta_{\Omega(s)}\}$. Intuitively, each θ-event corresponds to the occurrence of the action in a specific world and the ϵ-event represents the action not occurring. The event relations give observers access only to the θ-events, and give oblivious agents access only to the ϵ-event.

Furthermore, for every $i \in \mathcal{G}$, $\langle x, y\rangle \in R_i^{\Sigma}$ iff

$$\begin{array}{l}\langle x,y\rangle = \langle \theta_{\Omega(u)}, \theta_{\Omega(v)}\rangle \wedge i \in F(u) \text{ or}\\ \langle x,y\rangle = \langle \theta_{\Omega(u)}, \epsilon\rangle \wedge i \in O(u) \text{ or}\\ \langle x,y\rangle = \langle \epsilon, \epsilon\rangle \wedge i \in \mathcal{G}\end{array} \quad (5)$$

Note that the last line states that, for every agent, there is a loop at the event encoding the non-occurrence of the action. Event preconditions constrain events to occur only in worlds conforming to their observability partition. Therefore, for every $\theta_{\Omega(u)} \in E$,

$$pre(\theta_{\Omega(u)}) = \psi \wedge \Omega(u)$$

and

$$pre(\epsilon) = \top.$$

The substitution function alters valuations according to ontic effects. For every $\theta_{\Omega(u)} \in E$,

$$sub(\theta_{\Omega(u)}) = \{p \leftarrow \Psi^+(p,a) \vee (p \wedge \neg\Psi^-(p,a)) \mid p \in \mathcal{F}\}, \quad (6)$$

where

$$\Psi^+(p,a) = \bigvee_{\text{“}a \text{ causes } p \text{ if } \varphi\text{”} \in \mathcal{T}} \varphi, \quad (7)$$

and

$$\Psi^-(p,a) = \bigvee_{\text{“}a \text{ causes } \neg p \text{ if } \varphi\text{”} \in \mathcal{T}} \varphi. \quad (8)$$

For the epsilon event,

$$sub(\epsilon) = \varnothing. \quad (9)$$

That is, a fluent p will be true if an ontic effect makes it true, or if it was already true and no ontic effect makes it false.

Figure 3 depicts the event model induced by *move-marble* in the state presented on the left of Figure 1. Each square represents an event, double borders indicate designated events, and arrows represent the event relation for each agent. Event names are shown with their preconditions next to the events where $\Omega(s) = w$ and $\Omega(u) = \neg w$. The substitution for each event is as follows: $sub(\theta_{\Omega(s)}) = sub(\theta_{\Omega(u)}) = \{d \leftarrow \bot, w \leftarrow w\}$, indicating that the marble will be in the box if the action *move-marble* actual occurs, and $sub(\epsilon) = \emptyset$ that indicates that nothing changes if the action does not occur. The link labeled S from $\Omega(s)$ to $\Omega(u)$ exists because S is a full observer of the action *move-marbe* in s. However, there is no link labeled S from $\Omega(u)$ to $\Omega(s)$ because S is oblivious of the action *move-marbe* in u. On the other hand, there are bi-directional link labeled A between $\Omega(u)$ to $\Omega(s)$ because A is a full observer of the action *move-marbe* in both s and u.

Figure 4 shows the state resulting from the application of *move-marble*. The world labels are $1 = (s, \theta_{\Omega(s)})$, $2 = (u, \theta_{\Omega(u)})$, $3 = (s, \epsilon)$, and $4 = (u, \epsilon)$. The marble is in the box ($\neg d$), and Sally and Anne both know it, but Anne believes that Sally believes that the marble is not in the box, *i.e.*, $\mathbf{B}_A\neg d \wedge \mathbf{B}_S\neg d \wedge \mathbf{B}_A\mathbf{B}_S d$.

We next prove some properties of the transitions between states of ontic actions[6].

[6]Proofs for Propositions 1-3 are in Appendix 6 because these propositions are similar to propositions proved in (Baral et al. 2022).

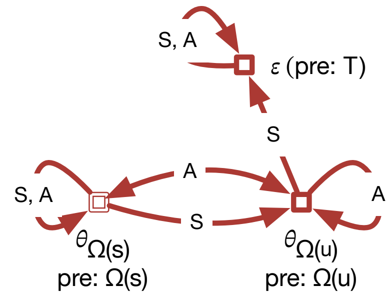

Figure 3: The event model induced by the application of the *move-marble* action in the state on the left of Figure 1.

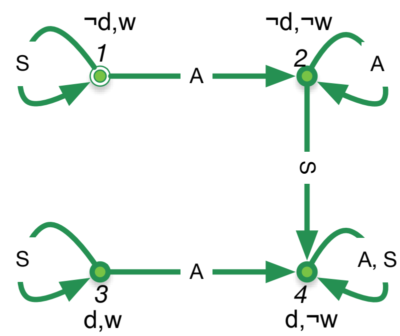

Figure 4: The state resulting after the execution of *move-marble* in the state on the left of Figure 1.

Proposition 1. *Assume that $\mathcal{T}$ contains the statement “a causes l if φ” (a is an ontic action), a is executable in $\langle M, s\rangle$, and its execution results in state $\langle M', s'\rangle$, then it holds that*

1. *Ontic effects alter the state if their conditions are met: if $\langle M, s\rangle \models \varphi$ then $\langle M', s'\rangle \models l$;*
2. *Observers learn about effects they believe occur: for every $i \in F(s)$, if $\langle M, s\rangle \models \boldsymbol{B}_i\varphi$ then $\langle M', s'\rangle \models \boldsymbol{B}_i l$;*
3. *Oblivious agents are unaffected by ontic events: for every $i \in O(s)$ and $\omega \in \mathcal{L}^{\mathcal{P}}.[\langle M', s'\rangle \models \boldsymbol{B}_i\omega$ iff $\langle M, s\rangle \models \boldsymbol{B}_i\omega]$;*
4. *Observers know that observers learn ontic effects: for every $i \in F(s)$ and $j \in F^i(M, s)$, if $\langle M, s\rangle \models \boldsymbol{B}_i\boldsymbol{B}_j\varphi$ then $\langle M', s'\rangle \models \boldsymbol{B}_i\boldsymbol{B}_j l$; and*
5. *Observers know that oblivious agents are unaffected: for $i \in F(s)$, $j \in O^i(M, s)$, $\omega \in \mathcal{L}^{\mathcal{P}}.[\langle M', s'\rangle \models \boldsymbol{B}_i\boldsymbol{B}_j\omega$ if $\langle M, s\rangle \models \boldsymbol{B}_i\boldsymbol{B}_j\omega]$.*

Proposition 1 shows that ontic effects conditionally alter the task environment as specified by $\mathcal{T}$, that observers learn about the action's effects upon the task environment and upon the mental states of other agents, and that oblivious agents do not. It is worth mentioning that Items 4 and 5 of this proposition are not discussed in (Baral et al. 2022). We will now turn our attention to sensing and announcement actions.

Event Models for Sensing and Announcement Actions
Let a be a sensing or announcement action whose sensed

or announced formula is φ, respectively. Furthermore, ψ be the precondition of a and $\langle M, s\rangle$ be a state where $M = \langle W, V, R_1, \dots, R_n\rangle$. Following m$\mathcal{A}^*$ we assume that $\langle M, s\rangle \models \psi$ if a is an announcement action[7]. The execution of a in $\langle M, s\rangle$ induces an event template $\langle \Sigma, \Gamma\rangle$ where $\Sigma = \langle E, R_1^\Sigma, \dots, R_n^\Sigma, pre, sub\rangle$ defined as follows. The set of events is

$$E = \{\theta_{\Omega(u)} \mid u \in W\} \cup \{\tau_{\Omega(u)} \mid u \in W\} \cup \{\epsilon\}.$$

For sensing actions, the designated events are

$$\Gamma = \{\theta_{\Omega(s)}, \tau_{\Omega(s)}\}.$$

For announcement actions, the designated event is

$$\Gamma = \{\theta_{\Omega(s)}\}.$$

Intuitively, each θ-event corresponds to the action occurring with a specific observability partition (the sensed formula or the announced formula is true), each τ-event expresses partial observers' uncertainty or false beliefs about an observed or announced formula given a specific observability partition, and the ϵ-event represents the action not occurring.

The relations in the event model Σ give full observers access from θ-events only to θ-events, and from τ-events only to τ-events. Partial observers also have access between θ-events and τ-events, expressing their uncertainty about what has been observed or announced. Oblivious agents access only the ϵ event. For each i in $\mathcal{G}$, $\langle x, y\rangle \in R_i^\Sigma$ iff

$$\begin{array}{l}
\langle x, y\rangle = \langle \theta_{\Omega(u)}, \theta_{\Omega(v)}\rangle \wedge i \in F(u) \cup P(u) \text{ or} \\
\langle x, y\rangle = \langle \tau_{\Omega(u)}, \tau_{\Omega(v)}\rangle \wedge i \in F(u) \cup P(u) \text{ or} \\
\langle x, y\rangle = \langle \theta_{\Omega(u)}, \tau_{\Omega(v)}\rangle \wedge i \in P(u) \text{ or} \\
\langle x, y\rangle = \langle \tau_{\Omega(u)}, \theta_{\Omega(v)}\rangle \wedge i \in P(u) \text{ or} \\
\langle x, y\rangle = \langle \theta_{\Omega(u)}, \epsilon\rangle \wedge i \in O(u) \text{ or} \\
\langle x, y\rangle = \langle \tau_{\Omega(u)}, \epsilon\rangle \wedge i \in O(u) \text{ or} \\
\langle x, y\rangle = \langle \epsilon, \epsilon\rangle
\end{array} \tag{10}$$

The event preconditions bind θ-events to worlds where the sensed or announced propositional formula φ holds, τ-events to worlds where φ does not hold, and both θ- and τ-events to worlds where precondition ψ holds and that conform to their respective agent observability partitions. For every θ-event $\theta_{\Omega(u)} \in E$

$$pre(\theta_{\Omega(u)}) = \psi \wedge \varphi \wedge \Omega(u),$$

for every τ-event $\tau_{\Omega(u)} \in E$

$$pre(\tau_{\Omega(u)}) = \psi \wedge \neg\varphi \wedge \Omega(u),$$

and

$$pre(\epsilon) = \top.$$

Since sensing and announcement actions have no ontic effects, for every event e in E,

$$sub(e) = \varnothing.$$

[7]This means that a is a truthful announcement. Dealing with untruthful announcements is an interested topic but is outside the cope of this paper.

Example 2 (Eavesdropping)**.** *A and B are in the same room and both do not know whether p and it is common knowledge between them. A receives a phone call from an outside agent who informs A that p is true. As usual, A believes that B does not pay attention to her day-to-day business and is therefore oblivious of her conversation. As such, A believes that B does not know about p after the phone call. However, B has a device secretly installed on A's phone and is monitoring her phone. Thus, B knows that p is true and knows that A believes that B does not know about p.*

This story represents the second-order false-belief task in the case of an announcement. It can be represented by the following statement.

- conversation *announces p*
- *A observes* conversation
- *B observes* conversation *if l*

where l denotes that A's phone has been hacked. The action conversation *occurs in a state depicted in Figure 5 where p, l are true in the true state of the world, A does not know whether p and has false belief about l. B knows l is true and knows that A does not know that.*

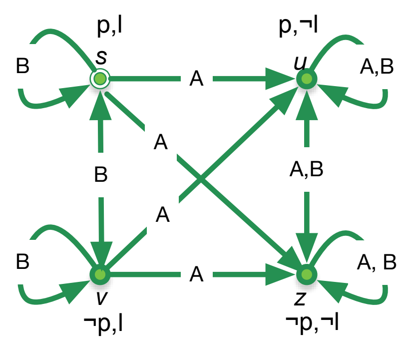

Figure 5: $\langle M, s\rangle$: State in which A receives the phone call (s is the true state of the world).

The event model for the occurrence of conversation *in* $\langle M, s\rangle$ *is shown in Figure 6. First, observe that* $\Omega(s) = \Omega(v) = l$ *because A and B are full observers in s and v and* $\Omega(u) = \Omega(z) = \neg l$ *because A is a full observer and B is oblivious in u and z. The precondition for* $\theta_{\Omega(s)}$, $\tau_{\Omega(s)}$, $\theta_{\Omega(u)}$, *and* $\tau_{\Omega(u)}$ *is* $p \wedge l$, $\neg p \wedge l$, $p \wedge \neg l$, *and* $\neg p \wedge \neg l$, *respectively. Observe that the event model does not contain links between events of the form* $\theta_{\Omega(x)}$ *and* $\tau_{\Omega(x)}$ *because both agents are full observers or oblivious in all worlds of* $\langle M, s\rangle$.

Figure 7 shows the state resulting from the occurrence of conversation *in the state given in Figure 5. Agents A and B both know p, but A believes that B does not know whether p.*

Similar to Proposition 1, we can prove the following propositions that show that the proposed method alter the beliefs of the agents in accordance to their observability.

Proposition 2. *Assume that* $\mathcal{T}$ *contains the statement "a determines* φ*" (a is a sensing action), a is executable in* $\langle M, s\rangle$*, and its execution results in* $\langle M', s'\rangle$*. It holds that*

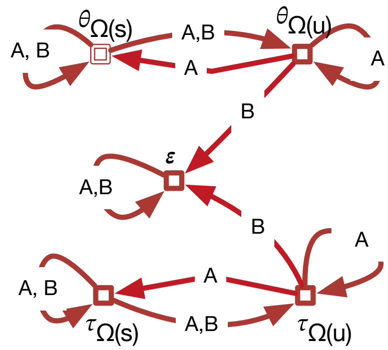

Figure 6: Event model induced by *conversation* in the state in Figure 5.

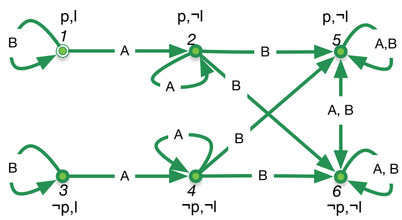

Figure 7: State after *conversation* occurred in the state in Figure 5.

1. *Full observers learn true sensed formulas: if* $\langle M, s\rangle \models \varphi^*$ *then* $\langle M', s'\rangle \models \boldsymbol{B}_i\varphi^*$ *for* $i \in F(s)$ *where* $\varphi^* \in \{\varphi, \neg\varphi\}$*;*
2. *Oblivious agents are unaffected by sensing events: for* $i \in O(s), \omega \in \mathcal{L}^{\mathcal{P}}.[\langle M', s'\rangle \models \boldsymbol{B}_i\omega$ *iff* $\langle M, s\rangle \models \boldsymbol{B}_i\omega]$*;*
3. *Full and partial observers learn that full observers learn sensed formulas: for* $i \in F(s) \cup P(s), j \in F^i(M, s).[\langle M', s'\rangle \models \boldsymbol{B}_i\big((\varphi \wedge \boldsymbol{B}_j\varphi) \vee (\neg\varphi \wedge \boldsymbol{B}_j\neg\varphi)\big)]$*;*
4. *Full and partial observers know that oblivious agents are unaffected: for* $i \in F(s) \cup P(s), j \in O^i(M, s), \omega \in \mathcal{L}^{\mathcal{P}}.[\langle M', s'\rangle \models \boldsymbol{B}_i\boldsymbol{B}_j\omega$ *if* $\langle M, s\rangle \models \boldsymbol{B}_i\boldsymbol{B}_j\omega]$*.*

The next proposition is similar to Proposition 2 but for announcement actions.

Proposition 3. *Assume that* $\mathcal{T}$ *contains the statement "a announces* φ*"(a is an announcement action), a is executable* $\langle M, s\rangle$*, and its execution results in state* $\langle M', s'\rangle$*. It holds that*

1. *Full observers learn announcements:* $\forall i \in F(s).[\langle M', s'\rangle \models \boldsymbol{B}_i\varphi]$*,*
2. *Oblivious agents are unaffected by announcements:* $\forall i \in O(s), \omega \in \mathcal{L}^{\mathcal{P}}.[\langle M', s'\rangle \models \boldsymbol{B}_i\omega$ *iff* $\langle M, s\rangle \models \boldsymbol{B}_i\omega]$*,*
3. *Full and partial observers learn that full observers learn announcements:* $\forall i \in F(s) \cup P(s), j \in F^i(M, s).[\langle M', s'\rangle \models \boldsymbol{B}_i\big((\varphi \wedge \boldsymbol{B}_j\varphi) \vee (\neg\varphi \wedge \boldsymbol{B}_j\neg\varphi)\big)]$*,*
4. *Full and partial observers know that oblivious agents are unaffected:* $\forall i \in F(s) \cup P(s), j \in O^i\langle M, s\rangle, \omega \in \mathcal{L}^{\mathcal{P}}.[\langle M', s'\rangle \models \boldsymbol{B}_i\boldsymbol{B}_j\omega$ *if* $\langle M, s\rangle \models \boldsymbol{B}_i\boldsymbol{B}_j\omega]$*.*

We observe that under the new semantics, it is no longer true that the sensed (or announced) formula becomes common knowledge among all full observers as in the original proposal of m$\mathcal{A}^*$. This is because under the new semantics, a full observer might not know about the observability of other agents (*e.g.*, A does not know that B is a full observer in Example 2).

4 Higher-Order False Beliefs

Propositions 1–3 show that under the new semantics, the transition function between states exhibit similar properties as the original m$\mathcal{A}^*$ semantics. The key difference between the proposed semantics and the original semantics is in higher order beliefs of agents. For example, Item 4 in Proposition 1 is not present in the theorem discussing properties of the semantics for ontic actions in (Baral et al. 2022) (Theorem 2); Item 5 in Proposition 1 is about the belief of a full observer about the belief of an oblivious agent with respect to its observability while Item 3 in Theorem 2 of (Baral et al. 2022) does not take into account the local observability of full observers; etc. Similar observations can be made for sensing and announcement actions, for example, because full observers might be uncertain about the observability of other full observers, the sensed (or announced) formula is no longer common knowledge among full observers under the new semantics (as stated in Theorem 3 of (Baral et al. 2022)). This is the main reason that m$\mathcal{A}^*$ cannot deal with the second order false-belief task, as demonstrated in Example 1. We first show that the new semantics indeed solves this problem.

Proposition 4. *Let* $\mathcal{T}$ *be a theory,* a *an action, and* $\langle M, s\rangle$ *a state. Assume that* ψ *is the precondition of* a *that is executable in* $\langle M, s\rangle$ (i.e., $\langle M, s\rangle \models \psi$*),* $i, j \in F(s)$*,* $j \in O^i(M, s)$*, and* $\langle M', s'\rangle$ *is the result of the execution of a in* $\langle M, s\rangle$*. Then,*

1. *If* $\mathcal{T}$ *contains* "a causes l if λ"*,* $\langle M, s\rangle \models \boldsymbol{B}_i\lambda \wedge \boldsymbol{B}_j\lambda$*, and* $\langle M, s\rangle \models \boldsymbol{B}_i\boldsymbol{B}_j\neg l$ *then* $\langle M', s'\rangle \models \boldsymbol{B}_i l \wedge \boldsymbol{B}_j l \wedge \boldsymbol{B}_i\boldsymbol{B}_j\neg l$*;*
2. *If* $\mathcal{T}$ *contains "a determines* φ*",* $\langle M, s\rangle \models \varphi^*$*, and* $\langle M, s\rangle \models \boldsymbol{B}_i\boldsymbol{B}_j\neg\varphi^*$ *then* $\langle M', s'\rangle \models \boldsymbol{B}_i\varphi^* \wedge \boldsymbol{B}_j\varphi^* \wedge \boldsymbol{B}_i\boldsymbol{B}_j\neg\varphi^*$ *for* $\varphi^* \in \{\varphi, \neg\varphi\}$*;*
3. *If* $\mathcal{T}$ *contains "a announces* φ*"and* $\langle M, s\rangle \models \boldsymbol{B}_i\boldsymbol{B}_j\neg\varphi$ *then* $\langle M', s'\rangle \models \boldsymbol{B}_i\varphi \wedge \boldsymbol{B}_j\varphi \wedge \boldsymbol{B}_i\boldsymbol{B}_j\neg\varphi$*;*
4. *If* $\mathcal{T}$ *contains "a determines* φ*",* $\langle M, s\rangle \models \varphi^*$*, and* $\langle M, s\rangle \models \boldsymbol{B}_i(\neg(\boldsymbol{B}_j\varphi^* \vee \boldsymbol{B}_j\neg\varphi^*))$ *then* $\langle M', s'\rangle \models \boldsymbol{B}_i\varphi^* \wedge \boldsymbol{B}_j\varphi^* \wedge \boldsymbol{B}_i(\neg(\boldsymbol{B}_j\varphi^* \vee \boldsymbol{B}_j\neg\varphi^*))$ *for* $\varphi^* \in \{\varphi, \neg\varphi\}$*; and*
5. *If* $\mathcal{T}$ *contains "a determines* φ*"and* $\langle M, s\rangle \models \boldsymbol{B}_i(\neg(\boldsymbol{B}_j\varphi \vee \boldsymbol{B}_j\neg\varphi))$ *then* $\langle M', s'\rangle \models \boldsymbol{B}_i\varphi \wedge \boldsymbol{B}_j\varphi \wedge \boldsymbol{B}_i(\neg(\boldsymbol{B}_j\varphi \vee \boldsymbol{B}_j\neg\varphi))$*.*

Proof. The first item follows from of Proposition 1: $\langle M', s'\rangle \models \mathbf{B}_i l \wedge \mathbf{B}_j l$ because $i, j \in F(s)$ and $\langle M, s\rangle \models \mathbf{B}_i\lambda \wedge \mathbf{B}_j\lambda$ (Item 2); and $\langle M', s'\rangle \models \mathbf{B}_i\mathbf{B}_j\neg l$ because $j \in O^i(M, s)$ and $\langle M, s\rangle \models \mathbf{B}_i\mathbf{B}_j\neg l$ (Item 5).

Similarly, the second (third) item follows from Items 1 and 4 of Proposition 2 (Proposition 3).

To prove Item 4, we note that $\langle M', s'\rangle \models \mathbf{B}_i\varphi^* \wedge \mathbf{B}_j\varphi^*$ because $i, j \in F(s)$ (Item 1 of Proposition 2). It remains to be shown that if $\langle M, s\rangle \models \mathbf{B}_i(\neg(\mathbf{B}_j\varphi^* \vee \mathbf{B}_j\neg\varphi^*))$ then $\langle M', s'\rangle \models \mathbf{B}_i(\neg(\mathbf{B}_j\varphi^* \vee \mathbf{B}_j\neg\varphi^*))$.

To continue, w.l.o.g., assume that $\langle M, s\rangle \models \varphi$. Observe that because a is executable in $\langle M, s\rangle$, we have that $s' = \langle s, \theta_{\Omega(s)}\rangle \in W'$ (W' is the set of worlds in M').

Consider u' and v' such that $\langle s', u'\rangle \in R'_i$ and $\langle u', v'\rangle \in R'_j$. By construction of M' and Equation (10), we can conclude that there are some worlds u and v in M such that $u' = \langle u, \theta_{\Omega(u)}\rangle$, $v' = \langle v, \epsilon\rangle$, $\langle s, u\rangle \in R_i$, and $\langle u, v\rangle \in R_j$. Because $\langle M, s\rangle \models \mathbf{B}_i(\neg(\mathbf{B}_j\varphi \vee \mathbf{B}_j\neg\varphi))$, we can conclude that there exists $\langle u, z\rangle \in R_j$ such that $(\langle M, v\rangle \models \varphi \Rightarrow \langle M, z\rangle \models \neg\varphi)$ and $(\langle M, v\rangle \models \neg\varphi \Rightarrow \langle M, z\rangle \models \varphi)$. This implies that $z' = \langle z, \epsilon\rangle \in W'$ and $\langle u', z'\rangle \in R'_j$. Because a is a sensing action, the valuation of the world x' is bisimilar to that of x for $x \in \{s, u, v, z\}$, *i.e.*, at every u' there exist v' and z' such that $\langle u', v'\rangle \in R'_j$, $\langle u', z'\rangle \in R'_j$, and $(\langle M', v'\rangle \models \varphi \Rightarrow \langle M', z'\rangle \models \neg\varphi)$ and $(\langle M', v'\rangle \models \neg\varphi \Rightarrow \langle M', z'\rangle \models \varphi)$. This holds for every u' and thus $\langle M, s\rangle \models \mathbf{B}_i(\neg(\mathbf{B}_j\varphi \vee \mathbf{B}_j\neg\varphi))$.
The proof for Item 5 is similar to the proof for Item 4. □

Proposition 4 shows that there are two possible types of second-order false beliefs. The first one is represented by Items 1–3 where an agent believes that another agent has a false belief about a property of the world (*e.g.*, Example 1: the second order false-belief formula is $\mathbf{B}_{Ann}\mathbf{B}_{Sally}d$). The second one is represented by Items 4–5 where an agent believes that another agent does not know whether a property holds while the latter indeed knows the truth value of the property (*e.g.*, Example 2: the second order false-belief formula is $\mathbf{B}_A\neg(\mathbf{B}_Bp \vee \mathbf{B}_B\neg p)$).

It is worth noticing that (Bolander 2018) developed a DEL formalism based on edge-conditioned event models to deal with the second order false-belief tasks and proposed two criteria, *robustness* and *faithfulness*, for such a formalism. More precisely, these two criteria are stated as follows:

- **Robustness:** The formalism should not only be able to deal with one or two selected false-belief tasks, but with as many as possible, with no strict limit on the order of belief attribution.
- **Faithfulness:** Each action of the false-belief story should correspond to an action in the formalism in a natural way, and it should be fairly straightforward, not requiring ingenuity, to find out what that action of the formalism is.

The proposed formalism is robust in that it can work with any m$\mathcal{A}^*$ domain, i.e., it is not specifically tailored to any example. It is also faithful since the event model for an action occurrence can be automatically computed given the m$\mathcal{A}^*$-domain and a Kripke model, and thus, does not require any special attention.

5 Discussion and Conclusion

The present paper is strongly related to formalisms that focus on second-order false beliefs (SCFB) in epistemic planning context or works that extend or propose alternative to m$\mathcal{A}^*$. (Bolander 2018) is probably the first paper[8] that discussed the SCFB task in the context of epistemic planning. It points out that the use of event-models in m$\mathcal{A}^*$ cannot address the SCFB task as well as the importance of observability of the form '*who sees who*' in reasoning about beliefs and knowledge. The main difference between the work by (Bolander 2018) and the present paper is in the use of DEL and action language. (Bolander 2018) introduced the notion of an edge-conditioned event model and showed that it can be used, together with the encoding of the observations into the states, to formalize the SCFB tasks in several examples. He also proposed the notions of robustness and faithfulness for characterizing formalisms that deal with SCFB tasks. We discuss these properties and show that SCFB tasks can be formalized using m$\mathcal{A}^*$ under the proposed semantics.

In a recent paper, (Engesser, Herzig, and Perrotin 2024) proposed a specification language for reasoning about actions with knowledge and belief, called *repetition-free epistemic-doxastic* (REDA), that can deal with SCFB. The investigation focuses on the fragment REDA$^{\leq 2}$, *i.e.*, the set of formulas of modal depth at most two. Unlike several action languages that only deal with one modal operator (belief), the language introduced in (Engesser, Herzig, and Perrotin 2024) works with two modal operators, belief and knowledge and assumes that the belief operator is serial, transitive, and euclidean. On the other hand, at the specification level, both REDA$^{\leq 2}$ and m$\mathcal{A}^*$ are similar in that they only allow for the specification of effects of actions using formulas of modal depth at most two (in m$\mathcal{A}^*$, effects are assumed to be propositional formula). As our focus in this paper is on higher-oder local observability, we considered only the belief operator. We note that the separation work by (Buckingham, Kasenberg, and Scheutz 2020) also considers two modal operators for an m$\mathcal{A}^*$-like language and several authors (*e.g.*, (Aucher 2008; Son et al. 2015; Son, Pham, and Pontelli 2024)) discussed conditions under which languages with event model based semantics can reason about knowledge and belief. As our investigation in this paper focuses on the construction of event-models to cope with SCFB, it is not clear to us whether the proposed semantics can maintain the KD45 property of states after action occurrences. It will be a topic of our future investigation.

In (Rajaratnam and Thielscher 2021), the authors proposed an action language, called *Dynamic Epistemic Representation* (DER), for representing and reasoning with event models for epistemic planning. In DER, observations are parameterized with agents which indicate the ownership of observations. Action effects, both on the state of the world or on the beliefs of agents, can be specified by a single type of statements of the form "a causes φ if ψ." DER also has an event-model based semantics, *i.e.*, each action specification is translated into an event-model. The results of the execution of an action in a state is then defined by the usual product update operator. We notice that DER, however, cannot work with the SCFB task as well because the construction of event models does not take into consideration agents' uncertainty about other agents' observability (*e.g.*, the event model constructed for the Sally-Anne example consists of only two events similar to that of m$\mathcal{A}^*$).

An edge-conditioned event-model based semantics for m$\mathcal{A}^*$ has been proposed in (Pham et al. 2022). It appears that this semantics can also deal with the SCFB tasks

[8]This is an extension of an earlier paper.

as it is a consequence of Proposition 1 in (Pham et al. 2022) even though a formal proof was not presented. The event-model constructed following this approach has only two events as in the original semantics of m$\mathcal{A}^*$. However, agents' accessibility relation between events are conditioned on the worlds. It is worth noticing that other modifications and extensions of m$\mathcal{A}^*$ have been proposed. However, these modifications aim at addressing the belief correction problem of m$\mathcal{A}^*$ (Izmirlioglu et al. 2022a; Izmirlioglu et al. 2022b); other consider different types of actions such as non-deterministic actions (Pham, Son, and Pontelli 2023b) or untruthful announcements (Pham, Son, and Pontelli 2022). Most of these extensions, however, employ edge-conditioned event-models. It will be interesting to see how these extensions can be considered under the semantics proposed in this paper. We leave this as a future research topic.

We observe that the proposed semantics can easily be implemented in systems that use m$\mathcal{A}^*$ as their specification language such as the system in (Le et al. 2018) by replacing the algorithm computing the event-models defined in (Baral et al. 2022) with the one proposed in Section 3. This will yield a system that can *reason* correctly with second-order false beliefs.

6 Conclusions and Future Work

We proposed a new semantics for m$\mathcal{A}^*$ that take into consideration higher-order action observability of agents in the construction of event-models encoding action occurrences. We proved that the new semantics changes agents' beliefs according to their observability and, more importantly, can properly deal with second-order false-belief tasks. We argued that the formalism—m$\mathcal{A}^*$ under this new semantics—satisfies two desirable criteria, *robustness* and *faithfulness*, as proposed by (Bolander 2018). We also illustrated the new definitions through the Sally-Anne second-order false-belief story as well as a new example and related to works formalizing second-order false-beliefs in epistemic planning context. Following the discussion by (Verbrugge 2023), we will utilize our formalism and focus our attention on developing computational tools for explaining behaviors in false-belief tasks and, potentially, detecting deceptive behaviors in the near future.

Acknowledgement

The first two authors would like to acknowledge the partial support from the ARO grant W911NF2220252. The third author would like to acknowledge the partial support from the NSF grants 1914635 and 2151254 and a contract from Wallaroo.

Appendix A: Technical Proofs

A.1. Ontic Actions

Let $\mathcal{T}$ be a theory, a an ontic action with precondition ψ and $\langle M, s\rangle$ a state such that $\langle M, s\rangle \models \psi$. Furthermore, assume that $\mathcal{T}$ contains the statement "a causes l if φ". In the following, we prove some lemma related to the occurrence of a in $\langle M, s\rangle$, which results in $\langle M', s'\rangle$. Let us denote the worlds of M and M' by W and W', respectively. Proposition 1 is proved by the following lemmae.

Lemma 1. *Let $x \in W$. If $\langle M, x\rangle \models \psi$ then $\langle x, \theta_{\Omega(x)}\rangle \in W'$. Furthermore, if $\langle M, x\rangle \models \psi \wedge \varphi$ then $\langle x, \theta_{\Omega(x)}\rangle \models l$.*

Proof. Because $\langle M, x\rangle \models \psi$ and $pre(\theta_{\Omega(x)}) = \psi \wedge \Omega(x)$, we have that $x' = \langle x, \theta_{\Omega(x)}\rangle \in W'$. By definition, the valuation of $\mathcal{P}$ assigned to x' is obtained by applying the substitution $sub(\theta_{\Omega(x)})$ to x. From Equation (6) we know that $l \leftarrow \varphi$ belongs to $sub(\theta_{\Omega(x)})$, and thus, if $\langle M, x\rangle \models \varphi$ then $\langle M', x'\rangle \models l$ which implies $x' \models l$ as l is a literal. ☐

Lemma 2. *Let $x \in W$. If $\langle M, x\rangle \models \psi$, $i \in F(x)$, and $\langle M, x\rangle \models \boldsymbol{B}_i\varphi$ then $x' = \langle x, \theta_{\Omega(x)}\rangle \in M'$ and $\langle M', x'\rangle \models \boldsymbol{B}_i l$.*

Proof. $x' = \langle x, \theta_{\Omega(x)}\rangle \in W'$ by Lemma 1. Consider $\langle x', u'\rangle \in R'_i$. By Equation (5) (first line), we can conclude that $u' \neq \epsilon$. Therefore $u' = \langle u, e\rangle$ for some $u \in W$. It follows from Equations (2) and (5) that $\langle x, u\rangle \in R_i$, $\langle \theta_{\Omega(s)}, e\rangle \in R_i^{\Sigma}$, $\langle M, u\rangle \models \psi$, and e is a θ-event, *i.e.*, $e = \theta_{\Omega(u)}$. This means that we apply $sub(\theta_{\Omega(u)})$ to u. Since $\langle M, x\rangle \models \mathbf{B}_i\varphi$ and $\langle x, u\rangle \in R_i$, we can derive that $\langle M, u\rangle \models \varphi$ (from definition of $\models$). Thus, $\langle M', u'\rangle \models l$ (Lemma 1). This holds for every u' such that $\langle x', u'\rangle \in R'_i$. Therefore, $\langle M', x'\rangle \models \mathbf{B}_i l$. ☐

Lemma 3. *Let $x \in W$. If $\langle M, x\rangle \models \psi$, and $i \in O(x)$ then $x' = \langle x, \theta_{\Omega(x)}\rangle \in M'$ and, for every formula $\omega \in \mathcal{L}^{\mathcal{P}}$, $\langle M, x\rangle \models \boldsymbol{B}_i\omega$ iff $\langle M', x'\rangle \models \boldsymbol{B}_i\omega$.*

Proof. $x' = \langle x, \theta_{\Omega(x)}\rangle \in W'$ by Lemma 1. Consider an agent $i \in O(x)$ and $\langle x', u'\rangle \in R'_i$. Again, from Equations (2) and (5) (second line) we can conclude that $u' = \langle u, \epsilon\rangle$ for some $u \in W$ and $\langle x, u\rangle \in R_i$. Because $\sigma(\epsilon) = \emptyset$, the valuation over $\mathcal{P}$ assigned to u' is bisimilar to that of u. This holds for every u' such that $\langle x', u'\rangle \in R'_i$. (*)

Now consider v such that $\langle x, v\rangle \in R_i$. Because $i \in O(s)$, we have that $\langle\langle x, \theta_{\Omega(s)}\rangle, \langle v, \epsilon\rangle\rangle \in R'_i$ and v is bisimilar to $\langle v, \epsilon\rangle$. This holds for every v such that $\langle x, v\rangle \in R_i$. (**)

(*) and (**) imply the conclusion of the lemma. ☐

Lemma 4. *Let $x \in W$. If $\langle M, x\rangle \models \psi$, $i \in F(x)$ and $j \in F^i(M, x)$, and $\langle M, x\rangle \models \boldsymbol{B}_i\boldsymbol{B}_j\varphi$ then $\langle M', x'\rangle \models \boldsymbol{B}_i\boldsymbol{B}_j l$ for $x' = \langle x, \theta_{\Omega(x)}\rangle$.*

Proof. Again, $x' = \langle x, \theta_{\Omega(x)}\rangle \in W'$ by Lemma 1. Consider some $\langle x', u'\rangle \in R'_i$. Similarly to Lemma 2, we have that $u' = \langle u, \theta_{\Omega(u)}\rangle$ for some $u \in W$ and $\langle x, u\rangle \in R_i$. Since $j \in F^i(M, x)$, it follows that $j \in F(u)$. If $\langle M, x\rangle \models \mathbf{B}_i\mathbf{B}_j\varphi$, then $\langle M, u\rangle \models \mathbf{B}_j\varphi$. Applying Lemma 2 for u, j, and l implies that $\langle M', u'\rangle \models \mathbf{B}_j l$. Thus, $\langle M', x'\rangle \models \mathbf{B}_i\mathbf{B}_j l$. ☐

Lemma 5. *Let $x \in W$. Assume $\langle M, x\rangle \models \psi$, $i \in F(x)$, $j \in O^i(M, x)$, and $\omega \in \mathcal{L}^{\mathcal{P}}$. Then, $\langle M', s'\rangle \models \boldsymbol{B}_i\boldsymbol{B}_j\omega$ if $\langle M, x\rangle \models \boldsymbol{B}_i\boldsymbol{B}_j\omega$.*

Proof. Again, $x' = \langle x, \theta_{\Omega(x)}\rangle \in W'$ by Lemma 1.

Consider some $\langle x', u'\rangle \in R'_i$. Similarly to Lemma 2, we have that $u' = \langle u, \theta_{\Omega(u)}\rangle$ for some $u \in W$ and $\langle x, u\rangle \in R_i$. Since $j \in O^i(M, x)$, it follows that $j \in O(u)$. If $\langle M, x\rangle \models \mathbf{B}_i\mathbf{B}_j\omega$, then $\langle M, u\rangle \models \mathbf{B}_j\omega$. Applying Lemma 3 for u, j, and ω implies that $\langle M', u'\rangle \models \mathbf{B}_j\omega$. Thus, $\langle M', x'\rangle \models \mathbf{B}_i\mathbf{B}_j\omega$. ☐

A.2. Sensing and Announcement Actions

Proof for Proposition 2.

1. Assume that $\langle M, s\rangle \models \varphi$. Because $\langle M, s\rangle \models \varphi$ and a is executable in $\langle M, s\rangle$, we have that the designated event s' of M' is $\langle s, \theta_{\Omega(s)}\rangle$. Consider agent $i \in F(s)$ and $\langle s', u'\rangle \in R'_i$. Since $s' = \langle s, \theta_{\Omega(s)}\rangle$ and $i \in F(s)$, Equation (10) (first line) implies that $u' = \langle u, \theta_{\Omega(u)}\rangle$ for some $u \in W$. This implies that $pre(\theta_{\Omega(u)}) \models \varphi$, necessarily $u \models \varphi$, and since $sub(\theta_{\Omega(u)}) = \varnothing$, $u' \models \varphi$. This holds for every u' such that $\langle s', u'\rangle \in R'_i$, which implies that $(M', s') \models \mathbf{B}_i\varphi$. (The proof for $\langle M, s\rangle \models \neg\varphi$ is similar).
2. Similar to the proof of Item 3 of Proposition 1.
3. Assume that $\langle M, s\rangle \models \varphi$. Let $i \in F(s) \cup P(s)$, and $j \in F^i(M, s)$. Similar arguments to the proof of Item 1 allow us to conclude that $s' = \langle s, \theta_{\Omega(s)}\rangle$ Consider $\langle s', u'\rangle \in R'_i$. By Equations (10) and (2), u' is either a θ-world or a τ-world, *i.e.*, either $u' = \langle u, \theta_{\Omega(u)}\rangle$ or $u' = \langle u, \tau_{\Omega(u)}\rangle$ for some $u \in W$ for some $u \in W$ and $\langle s, u\rangle \in R_i$. Because $j \in F^i(M, s)$, we have that $j \in F(u)$. Similar arguments to the proof of Item 1, we can show that if $u' = \langle u, \theta_{\Omega(u)}\rangle$, $\langle M', u'\rangle \models \varphi \wedge \mathbf{B}_j\varphi$; and if $u' = \langle u, \tau_{\Omega(u)}\rangle$, $\langle M', u'\rangle \models \neg\varphi \wedge \mathbf{B}_j\neg\varphi$. In other words, we have that $\langle M', s'\rangle \models \mathbf{B}_i((\varphi \wedge \mathbf{B}_j\varphi) \vee (\neg\varphi \wedge \mathbf{B}_j\neg\varphi))$.
   The proof is similar for the case $\langle M, s\rangle \models \neg\varphi$ holds.
4. The proof is similar to the proof of Item 5 of Proposition 1 with the observations that (*i*) s' can be either $\langle s, \theta_{\Omega(s)}\rangle$ or $\langle s, \tau_{\Omega(s)}\rangle$ (it depends on whether $\langle M, s\rangle \models \varphi$, as shown in the proof of Item 1); (*ii*) for every u' such that $\langle s', u'\rangle \in R'_i$, u' can be either $\langle u, \theta_{\Omega(u)}\rangle$ or $\langle u, \tau_{\Omega(u)}\rangle$ for some $u \in W$ such that $\langle s, u\rangle \in R_i$ as $i \in F(s) \cup P(s)$; and (*iii*) $j \in O^i(M, s)$ implies that $j \in O(u)$. □

Proof for Proposition 3. The proof of this proposition is almost identical to the proof for Proposition 2, without the case $\langle M, s\rangle \models \neg\varphi$ as we assume that the announcement is truthful. □

References

Andersen, M. B.; Bolander, T.; and Jensen, M. H. 2012. Conditional epistemic planning. In del Cerro, L. F.; Herzig, A.; and Mengin, J., eds., *Logics in Artificial Intelligence*, 94–106. Berlin, Heidelberg: Springer Berlin Heidelberg.

Aucher, G. 2008. Consistency preservation and crazy formulas in BMS. In Hölldobler, S.; Lutz, C.; and Wansing, H., eds., *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*, 21–33. Springer.

Baltag, A., and Moss, L. 2004. Logics for epistemic programs. *Synthese*.

Baltag, A.; Moss, L.; and Solecki, S. 1998. The logic of public announcements, common knowledge, and private suspicions. In *7th TARK*, 43–56.

Baral, C.; Gelfond, G.; Pontelli, E.; and Son, T. C. 2022. An action language for multi-agent domains. *Artificial Intelligence* 302:103601.

Bolander, T., and Birkegaard Andersen, M. 2011. Epistemic planning for single and multi-agent systems. *Journal of Applied Non-classical Logics - JANCL* 21:9–34.

Bolander, T. 2017. A gentle introduction to epistemic planning: The DEL approach. *Electronic Proceedings in Theoretical Computer Science* 243:1–22.

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

Braüner, T.; Blackburn, P.; and Polyanskaya, I. 2016. Second-order false-belief tasks: Analysis and formalization. In *International Workshop on Logic, Language, Information, and Computation*, 125–144. Springer.

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

Engesser, T.; Bolander, T.; Mattümller, R.; and Nebel, B. 2017. Cooperative epistemic multi-agent planning for implicit coordination. *Electronic Proceedings in Theoretical Computer Science* 243:75–90.

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

Fagin, R.; Halpern, J. Y.; Moses, Y.; and Vardi, M. Y. 1995. *Reasoning About Knowledge*. Cambridge, MA, London, England: MIT Press.

Izmirlioglu, Y.; Pham, L.; Son, T. C.; and Pontelli, E. 2022a. A review of the muddy children problem. In Dovier, A.; Montanari, A.; and Orlandini, A., eds., *AIxIA 2022 - Advances in Artificial Intelligence - XXIst International Conference of the Italian Association for Artificial Intelligence, AIxIA 2022, Udine, Italy, November 28 - December 2, 2022, Proceedings*, volume 13796 of *Lecture Notes in Computer Science*, 127–139. Springer.

Izmirlioglu, Y.; Pham, L.; Son, T. C.; and Pontelli, E. 2022b. A robust state transition function for multi-agent epistemic systems. In Marreiros, G.; Martins, B.; Paiva, A.; Ribeiro, B.; and Sardinha, A., eds., *Progress in Artificial Intelligence - 21st EPIA Conference on Artificial Intelligence, EPIA 2022, Lisbon, Portugal, August 31 - September 2, 2022, Proceedings*, volume 13566 of *Lecture Notes in Computer Science*, 688–700. Springer.

Le, T.; Fabiano, F.; Son, T. C.; and Pontelli, E. 2018. EFP and PG-EFP: Epistemic Forward Search Planners in Multi-Agent Domains. In *International Conference on Automated Planning and Scheduling (ICAPS)*. AAAI Press.

Muise, C.; Belle, V.; Felli, P.; McIlraith, S. A.; Miller, T.; Pearce, A. R.; and Sonenberg, L. 2022. Efficient multi-agent epistemic planning: Teaching planners about nested belief. *Artif. Intell.* 302:103605.

Pham, L.; Izmirlioglu, Y.; Son, T. C.; and Pontelli, E. 2022. A new semantics for action language mA^*. In Aydogan, R.; Criado, N.; Lang, J.; Sánchez-Anguix, V.; and Serramia, M., eds., *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*, 553–562. Springer.

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

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

Pham, L.; Son, T. C.; and Pontelli, E. 2023b. Reasoning about non-deterministic observability and hypothetical action occurrences in multi-agent domains. In Hong, J.; Lanperne, M.; Park, J. W.; Cerný, T.; and Shahriar, H., eds., *Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing, SAC 2023, Tallinn, Estonia, March 27-31, 2023*, 957–964. ACM.

Rajaratnam, D., and Thielscher, M. 2021. Representing and reasoning with event models for epistemic planning. In Bienvenu, M.; Erdem, E.; and Lakemeyer, G., eds., *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning (KR)*, 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*.

Son, T. C.; Pham, L.; and Pontelli, E. 2024. On Dealing with False Beliefs and Maintaining KD45 Property. In *AAMAS*.

Van Ditmarsch, H.; van Der Hoek, W.; and Kooi, B. 2007. *Dynamic epistemic logic*, volume 337. Springer Science & Business Media.

Verbrugge, R. 2023. Reasoning about reasoning about reasoning: From logic to the lab. Invited Talk — KRR 2023.

Wan, H.; Fang, B.; and Liu, Y. 2021. A general multi-agent epistemic planner based on higher-order belief change. *Artif. Intell.* 301:103562.