Reasoning with Untruthful Announcements

Loc Pham[1], Tran Cao Son[1], and Enrico Pontelli[1]

New Mexico State University, Las Cruces NM 88003, USA
{locpham, stran, epontell}@nmsu.edu

Abstract. Untruthful announcement is a significant part of multi-agent communications. Providing a formal account of such announcements is important for representing and reasoning about effects of actions and epistemic planning in multi-agent domains. This paper attacks the problem of dealing with lying and misleading announcements by defining update models for them. It also shows that these update models yield intuitive results when applying on a pointed Kripke structure for reasoning about the beliefs of agents.

Keywords: Lying and Misleading Announcement · Mutli-Agent System · Beliefs.

1 Introduction and Motivation

Communication between agents is an integral part of multi-agent systems ever since the research area was introduced. This led to the development of agent communication languages (e.g., KQML) for agents to communicate with each other. In most of these languages, performatives such as `ask` and `tell` allow agents to ask for and receive information from other agents. It general, it is assumed that agents within a system will provide each other, via `tell`, truthful information. This assumption appears to still hold in the era of *Agentic AI* [1, 7, 12, 15, 16], which has been hailed as the new frontier of AI with enormous potential. It is in this context, the study of untruthful announcements becomes ever more important. First, with autonomous agents capable of understanding natural languages and assimilating information from various sources, there is no guarantee that the information that an agent holds true is actually true. Therefore, an agent can, unintentionally, send false information to others. Second, as agentic AI agents can create their own goals, with their own utility functions, there is no guarantee that they would not intentionally lie to others to achieve their individual goals.

In this paper, we use *untruthful announcements* to refer to the act of an agent informing others a piece of information that is not true in the actual state of the world. We divide them into two groups: *misleading announcements* and *lies*. In the first group, the agents who make an announcement of φ, are uncertain about φ, they do not believe φ nor $\neg\varphi$. In the second group, the agents who make an announcement of φ, believe $\neg\varphi$. The goal of this paper is to develop a formalization of untruthful announcements that allows for the consideration of this type of action in epistemic planning. In this context, we would like to specify that an action is of the type announcement and leave the decision when a truthful/untruthful announcement should be executed for the planner. For example, we should be able to specify that "agent A informs agent B that a coin lies heads up" is an announcement; and, if the goal is to have B to have false belief about the state of the coin, then the planner can make sure that A knows that the coin lies tails up before the announcement is executed. In order to achieve this goal, we first present a way to specify misleading announcements and lying actions using a high-level action description, the language m$\mathcal{A}^*$ [3]. Under this view, all types of announcements are *announcements*. Whether or not an announcement is a truthful announcement, a misleading announcement, or a lie will be settled at the execution time, depending on the concrete state of the knowledge and beliefs of the agents. We then present a method for automatically constructing an edge-conditioned update model given a Kripke model and an occurrence of an announcement.

We note that studying the effects of lying and misleading announcements on the knowledge and beliefs of agents has been intensively investigated by logicians and philosophers (see, e.g., [11, 14, 19]) and a formal account of lying in dynamic epistemic logic has been the focus of several works such as [2, 8, 10, 9]. Most of these works treat lying announcements as public announcements, i.e., every agent is a full observer. Furthermore, with the exception of [9], it is assumed that the addressee believes the announced formula. In our view, this assumption is too strong as lying announcement is, first and foremost, a type of announcement, and thus, when it is made, there will be agents who are unaware of the lie, or who only partially aware about it. Therefore, treating lying announcements as public announcements leaves out several scenarios that should be considered. Another significant issue with earlier approaches is that they do not enable a simple integration of untruthful announcements into epistemic planning systems because they do not provide a method for constructing the event models used in their formalization. Our proposal is therefore significant different than earlier works. Furthermore, it provides a systematic way for the consideration of untruthful announcements in epistemic planning.

The paper is organized as follows. In Section 2 we briefly review the necessary notions. Afterwards, we introduce a running example and state our assumptions for our formalization. Sections 4 introduce the construction of the update model for untruthful announcements and prove several properties of the model. We then discuss possible extensions, relate them to other approaches, and provide final considerations.

2 Preliminary

Belief Formulae. A *multi-agent* domain $\langle \mathcal{AG}, \mathcal{F} \rangle$ includes a finite and non-empty set of agents $\mathcal{AG}$ and a set of fluents (atomic propositions) $\mathcal{F}$, used to encode the properties of the world. *Belief formulae* over $\langle \mathcal{AG}, \mathcal{F} \rangle$ are defined by the BNF: "$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid (\varphi \vee \varphi) \mid \mathbf{B}_i\varphi$" where $p \in \mathcal{F}$ is a fluent and $i \in \mathcal{AG}$. We refer to a belief formula which does not contain any occurrence of $\mathbf{B}_i$ as *a fluent formula*. In addition, for a formula φ and a non-empty set $\alpha \subseteq \mathcal{AG}$, $\mathbf{B}_\alpha\varphi$ and $\mathbf{C}_\alpha\varphi$ denote $\bigwedge_{i\in\alpha} \mathbf{B}_i\varphi$ and $\bigwedge_{k=1}^{\infty} \mathbf{B}_\alpha^k\varphi$, where $\mathbf{B}_\alpha^1\varphi$=$\mathbf{B}_\alpha\varphi$ and $\mathbf{B}_\alpha^k\varphi$=$\mathbf{B}_\alpha^{k-1}\mathbf{B}_\alpha\varphi$ for $k > 1$, respectively. $\mathcal{L}_{\mathcal{AG}}$ denotes the set of belief formulae over $\langle \mathcal{AG}, \mathcal{F}\rangle$.

Satisfaction of belief formulae is defined over *pointed Kripke structures* [13]. A Kripke structure M is a tuple $\langle S, \pi, \{\mathcal{B}_i\}_{i\in\mathcal{AG}}\rangle$, where S is a set of worlds (denoted by $M[S]$), $\pi : S \mapsto 2^{\mathcal{F}}$ is a function that associates an interpretation of $\mathcal{F}$ to each element of S (denoted by $M[\pi]$), and for $i \in \mathcal{AG}$, $\mathcal{B}_i \subseteq S \times S$ is a binary relation over S (denoted by $M[i]$). For convenience, we will often draw a Kripke structure M as a directed labeled graph, whose set of labeled nodes represents S and whose set of labeled edges contains $s \xrightarrow{i} t$ iff $(s, t) \in \mathcal{B}_i$; the label of each node is the name of the world and its interpretation is displayed as a text box next to it. For $u \in S$ and a fluent formula φ, $M[\pi](u)$ and $M[\pi](u)(\varphi)$ denote the interpretation associated to u via π and the truth value of φ with respect to $M[\pi](u)$. For a world $s \in M[S]$, (M, s) is a *pointed Kripke structure*, hereafter called a *state*. The satisfaction relation $\models$ between belief formulae and a state (M, s) is defined as follows:

1. $(M, s) \models p$ if p is a fluent and $M[\pi](s)(p)$ is true;
2. $(M, s) \models \neg\varphi$ if $(M, s) \not\models \varphi$;
3. $(M, s) \models \varphi_1 \wedge \varphi_2$ if $(M, s) \models \varphi_1$ and $(M, s) \models \varphi_2$;
4. $(M, s) \models \varphi_1 \vee \varphi_2$ if $(M, s) \models \varphi_1$ or $(M, s) \models \varphi_2$;
5. $(M, s) \models \mathbf{B}_i\varphi$ if $\forall t.[(s, t) \in \mathcal{B}_i \Rightarrow (M, t) \models \varphi]$.

We also use *update models* to describe transformations of (pointed) Kripke structures according to a predetermined transformation pattern. An update model is structured similarly to a pointed Kripke structure and it describes how to transform a pointed Kripke structure using an update operator defined in [2, 4].

Let us start with some preliminary definitions. An $\mathcal{L}_{\mathcal{AG}}$-substitution is a set $\{p_1 \to \varphi_1, \ldots, p_k \to \varphi_k\}$, where each p_i is a distinct fluent in $\mathcal{F}$ and each $\varphi_i \in \mathcal{L}_{\mathcal{AG}}$. $SUB_{\mathcal{L}_{\mathcal{AG}}}$ denotes the set of all $\mathcal{L}_{\mathcal{AG}}$-substitutions.

Definition 1 (Update Model). *Given a set $\mathcal{AG}$ of n agents, an* update model $\mathbf{\Sigma}$ *is a tuple $\langle \Sigma, R_1, \ldots, R_n, pre, sub\rangle$ where*

(i) Σ is a set, whose elements are called events*;*
(ii) each R_i is a binary relation on Σ;
(iii) $pre : \Sigma \to \mathcal{L}_{\mathcal{AG}}$ is a function mapping each event $e \in \Sigma$ to a formula in $\mathcal{L}_{\mathcal{AG}}$; and
(iv) $sub : \Sigma \to SUB_{\mathcal{L}_{\mathcal{AG}}}$ is a function mapping each event $e \in \Sigma$ to a substitution in $SUB_{\mathcal{L}_{\mathcal{AG}}}$.

An update instance ω *is a pair $(\mathbf{\Sigma}, e)$ where $\mathbf{\Sigma}$ is an update model $\langle \Sigma, R_1, \ldots, R_n, pre, sub\rangle$ and e, referred to as a* designated event*, is a member of Σ.*

Intuitively, an update model represents different views of an action occurrence which are associated with the observability of agents. Each view is represented by an event in Σ. The designated event is the one that agents who are aware of the action occurrence will observe. The relation R_i describes agent i's uncertainty on action execution—i.e., if $(\sigma, \tau) \in R_i$ and event σ is performed, then agent i may believe that event τ is executed instead. *pre* defines the action precondition and *sub* specifies the changes of fluent values after the execution of an action.

Definition 2 (Updates by an Update Model). *Let M be a Kripke structure and $\mathbf{\Sigma} = \langle \Sigma, R_1, \ldots, R_n, pre, sub \rangle$ be an update model. The* update *induced by $\mathbf{\Sigma}$ defines a Kripke structure $M' = M \otimes \mathbf{\Sigma}$, where:*

(i) $M'[S] = \{(s, \tau) \mid s \in M[S], \tau \in \Sigma, (M, s) \models pre(\tau)\}$;
(ii) $((s, \tau), (s', \tau')) \in M'[i]$ *iff* $(s, \tau), (s', \tau') \in M'[S]$, $(s, s') \in M[i]$ *and* $(\tau, \tau') \in R_i$;
(iii) For all $(s, \tau) \in M'[S]$ *and* $f \in \mathcal{F}$, $M'[\pi]((s, \tau)) \models f$ *iff* $f \rightarrow \varphi \in sub(\tau)$ *and* $(M, s) \models \varphi$.

The structure M' is obtained from the component-wise cross-product of the old structure M and the update model $\mathbf{\Sigma}$, by **(i)** removing pairs (s, τ) such that (M, s) does not satisfy the action precondition (checking for satisfaction of action's precondition), and **(ii)** removing links of the form $((s, \tau), (s', \tau'))$ from the cross product of $M[i]$ and R_i if $(s, s') \notin M[i]$ or $(\tau, \tau') \notin R_i$ (ensuring that each agent's accessibility relation is updated according to the update model). We note that the truth value of fluents in each world (s, τ) of M' is dictated by the substitution *sub*: if $f \rightarrow \varphi \in sub(\tau)$ and $(M, s) \models \varphi$ then f is true in (s, τ).

An *update template* is a pair $(\mathbf{\Sigma}, \Gamma)$, where $\mathbf{\Sigma}$ is an update model with the set of events Σ and $\Gamma \subseteq \Sigma$. The update of a state (M, s) given an update template $(\mathbf{\Sigma}, \Gamma)$ is a set of states, denoted by $(M, s) \otimes (\mathbf{\Sigma}, \Gamma)$, where

$$(M, s) \otimes (\mathbf{\Sigma}, \Gamma) = \{(M \otimes \mathbf{\Sigma}, (s, \tau)) \mid \tau \in \Gamma, (M, s) \models pre(\tau)\}$$

3 Specifying Announcement Actions

In this paper, we consider an announcement action (or an *announcement*, for short) as an action of an agent (or a group of agents) telling other agents a piece of information. In the context of planning, an action can only be executed under certain conditions. An announcement is of the form $a\langle\alpha\rangle$ where a represents an announcement and α is a group of agents who execute a. Following [3], we specify the announcements by statements of the form

$$\mathsf{a}\ \textbf{announces}\ \varphi \tag{1}$$

where a is an announcement and φ is a formula encoding the announced formula of a. Different from the original paper [3], for brevity of the presentation, we will assume that announcement can be executed in any state. In a multiagent environment, when an event happens, agents have different awareness of its occurrence. We specify the state of awareness, or the observability, of agents about action occurrences by statements of the form

$$z \ \ \textbf{observes} \ \mathsf{a} \ \textbf{if} \ \delta_z \tag{2}$$

$$z \ \ \textbf{aware\_of} \ \mathsf{a} \ \textbf{if} \ \theta_z \tag{3}$$

where z denotes an agent, δ_z and θ_z are formulae. Statement (2) indicates that agent z is a *full observer* of a if δ_z holds. Statement (3) states that agent z is a *partial observer* of a if θ_z holds. We will assume that for each pair of an agent z and an action a, if z occurs in both (2) and (3) then δ_z and θ_z are mutual exclusive, i.e., $\delta_z \rightarrow \neg\theta_z$ and $\theta_z \rightarrow \neg\delta_z$. Without the lost of generality, we assume that whenever z does not occur in a statement of the form (2) (resp. (3)) then z **observes** a **if** $\bot$ (resp. z **aware_of** a **if** $\bot$) is given.

The language m$\mathcal{A}^*$ assumes that whenever an announcement $\mathsf{a} = a\langle\alpha\rangle$ in the statement (1) is executed in a state (M, s), $(M, s) \models \mathbf{B}_\alpha\varphi$ holds. This means that the agents announce only what they believe in. As we have mentioned earlier, our focus in this paper is to study announcement actions that have not been considered in m$\mathcal{A}^*$. Specifically, our aim is to define update models for untruthful announcements.

For our discussion in this paper, we will use a variant of the *Concealed Coin* from [2] as a running example and denote this domain with D_{coin}. In this domain, we have $\mathcal{AG} = \{A, B, C, D, E\}$. The agents are in a large room and a coin lies on the table in the middle of the room. The set of fluents $\mathcal{F}$ for this domain consist of

- *head*: the coin lies heads up;
- $nextto(x, y)$: indicates that agent x and agent y are next to each other and can hear each other; and
- $close(x, y)$: encodes that x and y are close to each other and they can see each other.

We assume that $nextto(x, y)$ implies $close(x, y)$, i.e., whenever the two agents are next to each other, they are close to each other as well. We consider a single action template $head(v)\langle x\rangle$, which encodes that agent x announces "the truth value for heads being up is v" where v is either *True* or *False*, and is specified[1] is as follows:

$$head(v)\langle x\rangle \ \textbf{announces} \ head = v$$

The observability of agents is described by

$$y \ \ \textbf{observes} \ head(v)\langle x\rangle \ \textbf{if} \ nextto(x, y)$$

$$y \ \ \textbf{aware\_of} \ head(v)\langle x\rangle \ \textbf{if} \ close(x, y) \wedge \neg nextto(x, y)$$

where $x, y \in \mathcal{AG}$ and $x \neq y$. The proximity between agents can be changed by agents' actions (e.g., moving). We will state the agents' proximity whenever

[1] We take the liberty of adjusting the syntax of m$\mathcal{A}^*$ slightly and assume that $nextto(x, x)$ is always true.

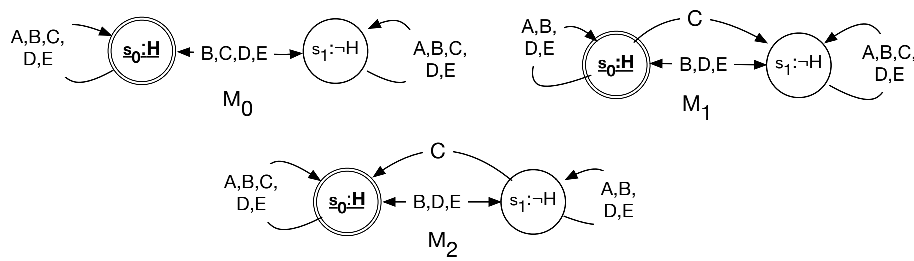

Fig. 1. Three states (M_0, s_0), (M_1, s_0), and (M_2, s_0)—H denotes *head*. A circle represents a world. All states have the same set of worlds $\{s_0, s_1\}$. A double circle represents the real state of the world (s_0). Labeled links indicate the accessibility relations of agents. Truth value of *head* in each world is given by H (true) or $\neg H$ (false).

this information is needed. We will also assume that the actions can always be executed.

Consider the states (M_0, s_0), (M_1, s_0), and (M_2, s_0) shown in Figure 1. These states represent different scenarios of beliefs of the agents. In all of them, it is common knowledge that A knows whether *head* or $\neg$*head* holds. In (M_1, s_0), C believes that $\neg$*head* holds. C believes that *head* holds in (M_2, s_0). None of B, D, or E knows whether *head* or $\neg$*head*. This knowledge is common among the agents. The ity between agents is the same in both worlds s_0 and s_1.

Consider an occurrence of the announcement $head(False)\langle A\rangle$ in (M_0, s_0), i.e., A announces that $\neg$*head*. In this case, A makes a false announcement since $(M_0, s_0) \models \mathbf{B}_A head$. On the other hand, if $head(False)\langle B\rangle$ occurs in (M_0, s_0), all we can say that B makes a misleading announcement since $(M_0, s_0) \models \neg\mathbf{B}_B head \wedge \neg\mathbf{B}_B\neg head$. This discussion shows that characterizing an announcement as truthful, lying, or misleading needs to take into consideration the state in which the announcement occurs. Formally, given a state (M, s) and an announcement $a\langle\alpha\rangle$ in (1), we say that $a\langle\alpha\rangle$ is

– *truthful* if $(M, s) \models \mathbf{B}_\alpha\varphi$;
– *misleading* if $(M, s) \models \neg\mathbf{B}_\alpha\varphi \vee \neg\mathbf{B}_\alpha\neg\varphi$; and
– *lying* if $(M, s) \models \mathbf{B}_\alpha\neg\varphi$.

From now on, whenever we refer to an untruthful announcement, we mean that the announcement is either misleading or a lying announcement.

Assumptions about Changes in Agents' Beliefs. Observe that when an announcement occurs, some agents are fully observant, partially observant, or oblivious. For instance, if A is next to all agents, then all agents will be fully observant. If only B is next to A and others are not close to A, then only B will hear A and others are oblivious; etc. Since an announcement does not change the world, it is up to the agents who are fully aware of the action occurrence to believe in the announced formula; we make the following assumptions on how beliefs of agents will change when an untruthful announcement occurs.

(**A1**) 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**) 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.

The assumption (**A1**) indicates that agents are "opinionated" about their own beliefs, while (**A2**) suggests that agents are eager to remove uncertainty in their beliefs. There can be finer or different classifications of agents' attitudes with respect to an announcement, e.g., agents' attitudes can depend on who makes the announcement and what type of information is announced. For example, in [9], agents can be credulous, skeptical, or revising reasoners, and they will update their beliefs based on their attitude. This is an interesting subject but outside the scope of this paper. We next formalize the update models for untruthful announcements.

Desirable Properties. Observe that we do not require that φ is true in the actual state. As such, it is possible that an untruthful announcement might result in some agents believe in a formula that is true in the actual world (e.g., a lie indirectly becomes a true announcement for some agents). From our assumptions (**A1**) and (**A2**), if i is an agent in $\mathcal{AG}$, we have the following desirable properties:

(**P1**) i is a *full observer* of a (i.e. $(M, s) \models \delta_i$). In this case, given the assumptions (**A1**) and (**A2**), we have the following sub-cases:

- $(M, s) \models (\mathbf{B}_i \mathbf{B}_\alpha \neg\varphi) \vee (\mathbf{B}_i(\neg\mathbf{B}_\alpha\neg\varphi \wedge \neg\mathbf{B}_\alpha\varphi)) \vee \mathbf{B}_i\neg\varphi \vee \mathbf{B}_i\varphi$: by (**A1**), the belief of i should not be changed since i knows α are untruthful or i believes whether φ
- $(M, s) \models \neg(\mathbf{B}_i\varphi \vee \mathbf{B}_i\neg\varphi) \wedge \neg\mathbf{B}_i\mathbf{B}_\alpha\neg\varphi \wedge \neg(\mathbf{B}_i(\neg\mathbf{B}_\alpha\neg\varphi \wedge \neg\mathbf{B}_\alpha\varphi))$: by (**A2**), the belief of i about φ should be changed and $\mathbf{B}_i\varphi$ is true after the untruthful announcement.

(**P2**) i is a *partial observer* of the action a (i.e. $(M, s) \models \theta_i$). As i is unaware of what is announced about the formula, her belief about φ does not change. However, she would assume that people who are fully observant know the truth value of the formula after the action occurrence.
(**P3**) i is not aware of the execution of the action a (i.e. $(M, s) \models \neg\delta_i \wedge \neg\theta_i$): nothing changes for i.

4 Formalizing Untruthful Announcements

Let us consider an occurrence of a untruthful announcement about φ, $\mathsf{a} = a\langle\alpha\rangle$, in a state (M, s). In other words, we have that a **announced** φ is a statement specified the action a and a occurs in (M, s). By our definition, $(M, s) \models \mathbf{B}_\alpha\neg\varphi$ or $(M, s) \models \neg(\mathbf{B}_\alpha\neg\varphi \vee \mathbf{B}_\alpha\varphi)$. We follow the approach in [3] and define the following sets of agents:

- The set of agents who are full observers and know whether φ. We denote this set of agents by F_{uk}[2] and define it by

$$F_{uk} = \{i \mid i \in \mathcal{AG}, (M, s) \models \delta_i \wedge (\mathbf{B}_i\varphi \vee \mathbf{B}_i\neg\varphi)\} \cup \{i \mid i \in \alpha \mid (M, s) \models (\mathbf{B}_i\varphi \vee \mathbf{B}_i\neg\varphi)\}$$

  This group of agents will not change their beliefs about φ regardless of whether they know that **a** is an untruthful announcement. This is by the assumption (**A1**).
- The set of agents who are full observers but do not know whether φ. There are two cases here:
  - The agent who know that α is making an untruthful announcement. We denote this set of agents by F_{ud}[3] and define it by

$$F_{ud} = \{i \mid i \in \mathcal{AG}, (M, s) \models \delta_i \wedge (\Gamma_i(M, s, \alpha) \wedge \neg(\mathbf{B}_i\varphi \vee \mathbf{B}_i\neg\varphi))\} \cup \{i \mid i \in \alpha, (M, s) \models \neg(\mathbf{B}_i\varphi \vee \mathbf{B}_i\neg\varphi))$$

    where
$$\Gamma_i(M, s, \alpha) \stackrel{def}{=} \mathbf{B}_i\mathbf{B}_\alpha\neg\varphi \vee (\mathbf{B}_i(\neg\mathbf{B}_\alpha\neg\varphi \wedge \neg\mathbf{B}_\alpha\varphi)).$$

    Intuitively, $\Gamma_i(M, s, \alpha)$ means that i knows that a is an untruthful announcement. Again, by the assumption (**A1**), these groups of agents will not change their beliefs about φ.
  - The set of agents who are full observers, who do not know that α is making an untruthful announcement. We denote this set of agents by F_c and define it by

$$F_c = \{i \mid i \in \mathcal{AG}, (M, s) \models \delta_i \wedge \neg\Gamma_i(M, s, \alpha) \wedge \neg\mathbf{B}_i\varphi \wedge \neg\mathbf{B}_i\neg\varphi\} \setminus \alpha.$$

    This group of agents, by (**A2**), will change to believe that φ is true.
- The set of agents who are partial observers. We denote this set of agents by P and define it by $P = \{i \mid i \in \mathcal{AG}, (M, s) \models \theta_i\}$.
- The set of agents who are oblivious and denoted by O is defined by $O = \{i \mid i \in \mathcal{AG}, (M, s) \models \neg\delta_i \wedge \neg\theta_i\} \setminus \alpha$.

Observe that we insist that α belongs to the set of full observers who do not change their beliefs about φ. This assumption can be eliminated by removing the part relating to α of the five sets. In the later part of this section, we will show an example that this assumption is reasonable. By our assumptions about δ_i and θ_i, it is easy to see that F_{ud}, F_{uk}, F_c, P, and O are pairwise disjoint. From the formalization of announcement in [3] and the desirable properties (**P1**)-(**P3**) we have that the execution of an untruthful announcement of φ, a, creates the following events:

- an event σ that is the designated event, which is the true event that occurs;

[2] The subscript $_{uk}$ stands for "unchanged in belief" and "know whether φ".

[3] The subscript $_{ud}$ stands for "unchanged in belief" and "don't know whether φ".

– a full observer, who knows that φ holds, will consider that an event μ which has the precondition φ occurs;
– a full observer, who knows that $\neg\varphi$ holds, will consider that an event ξ which has the precondition $\neg\varphi$ occurs;
– a full observer, who does not know whether φ or $\neg\varphi$ holds. In this case, we have two situations
  - if the observer knows that α is untruthful, then nothing changes, and the observer would consider that either μ or ξ occurs.
  - if full observer knows that α is untruthful, considers that an event τ happens that causes the agent to change its belief in φ to be true. The precondition of this event is the formula φ.
– a partial observer, who is aware of the announcement, but do not know the details, will believe the full observers know whether φ but cannot know their truth values. Following [3], we will create two events ρ and λ whose preconditions are φ and $\neg\varphi$, respectively; and
– an oblivious agent, who is not aware of the announcement, will consider ϵ, an event denoting that nothing has happened.

Given the above events, what are the accessibility relations of the agents? Let us consider an agent i. We have the following scenarios:

– σ is the true event that occurs. What will be the event that i believes to have occurred? This depends on the agent's classification.
  - If $i \in F_{uk}$, then i could think that either μ or ξ could have occurred. Essentially, in the view of i, its beliefs do not change.
  - Similarly, if $i \in F_{ud}$, then i could think that either μ or ξ could have occurred.
  - If $i \in F_c$, then i thinks that τ occurred.
  - If $i \in P$, then i thinks that either ρ or λ occurred.
  - If $i \in O$, then i thinks that ϵ occurred.
– Similarly, if μ is the true event that occurs, then we can easily say that for $i \in P$, i thinks that either ρ or λ occurred; and for $i \in O$, i thinks that ϵ occurred. On the other hand, for $i \in F_{uk}$, then i only thinks that μ occurred; for $i \in F_{ud}$, then i thinks that μ or ξ occurred.

Similarly, we can derive the accessibility relations from other events for each agent $i \in \mathcal{AG}$. This derivation is summarized in the following definition.

Definition 3 (Update Model for Untruthful Announcement). *Let* $\mathsf{a} = a\langle\alpha\rangle$ *be an untruthful announcement of formula* φ *and* (M, s) *be a state such that* $(M, s) \models \mathbf{B}_\alpha\neg\varphi$ *or* $(M, s) \models \neg(\mathbf{B}_\alpha\neg\varphi \vee \mathbf{B}_\alpha\varphi)$*. The update model for* a *in* (M, s)*,* $\omega(\mathsf{a}, (M, s))$*, is defined by* $\langle \Sigma, \{R_i\}_{i\in\mathcal{AG}}, pre, sub\rangle$ *as follows:*

○ $\Sigma = \{\sigma, \mu, \xi, \tau, \rho, \lambda, \epsilon\}$;
○ R_i *is defined as follows:*
  - *for* $i \in F_{uk}$, $R_i = \{(\sigma, \mu), (\sigma, \xi), (\xi, \xi), (\mu, \mu), (\rho, \rho), (\lambda, \lambda), (\epsilon, \epsilon)\}$.

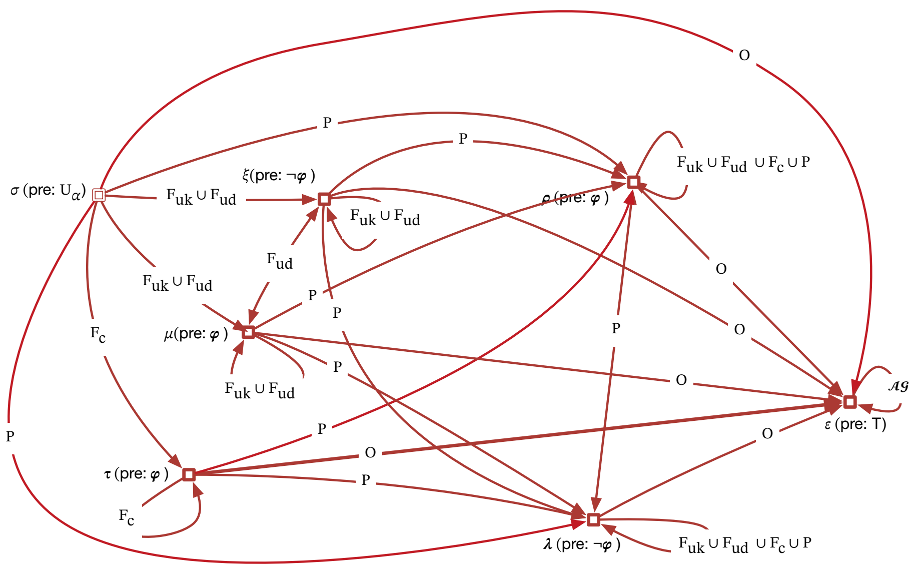

Fig. 2. Update model for an untruthful announcement of φ - U_α (precondition of σ) denotes $\mathbf{B}_\alpha \neg\varphi \vee \neg(\mathbf{B}_\alpha\varphi \vee \mathbf{B}_\alpha\neg\varphi)$. Nodes are the events of the update model and links represent the set R_i for $i \in \mathcal{AG}$. A node in double rectangles is the designated event.

- *for* $i \in F_{ud}$, $R_i = \{(\sigma,\mu),(\sigma,\xi),(\xi,\xi),(\mu,\mu),(\xi,\mu),(\mu,\xi),(\rho,\rho),(\lambda,\lambda),(\epsilon,\epsilon)\}$.
- *for* $i \in F_c$, $R_i = \{(\sigma,\tau),(\tau,\tau),(\rho,\rho),(\lambda,\lambda),(\epsilon,\epsilon)\}$.
- *for* $i \in UF$, $R_i = \{(\sigma,\tau),(\rho,\rho),(\lambda,\lambda),(\epsilon,\epsilon)\}$.
- *for* $i \in P$, $R_i = \{(\sigma,\rho),(\tau,\rho),(\mu,\rho),(\xi,\rho),(\sigma,\lambda),(\tau,\lambda),(\mu,\lambda),(\xi,\lambda),(\mu,\lambda),(\lambda,\mu),(\rho,\rho),(\lambda,\lambda),(\epsilon,\epsilon)\}$.
- *for* $i \in O$, $R_i = \{(\eta,\epsilon) \mid \eta \in \Sigma\}$

◦ *The preconditions pre are:*

- $pre(\sigma) = \mathbf{B}_\alpha\neg\varphi \vee (\neg(\mathbf{B}_\alpha\varphi \vee \mathbf{B}_\alpha\neg\varphi))$;
- $pre(\tau) = \varphi$;
- $pre(\rho) = \varphi$;
- $pre(\lambda) = \neg\varphi$;
- $pre(\chi) = \neg\varphi$;
- $pre(\mu) = \varphi$;
- $pre(\epsilon) = \top$.

◦ $sub(x) = \emptyset$ *for each* $x \in \Sigma$.

The designated event of $\omega(\mathsf{a},(M,s))$ *is* σ.

Having defined the update model $\omega(\mathsf{a},(M,s))$ for an untruthful announcement a in (M,s), we define the state resulting from the occurrence of a in (M,s) as $(M,s) \otimes (\omega(\mathsf{a},(M,s)),\sigma)$.

We will illustrate the execution of $head(False)\langle A\rangle$ in (M_0,s_0) in Figure 1. It is easy to see that this is a lying announcement. For illustration purpose, let assume that A, B and C are next to each other, while D is close to them (but is not next to them) and E is not close to everyone.

Example 1. Let us consider the occurrence of $head(False)\langle A\rangle$ in (M_0,s_0) (Fig. 1) where A, B and C are *full observers*, D is *partial observer* and E is *oblivious*

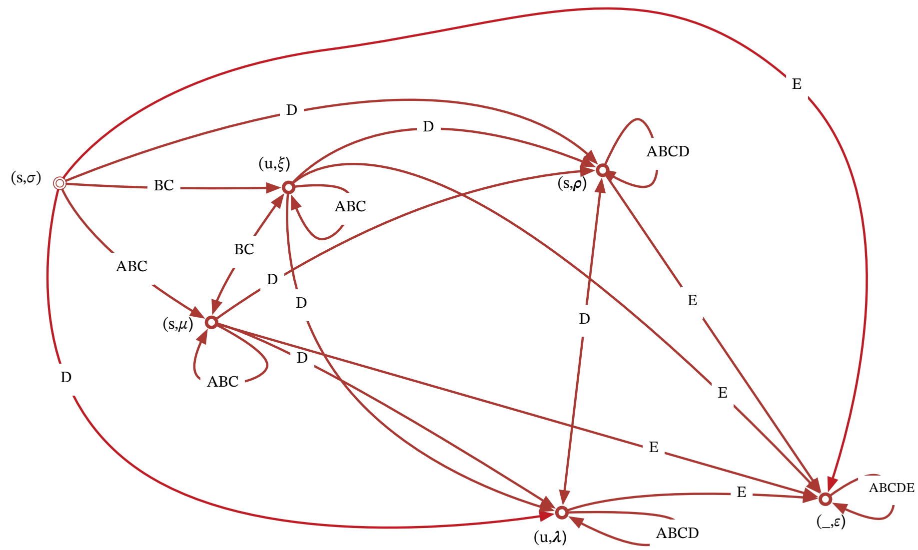

Fig. 3. $(M_0, s_0) \otimes (\omega(head(False)\langle A \rangle, (M_0, s_0)), \sigma)$ – $(\_, \epsilon)$ denotes (s, ϵ) or (u, ϵ)

$(nextto(A, B), nextto(A, C), nextto(B, C), close(A, D), close(B, D), close(C, D)$ are true in both s_0 and s_1). The update model $\omega(head(False)\langle A \rangle, (M_0, s_0))$ can be obtained from Fig. 3 with $\alpha = \{A\}$ and $\varphi = \neg head$. For this example, we have that $F_{uk} = \{A\}$, $F_{ud} = \{B, C\}$, $F_c = \emptyset$, $P = \{D\}$, and $O = \{E\}$.

The result (M', s') (where $s' = (s, \sigma)$) of this announcement is given in Fig. 3. As we can observe from the result, the agents have the following knowledge:

- A's belief does not change. A knows that B and C will not change their belief about *head*.
- B's belief does not change but so its belief about A's belief about *head*.
- D is uncertain about *head* or $\neg head$ but believes that A, B, and C share the same belief about *head* or $\neg head$ ($(M', s') \models \neg(\mathbf{B}_D head \vee \mathbf{B}_D \neg head) \wedge \mathbf{B}_D(\mathbf{B}_{\{A,B,C\}} \neg head \vee \mathbf{B}_{\{A,B,C\}} head)$).
- E's belief does not change.

We will next discuss the properties of the update model $\omega(\mathsf{a}, (M, s))$. In the rest of this section, we assume that for agents x, y, [x **observes** a **if** δ_x] and [x **aware_of** a **if** θ_x], [y **observes** a **if** δ_y], and [y **aware_of** a **if** θ_y] are given. Let $(M's') = (M, s) \otimes \omega(\mathsf{a}, (M, s))$.

We note that the update models for untruthful announcement guarantee that agents in $\mathcal{AG}$ do not become ignorant after the occurrence of the announcement if they are not ignorant before the occurrence of the announcement, i.e., in (M, s). This is proved in the following proposition.

Proposition 1. *For every agent $x \in \mathcal{AG}$, there exists some state $u' \in M'[S]$ such that $(s', u') \in M'[x]$.*

Proof. Because x is not ignorant in (M, s) there exists some u such that $(s, u) \in M[x]$.

Clearly $x \in O$, either $((s, \sigma), (u, \epsilon))$ is in $M'[x]$.
For $x \in P$, either $((s, \sigma), (u, \rho))$ or $((s, \sigma), (u, \lambda))$ is in $M'[x]$.
For $x \in F_{uk} \cup F_{ud}$, either $((s, \sigma), (u, \xi))$ or $((s, \sigma), (u, \mu))$ is in $M'[x]$.
For $x \in F_c$, $((s, \sigma), (u, \tau))$ is in $M'[x]$.
The different cases prove the proposition. □

We prove that the proposed formalization for untruthful announcement satisfying the properties (**P1**)-(**P3**) in the next propositions. Let $\mathsf{a} = a\langle\alpha\rangle$ be the occurrence of an untruthful announcement of φ and $\omega(\mathsf{a}, (M, s))$ be given in Definition 3. Assume that $(M', s') = (M, s) \otimes (\omega(\mathsf{a}, (M, s)), \sigma)$. Consider an arbitrary agent x such that [x **observes** a **if** δ_x] and [x **aware_of** a **if** θ_x] are given. We can prove the following:

Proposition 2. *If* $(M, s) \models \delta_x$[4] *or* $x \in \alpha$, *then*

- *if* $(M, s) \models (\mathbf{B}_x\varphi \vee \mathbf{B}_x\neg\varphi)$ *then* $(M', s') \models (\mathbf{B}_x\varphi \vee \mathbf{B}_x\neg\varphi)$.
- *if* $(M, s) \models \Gamma_x(M, s, \alpha) \wedge \neg(\mathbf{B}_x\varphi \vee \mathbf{B}_x\neg\varphi)$ *then* $(M', s') \models \neg(\mathbf{B}_x\varphi \vee \mathbf{B}_x\neg\varphi)$.
- *if* $(M, s) \models \neg\Gamma_x(M, s, \alpha) \wedge \neg(\mathbf{B}_x\varphi \vee \mathbf{B}_x\neg\varphi)$ *then* $(M', s') \models \mathbf{B}_x\varphi$.

Proof. By definition, we have that $s' = (s, \sigma)$.

- The first item has two subcases:
  - Consider the case $(M, s) \models \mathbf{B}_x\varphi$. This means that for every $(s, u) \in M[x]$, $(M, u) \models \varphi$.
    Since $(M, s) \models \delta_x$, it holds that $x \in F_{ud}$.
    Consider $u' \in M'[S]$ such that $(s', u') \in M'[x]$, we have that $u' = (u, \mu)$ for some $u \in M[S]$ and $(s, u) \in M[x]$. This implies $(M', u') \models \varphi$. Therefore, $(M', s') \models \mathbf{B}_x\varphi$;
  - Similar to the above case, we can show that if $(M, s) \models \mathbf{B}_x\neg\varphi$ then $(M', s') \models \mathbf{B}_x\neg\varphi$. The difference lies in that $u' = (u, \xi)$.
- Assume that $(M, s) \models \Gamma_x(M, s, \alpha) \wedge \neg(\mathbf{B}_x\varphi \vee \mathbf{B}_x\neg\varphi)$. In this case, $x \in F_{ud}$. $(M, s) \models \neg(\mathbf{B}_x\varphi \vee \mathbf{B}_x\neg\varphi)$ then there exists at least 2 worlds u_1, u_2 such that $(s, u_1) \in M[x]$ and $(s, u_2) \in M[x]$, and $(M, u_1) \models \varphi$ and $(M, u_2) \models \neg\varphi$. This means that for $u'_1 = (u_1, \mu)$ and $u'_2 = (u_2, \xi)$, $(s', u'_1) \in M'[x]$ and $(s', u'_2) \in M'[x]$. This implies that $(M', u'_1) \models \varphi$ and $(M', u'_2) \models \neg\varphi$. Therefore, $(M', s') \models \neg(\mathbf{B}_x\varphi \vee \mathbf{B}_x\neg\varphi)$.
- Assume that $(M, s) \models \neg\Gamma_x(M, s, \alpha) \wedge \neg\mathbf{B}_x\varphi \wedge \neg\mathbf{B}_x\neg\varphi$. In this case, $x \in F_c$. This implies that for every $u' \in M'[S]$ such that $(s', u') \in M'[x]$, we have that $u' = (u, \tau)$ for some $u \in M[S]$ and $(s, u) \in M[x]$ and $(M, u) \models \varphi$. Note that because $(M, s) \models \neg\mathbf{B}_x\varphi \wedge \neg\mathbf{B}_x\neg\varphi$ there exists at least on $(s, u) \in M[x]$ such that $(M, u) \models \varphi$, and thus, this implies that $(M', s') \models \mathbf{B}_x\varphi$.

□

[4] Recall that by the assumption of mutual exclusive between δ_x and θ_x, we have that $(M, s) \models \neg\theta_x$.

Observe that this proposition deals with full observers. In the first item, x knows the truth value of φ. As such, x's belief about φ does not change. In the second item, x does not change its belief because it knows that a is an untruthful announcement even though it is uncertain about $\neg\varphi$. The last item indicates that x change its belief about φ because it does not realize that the announcement is an untruthful announcement and is uncertain about φ. The next proposition is about partial observers and oblivious agents.

Proposition 3. *If* $(M, s) \models \theta_x$ *and* $x \notin \alpha$*, then*

- *if* $(M, s) \models \mathbf{B}_x\varphi$ *then* $(M', s') \models \mathbf{B}_x\varphi$*;*
- *if* $(M, s) \models \mathbf{B}_x\neg\varphi$ *then* $(M', s') \models \neg\mathbf{B}_x\varphi$*; and*
- *if* $(M, s) \models \neg(\mathbf{B}_x\varphi \vee \mathbf{B}_x\neg\varphi)$ *then* $(M', s') \models \neg(\mathbf{B}_x\varphi \vee \mathbf{B}_x\neg\varphi)$*; and*
- *for a fluent formula* η*, if* $(M, s) \models \mathbf{B}_x\eta$ *then* $(M', s') \models \mathbf{B}_x\eta$*.*

Proof. Again, we have $s' = (s, \sigma)$. Since $(M, s) \models \theta_x$ and $x \notin \alpha$, we have that $x \in P$. The construction of the update model implies that for every $u' \in M'[S]$ such that $(s', u') \in M'[x]$, it holds that $u' = (u, \lambda)$ or $u' = (u, \rho)$ for some $u \in M[S]$ and $(s, u) \in M[x]$. This allows us to conclude that $v \in M[S]$ such that $(s, v) \in M[x]$ iff (v, λ) or (v, ρ) in $M'[x]$. This implies the above four properties □

The above proposition shows a partial observer does not change its beliefs about the true state of the world. Now, consider $y \in \mathcal{AG}$ such that [y **observes** a **if** δ_y] and [y **aware_of** a **if** θ_y] are given. We can show that

Proposition 4. *If* $(M, s) \models \theta_x$*,* $x \notin \alpha$*, and* $(M, s) \models \delta_y$ *or* $y \in \alpha$ *then* $(M', s') \models \mathbf{B}_x(\mathbf{B}_y\varphi \vee \mathbf{B}_y\neg\varphi)$

Proof. Consider $u', v' \in M'[S]$ such that $(s', u') \in M'[x]$, $(u', v') \in M'[y]$. Since $(M, s) \models \theta_x$ and $(M, s) \models \mathbf{B}_x\delta_y$, it holds that $u' = (u, \lambda)$ and $v' = (v, \lambda)$ (or $u' = (u, \rho)$ and $v' = (v, \rho)$) for some $u, v \in M[S]$, $(s, u) \in M[x]$ and $(u, v) \in M[y]$. By the definition of the update model, this implies $(M', s') \models \mathbf{B}_x(\mathbf{B}_y\varphi \vee \mathbf{B}_y\neg\varphi)$. □

The above proposition shows that partial observers know that full observers know whether φ. The next proposition discusses the beliefs of oblivious agents.

Proposition 5. *If* $(M, s) \models \neg(\delta_y \vee \theta_y)$ *and* $y \notin \alpha$ *then for any fluent formula* η*,*

- $(M, s) \models \mathbf{B}_y\eta$ *iff* $(M', s') \models \mathbf{B}_y\eta$*; and*
- *for every* $x \in \mathcal{AG}$ *and ,* $(M, s) \models \mathbf{B}_x\mathbf{B}_y\eta$ *iff* $(M', s') \models \mathbf{B}_x\mathbf{B}_y\eta$*.*

Proof. Because $(M, s) \models \neg(\delta_y \vee \theta_y)$ and $y \notin \alpha$, $y \in O$. The proof of the first item is similar to the proof in Proposition 3. For the second item, observe that for $u', v' \in M'[S]$ such that $(s', u') \in M'[x]$, $(u', v') \in M'[y]$, it holds that $u' = (u, \xi)$ and $v' = (v, \epsilon)$ for $\xi \in \Sigma$, and $u, v \in M[S]$. The construction of the update model implies the conclusion of the proposition. □

We conclude this section with an example of a misleading announcement that illustrates the point to insist that α belongs to the set of full observers.

Example 2. Let us consider the state (M_3, s_0) described in Figure 4.

This state represents a scenario where the coin in the box is *heads* up but nobody in the room knows this fact. B, D, and A knows that none of them knows the state of the coin. However, C and E believe that A and B could know the state of the coin but do not know which one. Consider a case where A, B and C are next to each other while D and E are close to them.

Assume that A executes the action $head(False)\langle A\rangle$ in (M_3, s_0). Because no agent knows the truth value of $head$, $F_{uk} = \emptyset$.

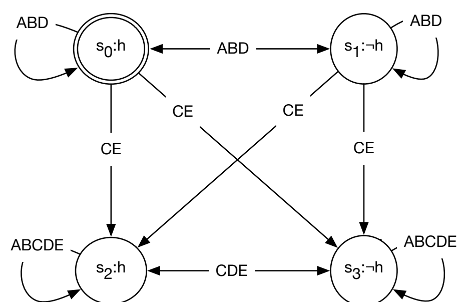

Fig. 4. State (M_3, s_0)

A knows that the announcement is a misleading one and does not know whether $head$. B realizes that the announcement is untruthful but also does not know whether $head$, therefore $F_{ud} = \{A, B\}$.

Furthermore, C believes that A knows whether φ and cannot recognize that A makes an untruthful announcement. Thus, $F_c = \{C\}$.

D and E are close but not next-to A, therefore $P = \{D, E\}$.

The result of the execution of $head(False)\langle A\rangle$ in (M_3, s_0) is shown in Figure 5. There are some green colored worlds (e.g.,) (s_0, τ)that are not reachable from (s_0, σ) and thus the accessibility relations from these worlds will not influence the beliefs of the agents. As such, the links from these worlds are missing for ease of reading of the graph. We can observe the following:

- A, B and D still maintain their uncertainty about the coin since all of them do not know whether $head$ and also realize that A makes an untruthful announcement, a misleading one indeed, since $(M_3, s_0) \models \mathbf{B}_B\neg(\mathbf{B}_A head \vee \mathbf{B}_A\neg head)$.
- C thinks that makes a truthful announcement and believes that the coin lies heads up.
- E and D believe that A, B, and C know whether the coin lies *heads* or *tails* up but cannot distinguish it.

In the above example, A belongs to the set F_{ud} because of it is the executor of the announcement and does not know the truth value of $head$. If the definitions of F_{ud} and F_{uk} do not distinguish between the agents in α, A would have been classified as an agent which needs to change its belief about $head$, which is rather counter intuitive.

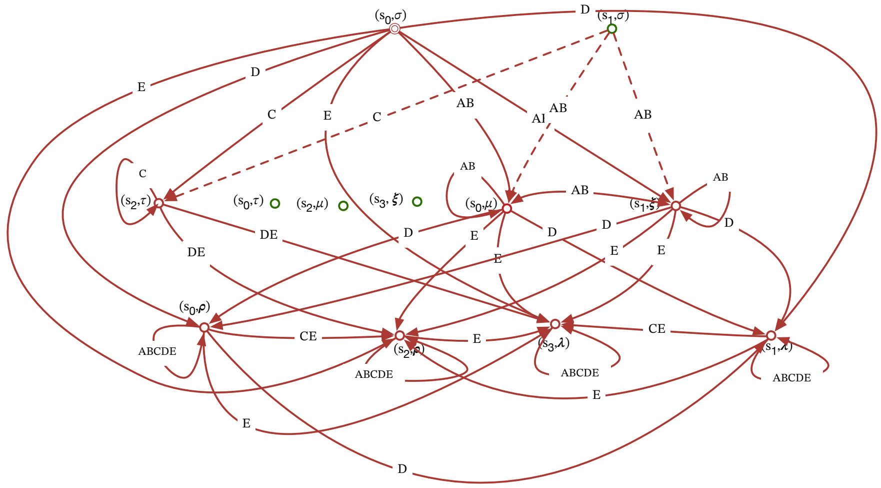

Fig. 5. $(M_3, s_0) \otimes (\omega(head(False)\langle A \rangle, (M_3, s_0)), \sigma)$

5 Discussion and Related Work

As we have mentioned in the introduction, there is a large body of research on lying and misleading announcements. Precise definition of "what is a lie?" is the subject of intensive research by logicians and philosophers (see, e.g., [11, 14, 19, 2, 8, 10, 9]). The key difference between previous works and our proposal in formalizing lying announcement lies in that they treat lying announcements as public announcements, i.e., every agent is a full observer. Furthermore, with the exception of [9], it is assumed that the addressee believes the announced formula. In contrast, we consider lying announcements similarly to a private announcement, with partial observers and oblivious agents. We also allow agents to update their beliefs based on what they believe before the announcement. In our view, considering the different classes of agents is a more realistic approach. Furthermore, the proposed reasoning mechanism can be used by an epistemic multi-agent planner to predict what it needs to do to get an agent A to believe in some lying announcement (e.g., execution of actions that change the belief of A about an announced formula before the announcement).

Comparing to the three types of agents considered in [9], our assumptions (**A1**) and (**A2**) model *skeptical* agents who believe what they are told only if that is consistent with their current beliefs since the present proposal forces a minimal set of agents to change their beliefs. How to do it is not a focus of this paper and we leave this, as well as, the modeling of *credulous* agents and we leave this for the future work.

The present work is closely related to the extension of m$\mathcal{A}^*$ for dealing with untruthful announcements proposed in [17, 18]. The research proposed in these works is in the same direction as they share the same goal, allowing the consideration of untruthful announcements in epistemic planning. Furthermore, the present work adopts the assumptions (**A1**) and (**A2**) and tries to enforce the properties (**P1**)-(**P3**). However, we do not use edge-conditioned update models and the set of events of the update model defined in this paper is also different than the set of events developed in [17, 18].

We note that in order to allow belief formulae to be announced, the only change needed in Definitions 3 is that φ is a belief formula. We omit the discussion on how the update models can be used in extending the transition function Φ of m$\mathcal{A}^*$ to allow the new types of announcements even though it is straightforward. Furthermore, for the purpose of planning, it might be useful to require that the announced formula φ is false in the actual world, i.e., $(M, s) \models \neg\varphi \wedge \mathbf{B}_\alpha \neg\varphi$. This can be dealt with by replacing the precondition of σ with $\neg\varphi \wedge U_\alpha$ (see note on U_α in Figure 2).

It is worth mentioning that our goal is to develop update models for untruthful announcement within the language m$\mathcal{A}^*$, thereby providing the basics for the development of a specification language for epistemic planning systems. It is therefore different from approaches dealing with false beliefs using sensing actions or via belief revisions, such as [3, 5]. As such, we do not distinguish between beliefs and knowledge. This issue can be addressed by the proposal introduced in [6], which also deals with false announcements. Their approach differs from ours in that it directly manipulates the two accessibility relations and does not employ update models. A deeper analysis of the differences and similarities between the two approaches is an interesting task that we leave as a future work.

Finally, let us notice that the time for computing the edge-conditioned update model for an untruthful announcement a in a state (M, s) is linear in the size of $|M| \times |\mathcal{AG}|$ where $|M|$ is the size of M that includes the number of worlds, the number of links in M, and $|\mathcal{AG}|$ is the number of agents.

Proposition 6. *The time complexity of computing $\omega(\mathsf{a}, (M, s))$ is $O(|M| \times |\mathcal{AG}|)$.*

Proof. Computing $M'[S]$ is at most $O(7 * |M|)$ since there are only seven events in $\omega(\mathsf{a}, (M, s))$. Similarly, computing $M'[x]$ will require at most $O(7 * |M|)$ since for each $(s, u) \in M[x]$ and (s, λ) in $M'[S]$, we need to check whether $((s, \lambda), (u, \eta)) \in M'[x]$ for $\eta \in \Sigma$. Hence, computing $\omega(\mathsf{a}, (M, s))$ is $O(|M| \times |\mathcal{AG}|)$. □

6 Conclusion and Future Work

In this paper, we presented a formal account for representing and reasoning about untruthful announcements. We provided a uniformed update model for both lying and misleading announcements and investigated key properties of such update models. The presented work could be viewed as an continuation of the language m$\mathcal{A}^*$ but differs significant from earlier treatments of lying and misleading announcements in m$\mathcal{A}^*$ which employ edge-conditioned update models. We show that the new proposal satisfies desirable properties at the first belief level and does not investigate the behavior of the update model with respect to higher-order beliefs of agents. This will be one of our works in the near future. Another direction that we intend to pursuit is to consider attitudes/trust of agents given an announcement.

References

1. AAAI-Spring-Symposium: Towards Agentic AI for Science: Hypothesis Generation, Comprehension, Quantification, and Validation (2025), https://aaaiagenticai.github.io
2. Baltag, A., Moss, L.: Logics for epistemic programs. Synthese (2004)
3. Baral, C., Gelfond, G., Pontelli, E., Son, T.C.: An action language for multi-agent domains. Artif. Intell. **302**, 103601 (2022). https://doi.org/10.1016/j.artint.2021.103601, https://doi.org/10.1016/j.artint.2021.103601
4. van Benthem, J., van Eijck, J., Kooi, B.P.: Logics of communication and change. Inf. Comput. **204**(11), 1620–1662 (2006)
5. Bolander, T.: Seeing Is Believing: Formalising False-Belief Tasks in Dynamic Epistemic Logic, pp. 207–236. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-319-62864-6_8, https://doi.org/10.1007/978-3-319-62864-6_8
6. Buckingham, D., Kasenberg, D., Scheutz, M.: Simultaneous representation of knowledge and belief for epistemic planning with belief revision pp. 172–181 (9 2020). https://doi.org/10.24963/kr.2020/18, https://doi.org/10.24963/kr.2020/18
7. Buehler, M.J.: Towards Agentic AI for Science Hypothesis Generation, Comprehension, Quantification, and Validation. In: Companion Proceedings of the ACM on Web Conference 2025. p. 1643–1644. WWW '25, Association for Computing Machinery, New York, NY, USA (2025). https://doi.org/10.1145/3701716.3718485, https://doi.org/10.1145/3701716.3718485
8. van Ditmarsch, H., van der Hoek, W., Kooi, B.: Dynamic Epistemic Logic. Springer (2007)
9. van Ditmarsch, H.: Dynamics of lying. Synth. **191**(5), 745–777 (2014). https://doi.org/10.1007/s11229-013-0275-3, https://doi.org/10.1007/s11229-013-0275-3
10. van Ditmarsch, H., van Eijck, J., Sietsma, F., Wang, Y.: On the logic of lying. In: van Eijck, J., Verbrugge, R. (eds.) Games, Actions and Social Software - Multidisciplinary Aspects, Lecture Notes in Computer Science, vol. 7010, pp. 41–72. Springer (2012). https://doi.org/10.1007/978-3-642-29326-9_4
11. van Ditmarsch, H., Hendriks, P., Verbrugge, R.: Editors' review and introduction: Lying in logic, language, and cognition. Top. Cogn. Sci. **12**(2), 466–484 (2020). https://doi.org/10.1111/tops.12492, https://doi.org/10.1111/tops.12492
12. Durante, Z., Huang, Q., Wake, N., Gong, R., Park, J.S., Sarkar, B., Taori, R., Noda, Y., Terzopoulos, D., Choi, Y., Ikeuchi, K., Vo, H., Fei-Fei, L., Gao, J.: Agent AI: Surveying the Horizons of Multimodal Interaction (2024), https://arxiv.org/abs/2401.03568
13. Fagin, R., Halpern, J., Moses, Y., Vardi, M.: Reasoning about Knowledge. MIT press (1995)
14. Mahon, J.E.: The definition of lying and deception. In: Zalta, E. (ed.) Stanford Encyclopedia of Philosophy (2015)
15. Marr, B.: Agentic AI: The Next Big Breakthrough That's Transforming Business And Technology, https://www.forbes.com/sites/bernardmarr/2024/09/06/agentic-ai-the-next-big-breakthrough-thats-transforming-business-and-technology/
16. Murugesan, S.: The Rise of Agentic AI: Implications, Concerns, and the Path Forward. IEEE Intelligent Systems (2025)

17. Pham, L., Son, T.C., Pontelli, E.: Update models for lying and misleading announcements. In: Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing. SAC '22, New York, NY, USA (2022)
18. Pham, L., Son, T.C., Pontelli, E.: Planning in multi-agent domains with untruthful announcements. In: Koenig, S., Stern, R., Vallati, M. (eds.) Proceedings of the Thirty-Third International Conference on Automated Planning and Scheduling, Prague, Czech Republic, July 8-13, 2023. pp. 334–342. AAAI Press (2023). https://doi.org/10.1609/ICAPS.V33I1.27211, https://doi.org/10.1609/icaps.v33i1.27211
19. Sakama, C., Caminada, M., Herzig, A.: A logical account of lying. In: Janhunen, T., Niemelä, I. (eds.) Logics in Artificial Intelligence - 12th European Conference, JELIA 2010, Helsinki, Finland, September 13-15, 2010. Proceedings. Lecture Notes in Computer Science, vol. 6341. Springer (2010)