Multi-Agent Epistemic Explanatory Diagnosis via Reasoning about Actions

Quan Yu[1,2] **Ximing Wen**[1,3] **Yongmei Liu**[1]
[1]Dept. of Computer Science,
Sun Yat-sen University, Guangzhou 510006, China
[2]Qiannan Normal College for Nationalities, Guizhou 558000, China
[3]Guangdong Institute of Public Administration, Guangzhou 510053, China
{yuquan3,wenxim}@mail2.sysu.edu.cn, ymliu@mail.sysu.edu.cn

Abstract

The task of explanatory diagnosis conjectures actions to explain observations. This is a common task in real life and an essential ability of intelligent agents. It becomes more complicated in multi-agent scenarios, since agents' actions may be partially observable to other agents, and observations might involve agents' knowledge about the world or other agents' knowledge or even common knowledge of a group of agents. For example, we might want to explain the observation that p does not hold, but Ann believes p, or the observation that Ann, Bob, and Carl commonly believe p. In this paper, we formalize the multi-agent explanatory diagnosis task in the framework of dynamic epistemic logic, where Kripke models of actions are used to represent agents' partial observability of actions. Since this task is undecidable in general, we identify important decidable fragments via techniques of reducing the potentially infinite search spaces to finite ones of epistemic states or action sequences.

1 Introduction

The task of explanatory diagnosis conjectures what actions occurred to explain observations of dynamic systems. This is a common task in real life and an essential ability of intelligent agents. For example, on a murder scene, a detective speculates on what happened based on his observations of the scene; in a factory, an engineer tries to determine the internal operation of a machine according to the external observations. Explanatory diagnosis becomes more complicated in multi-agent scenarios, since agents' actions may be partially observable to other agents, and observations might involve agents' knowledge about the world or other agents' knowledge or even common knowledge of a group of agents. In this paper, we use the words "knowledge" and "epistemic" in the broad sense.

For example, consider a simple scenario involving three persons Ann, Bob and Carl working in the same office. There is a desk in the office and each person owns a drawer of the desk. Ann has a valuable watch in her drawer, and the three persons commonly know that. The door of the office can only be opened by Ann, Bob and Carl. Ann left the office and closed the door as usual. When she comes back, she finds that, contrary to her belief, her watch is not in her drawer. Now Ann wants to know what happened.

McIlrailth [1998] presented a formal characterization of explanatory diagnosis in the language of the situation calculus, one of the most popular languages for reasoning about actions. Sohrabi *et al.* [2010] established a formal correspondence between explanatory diagnosis and planning, and showed how modern planning techniques can be exploited to generate explanatory diagnoses. Further, they [2011] explored generating preferred explanations via planning.

In the big picture, explanatory diagnosis is a generalization of model-based diagnosis from static systems to dynamic systems. Reiter [1987] laid an elegant theoretical foundation for model-based diagnosis, which is the task of locating faulty components of a system based on a description of the correct behavior of the system and an observation of its aberrant behavior. In control theory, a very influential work about model-based diagnosis of dynamic systems is the one by Sympath *et al.* [1995] where they modeled discrete event systems as finite state automata and characterized diagnosis as a reachability analysis problem. Pencolé and Cordier [2005] proposed a formal framework for the decentralized diagnosis of large-scale discrete event systems. In the past decade, Grastien, Rintanen and colleagues [Rintanen and Grastien, 2007; Grastien *et al.*, 2007] showed how the diagnosing problems of discrete event systems can be translated into the propositional satisfiability problem and solved by modern SAT solvers.

The most influential logic framework for reasoning about actions in the multi-agent case is dynamic epistemic logic (DEL) [van Ditmarsch *et al.*, 2007]. An important concept in DEL is that of an event model, which is a Kripke model of events, representing the agents' uncertainty about the current event. By the product update operation, an event model may be used to update a Kripke model. Recently, Bolander and Anderson [2011] explored multi-agent epistemic planning based on DEL. They showed that single-agent epistemic planning is decidable, but multi-agent epistemic planning is undecidable even without common knowledge. Meanwhile, Löwe *et al.* [2011] showed that for a special type of event models, multi-agent epistemic planning is decidable.

In this paper, we formalize the multi-agent epistemic explanatory diagnosis task in DEL. Since this task is undecidable in general, we identify two important decidable frag-

ments of it by restricting our attention to propositional actions, *i.e.*, actions whose preconditions do not involve knowledge, or purely epistemic actions, *i.e.*, actions which do not effect world change. Our first result is that when observations do not involve common knowledge and all actions are propositional, explanatory diagnosing is decidable. Secondly, we identify a wide variety of special types of propositional purely epistemic actions, and show that when all actions are of these types, explanatory diagnosing is decidable. Our decidability results are achieved via techniques of reducing the potentially infinite search spaces to finite ones of epistemic states or action sequences.

2 Preliminaries

In this section, we present the syntax and semantics of multi-agent epistemic logic, and introduce the concepts of event models and product update from dynamic epistemic logic.

2.1 Multi-agent epistemic logic

We fix a finite set $\mathcal{A}$ of agents and a finite set of atoms $\mathcal{P}$. We use $|S|$ for the cardinality of a set S.

Definition 2.1. The language $\mathcal{L}_{KC}$ of multi-agent epistemic logic with common knowledge is generated by the BNF:

$$\varphi ::= p \mid \neg\phi \mid (\phi \wedge \psi) \mid K_a\phi \mid C_{\mathcal{B}}\phi,$$

where $p \in \mathcal{P}$, $a \in \mathcal{A}$, $\mathcal{B} \subseteq \mathcal{A}$, and $\phi, \psi \in \mathcal{L}_{KC}$. We use $\mathcal{L}_K$ for the language without the $C_{\mathcal{B}}$ operator.

Definition 2.2. A frame is a structure (W, R), where

- W is a finite non-empty set of possible worlds;
- For each agent $a \in \mathcal{A}$, R_a is a binary relation on W, called the accessibility relation for a.

We say R_a is serial if for any $w \in W$, there is $w' \in W$ such that wR_aw'. We say R_a is Euclidean if whenever wR_aw_1 and wR_aw_2, we have $w_1R_aw_2$. A frame whose accessibility relations are equivalence relations is called an S5 frame, and a frame whose accessibility relations are transitive and Euclidean (resp. serial, transitive and Euclidean) is called a K45 (resp. KD45) frame.

Definition 2.3. A Kripke model is a triple $M = (W, R, V)$, where (W, R) is a frame, and V is a valuation map, which maps each $w \in W$ to a subset of $\mathcal{P}$.

Definition 2.4. An epistemic state, or an e-state in short, is a pair $s = (M, w)$, where M is a Kripke model and w is a world of M, called the actual world.

Definition 2.5. Let $s = (M, w)$ be an epistemic state where $M = (W, R, V)$. We interpret formulas in $\mathcal{L}_{KC}$ by induction:

- $M, w \models p$ iff $p \in V(w)$;
- $M, w \models \neg\phi$ iff $M, w \not\models \phi$;
- $M, w \models \phi \wedge \psi$ iff $M, w \models \phi$ and $M, w \models \psi$;
- $M, w \models K_a\phi$ iff for all v s.t. wR_av, $M, v \models \phi$;
- $M, w \models C_{\mathcal{B}}\phi$ iff for all v s.t. $wR_{\mathcal{B}}v$, $M, v \models \phi$, where $R_{\mathcal{B}}$ is the transitive closure of the union of R_a for $a \in \mathcal{B}$.

Example 1. Consider the watch example. We use p_a, p_b and p_c to denote that the watch is in Ann's, Bob's and Carl's drawer, respectively, and p_d to denote that the watch is on the surface of the desk. Then the initial e-state is s_0 where there is only one world w_0, $V(w_0) = \{p_a\}$, and all accessibility relations are reflexive. Then we have $s_0 \models C_{\{a,b,c\}}p_a$, *i.e.*, all agents commonly know that the watch is in Ann's drawer.

2.2 Event models and product update

Event models and product update are two important concepts of dynamic epistemic logic. Intuitively, an event model is a Kripke model of events, representing the agents' uncertainty about the current event.

Definition 2.6. An event model is a tuple $\mathcal{E} = (E, \rightarrow, pre, post)$, where

- $(E, \rightarrow)$ is a frame, and an $e \in E$ is called an event;
- For each $e \in E$, $pre(e) \in \mathcal{L}_{KC}$ is its precondition;
- For each $e \in E$, $post(e)$ is its postcondition, and it is in the form of a conjunction of atoms or their negations.

Suppose $\mathcal{P} = \{p, q, r\}$ and $post(e) = p \wedge \neg q$. It means that the execution of e makes p true, q false, and r unchanged.

We let $e_\top$ denote a special event such that $pre(e_\top) = post(e_\top) = true$. Intuitively, $e_\top$ means nothing happens.

Note that the above definition of $post(e)$ follows that in [Bolander and Andersen, 2011]. This form of postconditions is restrictive in that it can only represent context-free actions. Usually, $post(e)$ is defined as a mapping from $\mathcal{P}$ to $\mathcal{L}_{KC}$, and this general form of postconditions can represent context-dependent or context-sensitive actions such as flipping a switch. As shown by Van Ditmarsch and Kooi [2008], for any event model $\mathcal{E}$ with general postconditions, we can construct an equivalent event model $\mathcal{E}'$ as defined in Definition 2.6. However, the number of events in $\mathcal{E}'$ is 2^n times the number of events in $\mathcal{E}$, where $n = |\mathcal{P}|$.

Definition 2.7. An action is a tuple $\alpha{=}(\mathcal{E}, e, actr, cost)$, where

- $\mathcal{E}$ is an event model;
- $e \in E$ is the actual event;
- For each $e \in E$, $actr(e) \in \mathcal{A}$ is the performer of e;
- For each $e \in E$, $cost(e)$, a real number, is the cost of e.

We often omit the $actr$ and $cost$ parts of an action.

Example 2. Consider the watch example. To describe the scenario that Bob moves the watch from Ann's drawer to the surface of the desk while Ann is away, we use action $\alpha_1 = (\mathcal{E}_1, e_1, actr, cost)$ where $\mathcal{E}_1 = (E, \rightarrow, pre, post)$, and

- $E = \{e_\top, e_1\}$, $pre(e_1) = p_a$, $post(e_1) = \neg p_a \wedge p_d$,
- $\rightarrow_a = \{(e_1, e_\top), (e_\top, e_\top)\}$, $\rightarrow_b$ and $\rightarrow_c$ are identities,
- $actr(e_\top) = actr(e_1) = b$, $cost(e_\top) = 0$, $cost(e_1) = 1$.

Intuitively, e_1 is the event of moving the watch and e_1 is the actual event, e_1 is observable to Bob and Carl but not to Ann. We illustrate $\mathcal{E}_1$ and α_1 with Figure 1 where we use a solid dot to represent the actual event.

Below we define the product update operation.

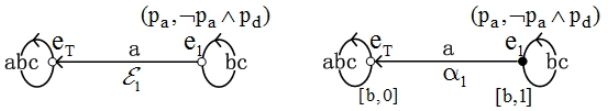

Figure 1: The event model and action model in Example 2

Definition 2.8. Given an epistemic state $s = (M, w)$ and an action $\alpha = (\mathcal{E}, e)$, α is applicable in s if $M, w \models pre(e)$.

Definition 2.9. Given an e-state $s = (M, w)$ and an action $\alpha = (\mathcal{E}, e)$, where $M = (W, R, V)$ and $\mathcal{E} = (E, \rightarrow, pre, post)$, when α is applicable in s, the product of s and α, denoted by $s \otimes \alpha$, is a new e-state $s' = (M', w')$, where $M' = M \otimes \mathcal{E} = (W', R', V')$, $w' = (w, e)$, and

- $W' = \{(w, e) \in W \times E \mid M, w \models pre(e)\}$;
- $(w, e)R'_a(w', e')$ iff wR_aw' and $e \rightarrow_a e'$;
- For each $(w, e) \in W'$, $V'((w, e)) = \{p \in \mathcal{P} \mid p \in V(w)$ and $post(e) \not\models \neg p$, or $post(e) \models p\}$.

Intuitively, (w, e) is the world resulting from doing e in w.

Example 3. Consider the e-state s_0 from Example 1 and the action α_1 from Example 2. Since $s_0 \models p_a$, α_1 is applicable in s_0. The update of s_0 by α_1, $s_0 \otimes \alpha_1$, is a new e-state s_1 where there are two worlds: $(w_0, e_\top)$ and (w_0, e_1), $V((w_0, e_\top)) = \{p_a\}$, $V((w_0, e_1)) = \{p_d\}$, and the accessibility relations are inherited from those of α_1. This is illustrated with Figure 2.

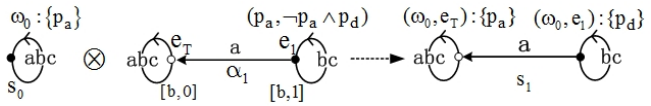

Figure 2: The update of epistemic state in Example 3

Definition 2.10. Let s be an e-state, and $\sigma = \alpha_1 \dots \alpha_n$ an action sequence. We define s_i, $i \leq n$ as follows: $s_0 = s$; for $i < n$, if s_i is defined and α_{i+1} is applicable in s_i, then $s_{i+1} = s_i \otimes \alpha_{i+1}$, otherwise s_{i+1} is undefined. If s_n is defined, we say that σ is applicable in s and define $s \otimes \sigma = s_n$.

Finally, we introduce different types of event models from the DEL literature, and we say an action is of type X if its event model is of type X.

Definition 2.11. An event model is public if it has a single event and all accessibility relations are reflexive.

So there is a single event point, observable to all agents.

Definition 2.12. An event model is (purely) epistemic if all postconditions are $true$. Hence we omit the $post$ part.

So an epistemic event model doesn't change the world.

Definition 2.13. An event model is propositional if all preconditions are propositional formulas.

So all preconditions only depend on the world state.

Definition 2.14. An event model is globally deterministic if all preconditions are pairwise inconsistent, *i.e.*, whenever e and e' are different, $pre(e)$ and $pre(e')$ are inconsistent.

Intuitively, in any world, at most one event can happen.

Definition 2.15. A sensing event model is a globally deterministic epistemic event model where the preconditions are collectively exhaustive, *i.e.*, the disjunction of all preconditions is a valid formula.

Definition 2.16. A secret communication event model is an epistemic event model $(\{e, e_\top\}, \rightarrow, pre)$, where there exists a subset $\mathcal{B}$ of $\mathcal{A}$ such that for all $a \in \mathcal{B}$, $\rightarrow_a$ is identity, and for all $a \notin \mathcal{B}$, $\rightarrow_a = \{e, e_\top\} \times \{e_\top\}$.

So agents in $\mathcal{B}$ are aware that $\phi = pre(e)$ is communicated, but others think that nothing happens. This definition corresponds to the group update program $U_{\mathcal{B}}?\phi$ in Gerbrandy's PhD thesis [1999].

3 Multi-agent epistemic explanatory diagnosis

In this section, we present the formal definition of multi-agent epistemic explanatory diagnosis, and propose several criteria for preferred diagnoses.

Definition 3.1. A diagnosis problem is a tuple $P = (\Delta, s_0, \phi_o)$, where Δ is a finite set of actions, s_0 is the initial epistemic state, and $\phi_o \in \mathcal{L}_{KC}$ is the observation.

Definition 3.2. An explanatory diagnosis for a diagnosis problem $P = (\Delta, s_0, \phi_o)$ is a finite sequence σ of actions from Δ such that σ is applicable in s_0, and $s_0 \otimes \sigma \models \phi_o$.

Thus σ is an explanatory diagnosis if in s_0, it is possible to execute the actions in σ one by one, and in the resulting epistemic state, the observation holds.

Example 4. The diagnosis problem for the watch example is $P = (\Delta, s_0, \phi_o)$, where $\phi_o = \neg p_a \wedge K_a p_a$, s_0 is from Example 1, and we let Δ be the set of the following actions:

- α_1 (α_2): Bob (Carl) moves the watch from Ann's drawer to the surface of the desk (while Ann is away);
- α_3 (α_4): Bob (Carl) moves the watch from the surface of the desk to Ann's drawer (while Ann is away);
- α_5 (α_6): Bob (Carl) moves the watch from the surface of the desk to Bob's (Carl's) drawer (while Ann is away);

Then the following are some solutions for P:
$\sigma_1 = \alpha_1\alpha_5$, $\sigma_2 = \alpha_1\alpha_3\alpha_1\alpha_5$, $\sigma_3 = \alpha_1\alpha_3\alpha_2\alpha_6$.
Figure 3 illustrates the evolution of e-states as the sequence $\alpha_1\alpha_5$ is performed. Clearly, $s_2 \models \neg p_a \wedge K_a p_a$.

As we can see from the above example, for a diagnosis problem, there might very well be many explanatory diagnoses, and there is a need to distinguish between diagnoses of different quality. In all cases, we would prefer diagnoses which avoid conjecturing actions which do not account for the observation. In addition, there may be domain-specific information which we would like to consider in determining preferred diagnoses.

Definition 3.3. Let $\prec$ be a preference relation between diagnoses. Given a diagnosis problem P, σ is a preferred diagnosis for P if σ is a diagnosis for P and there does not exist another diagnosis σ' for P such that $\sigma' \prec \sigma$.

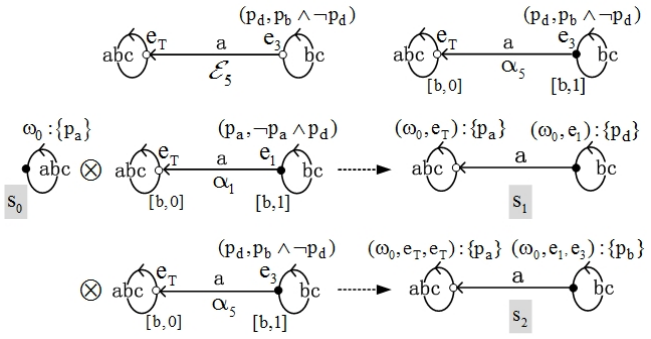

Figure 3: The evolution of epistemic states in Example 4

Next, we propose three domain-independent preference criteria for diagnoses. In many situations, the most important criterion is the length of the diagnosis. There are two other useful criteria: one prefers diagnoses involving fewer agents, and the other prefers diagnoses of lower-cost.

To formalize these criteria, we introduce some notation. Let σ be a finite sequence of actions $\alpha_1 \dots \alpha_n$, where $\alpha_i = (\mathcal{E}_i, e_i, actr_i, cost_i)$, $i = 1, \dots, n$. We use $\|\sigma\|$ to denote the length of the sequence σ. We define $actrs(\sigma)$ as the set $\{actr_i(e_i) \mid i = 1, \dots, n\}$, and $cost(\sigma)$ as $\Sigma_{i=1}^{n} cost_i(e_i)$.

Definition 3.4. Given a diagnosis problem P and two diagnoses σ_1 and σ_2 for it,

1. σ_1 is preferred to σ_2 wrt length if $\|\sigma_1\| < \|\sigma_2\|$;
2. σ_1 is preferred to σ_2 wrt simplicity if $|actrs(\sigma_1)| < |actrs(\sigma_2)|$;
3. σ_1 is preferred to σ_2 wrt cost if $cost(\sigma_1) < cost(\sigma_2)$.

Considering Example 4, we have: σ_1 is preferred to σ_2 and σ_3 wrt length; σ_1 and σ_2 are preferred to σ_3 wrt simplicity.

From our definition, it is easy to see that computing an explanatory diagnosis is analogous to generating a plan to achieve a goal. However, there is a difference between explanatory diagnosis and planning. In planning, a goal might not be achievable. But in diagnosis, an observation is a property of the current e-state, which results from the initial e-state by the execution of a sequence of actions. So if a diagnosis problem is correctly specified, a solution always exists. But a diagnosis problem might be incorrectly specified and if so, a solution may not exist. In this paper, we study the decidability issue of the following problem:

Definition 3.5. *The explanatory diagnosing problem is to decide if a given diagnosis problem has a solution.*

In fact, our definition of explanatory diagnoses is the same as that of epistemic planning solutions by Bolander and Anderson [2011] except that they use local epistemic states. A local epistemic state for agent a is a pair (M, W_a), where $M = (W, R, V)$ is a Kripke model, and $W_a \subseteq W$ is closed under R_a, *i.e.*, if $w \in W_a$ and wR_aw', then $w' \in W_a$. They showed that multi-agent epistemic planning is undecidable even without common knowledge. So in general, explanatory diagnosing is undecidable. In the next two sections, we identify important decidable fragments of the problem.

4 A decidability result in the absence of common knowledge

In this section, we present our result that in the absence of common knowledge, when we only allow propositional actions, explanatory diagnosing is decidable. Our result is inspired by two decidability results from [Bolander and Andersen, 2011] concerning multi-agent epistemic planning. So we introduce their results first.

The basic idea of their results is this. Given a multi-agent epistemic planning problem, we perform a search in the space of e-states. In general, this is an infinite search space. Under conditions ensuring that the number of non-isomorphic e-states is finite, planning is decidable. Better yet, under conditions guaranteeing that the number of non-equivalent e-states is finite, planning is also decidable. In the following, we rephrase their results in the context of explanatory diagnosis, and briefly present the proofs.

Proposition 4.1. *Given a natural number n, the number of non-isomorphic e-states with exactly n worlds is $\leq n \cdot 2^{n^2}$.*

Theorem 4.2. *Explanatory diagnosing is decidable when only globally deterministic actions are allowed.*

Proof. Given such a diagnosis problem, we perform a search in the space of e-states. For a globally deterministic event model, in any world, at most one event can happen. Hence when we update an e-state by an action, the number of worlds won't increase. Suppose that the initial e-state has n worlds. Then any possible e-state has no more than n worlds. By Proposition 4.1, the e-state space is finite. □

Examples of globally deterministic actions are public actions and sensing actions.

We now introduce the well-known concept of bisimulation, which is needed in the rest of this paper. For a reference on modal logic, see for example [Blackburn *et al.*, 2002].

Definition 4.3. Let $s = (M, w)$ and $s' = (M', w')$ be two e-states, where $M = (W, R, V)$ and $M' = (W', R', V')$. A bisimulation between s and s' is a relation $\rho \subseteq W \times W'$ s.t. $w \rho w'$, and whenever $u \rho u'$, $V(u) = V'(u')$, and

- The forth condition: for all agents $a \in \mathcal{A}$, if uR_av, then there is v' (called the forth witness) s.t. $u'R'_av'$ and $v \rho v'$;
- The back condition: for all agents $a \in \mathcal{A}$, if $u'R'_av'$, then there is v (called the back witness) s.t. uR_av and $v \rho v'$.

We say that s and s' are bisimilar, written $s \underline{\leftrightarrow} s'$, if there is a bisimulation relation between s and s'.

A nice property of bisimilar epistemic states is that they agree on all epistemic formulas:

Proposition 4.4. *Let s and s' be two epistemic states s.t. $s \underline{\leftrightarrow} s'$. Then for any $\phi \in \mathcal{L}_{KC}$, $s \models \phi$ iff $s' \models \phi$.*

Let $M = (W, R, V)$ be a Kripke model. We can define an equivalence relation on W as follows: $w \underline{\leftrightarrow} v$ iff $M, w \underline{\leftrightarrow} M, v$. Then we can construct the quotient structure of M wrt $\underline{\leftrightarrow}$, which turns out to be bisimilar to M.

Definition 4.5. Given a Kripke model $M = (W, R, V)$, the bisimulation contraction of M is the quotient structure $M_{\underline{\leftrightarrow}} = (W', R', V')$, where

- $W' = \{[w]_{\underline{\leftrightarrow}} \mid w \in W\}$, here $[w]_{\underline{\leftrightarrow}} = \{v \in W \mid (M,w)\underline{\leftrightarrow}(M,v)\}$;
- For all agents $a \in \mathcal{A}$, $[w]_{\underline{\leftrightarrow}} R'_a [v]_{\underline{\leftrightarrow}}$ iff there are $w' \in [w]_{\underline{\leftrightarrow}}$ and $v' \in [v]_{\underline{\leftrightarrow}}$ such that $w' R_a v'$;
- For all $w \in W$, $V'([w]_{\underline{\leftrightarrow}}) = V(w)$.

Proposition 4.6. $(M_{\underline{\leftrightarrow}}, [w]_{\underline{\leftrightarrow}})\underline{\leftrightarrow}(M,w)$.

We are now ready to present the second result.

Proposition 4.7. *In the single-agent case, the number of non-bisimilar S5, KD45 or K45 e-states is $\leq 2^n \cdot 2^{2^n}$, here $n = |\mathcal{P}|$.*

Proof. In the single-agent case, it is easy to show: (1) Each S5 e-state is bisimilar to an e-state $s = ((W,R,V),w)$, where $R = W \times W$, and V is an injection. (2) Each KD45 or K45 e-state is bisimilar to an e-state $s = ((\{w\} \cup W, R, V), w)$, where $W \neq \emptyset$ in the case of KD45, $R = (\{w\} \cup W) \times W$, and the restriction of V to W is an injection. □

Theorem 4.8. *Single-agent explanatory diagnosing is decidable when the frames are all S5, all KD45, or all K45.*

Proof. It is easy to show that bisimilarity of e-states is preserved under product update. So by Propositions 4.4 and 4.6, when we perform search of the e-state space, we can replace each e-state by its bisimulation contraction. By Proposition 4.7, we get a finite search space. □

Now we proceed to present our result. Theorem 4.8 holds because there are a finite number of non-bisimilar e-states, which does not hold in the multi-agent case. However, if the observations do not contain common knowledge, a weak notion of bisimilarity would suffice. Below we present the definition of k-bisimulation and the property that two k-bisimilar e-states agree on all formulas of $\mathcal{L}_K$ with modal depth bounded by k from [Blackburn *et al.*, 2002]. We will show that the k-bisimulation contraction of a Kripke model is k-bisimilar to itself, there are a finite number of non-k-bisimilar e-states, and k-bisimilarity is preserved under update by propositional actions. With these results, we can show that explanatory diagnosing is decidable when observations do not contain common knowledge and only propositional actions are allowed.

Definition 4.9. Let $s = (M,w)$ and $s' = (M',w')$ be two e-states, where $M = (W,R,V)$ and $M' = (W',R',V')$. We say that s and s' are k-bisimilar, written $s\underline{\leftrightarrow}_k s'$, if $V(w) = V'(w')$, and either $k = 0$ or the following conditions hold:

- The forth condition: for all agents $a \in \mathcal{A}$, if $wR_a v$, then there is v' s.t. $w' R'_a v'$ and $(M,v)\underline{\leftrightarrow}_{k-1}(M',v')$;
- The back condition: for all agents $a \in \mathcal{A}$, if $w' R'_a v'$, then there is v s.t. $wR_a v$ and $(M,v)\underline{\leftrightarrow}_{k-1}(M',v')$.

Clearly, for any $k \geq 0$, if $s\underline{\leftrightarrow}_{k+1}s'$, then $s\underline{\leftrightarrow}_k s'$.

The modal depth of a formula ϕ, denoted by $md(\phi)$, is the depth of nesting of modal operators in ϕ.

Proposition 4.10. *Let s and s' be two e-states s.t. $s\underline{\leftrightarrow}_k s'$. Then for any $\phi \in \mathcal{L}_K$ s.t. $md(\phi) \leq k$, $s \models \phi$ iff $s' \models \phi$.*

Following the idea of bisimulation contraction, we define:

Definition 4.11. The k-bisimulation contraction of a Kripke model M is the quotient structure of M wrt the $\underline{\leftrightarrow}_k$ relation.

The proposition below shows that the k-bisimulation contraction of a Kripke model M is k-bisimilar to M itself.

Proposition 4.12. *Let (M,w) be an e-state. Then for any j,k such that $j \geq k \geq 0$, $(M_{\underline{\leftrightarrow}_j}, [w]_{\underline{\leftrightarrow}_j})\underline{\leftrightarrow}_k(M,w)$.*

Proof. Let $M = (W,R,V)$. To simplify the notation, we denote $M_{\underline{\leftrightarrow}_j}$ by M_j, and $[w]_{\underline{\leftrightarrow}_j}$ by $[w]_j$. We prove by induction on k. The base case is obvious. Assume that the statement holds for k. Then by transitivity of $\underline{\leftrightarrow}_k$, we have for any $j_1, j_2 \geq k$, $(M_{j_1}, [w]_{j_1})\underline{\leftrightarrow}_k(M_{j_2}, [w]_{j_2})$. We prove that the statement holds for $k+1$. Let $j \geq k+1$. Let $M_j = (W', R', V')$. First, $V'([w]_j) = V(w)$. (1) Suppose $wR_a v$. Then $[w]_j R'_a [v]_j$. By induction hypothesis, $(M_j, [v]_j)\underline{\leftrightarrow}_k(M,v)$. (2) Suppose $[w]_j R'_a [v]_j$. Then there are $w^* \in [w]_j$ and $v^* \in [v]_j$ s.t. $w^* R_a v^*$. Since $(M,w)\underline{\leftrightarrow}_j(M,w^*)$, there is v' s.t. $wR_a v'$ and $(M,v')\underline{\leftrightarrow}_{j-1}(M,v^*)$. Since $(M,v)\underline{\leftrightarrow}_j(M,v^*)$, $(M,v)\underline{\leftrightarrow}_{j-1}(M,v^*)$. So $(M,v)\underline{\leftrightarrow}_{j-1}(M,v')$, hence $[v]_{j-1} = [v']_{j-1}$. By induction hypothesis, $(M_{j-1}, [v']_{j-1})\underline{\leftrightarrow}_k(M,v')$, and $(M_j, [v]_j)\underline{\leftrightarrow}_k (M_{j-1}, [v]_{j-1})$. Since $[v]_{j-1} = [v']_{j-1}$, we have $(M_j, [v]_j)\underline{\leftrightarrow}_k(M,v')$. Thus, $(M_j, [w]_j)\underline{\leftrightarrow}_{k+1}(M,w)$. □

Visser [1996] showed that in the single-agent case, the number of $\underline{\leftrightarrow}_k$ equivalence classes of any Kripke model is bounded by a number which only depends on k. In the following, we extend his result to the multi-agent case:

Proposition 4.13. *Let M be a Kripke model. We use $F_k(M)$ to denote the number of $\underline{\leftrightarrow}_k$ equivalence classes of worlds of M. Suppose $|\mathcal{A}| = m$ and $|\mathcal{P}| = n$. Let $f(0) = 2^n$, and $f(i+1) = 2^{mf(i)+n}$ for $i \geq 0$. Then $F_k(M) \leq f(k)$.*

Proof. Let $M = (W,R,V)$. We prove by induction on k. Base case: $k = 0$. For any $w_1, w_2 \in W$, if $V(w_1) = V(w_2)$, then $M, w_1 \underline{\leftrightarrow}_0 M, w_2$. There are at most 2^n different possibilities for $V(w), w \in W$. Thus $F_0(M) \leq 2^n = f(0)$. Inductive step: Assume that the statement holds for k. We prove that it holds for $k+1$. For any $w \in W$, $a \in \mathcal{A}$, let $\mathcal{R}_a(w)$ denote the set $\{[w']_{\underline{\leftrightarrow}_k} \mid wR_a w'\}$. Then we have: for any $w_1, w_2 \in W$, if $V(w_1) = V(w_2)$ and $\mathcal{R}_a(w_1) = \mathcal{R}_a(w_2)$ for all $a \in \mathcal{A}$, then $(M,w_1)\underline{\leftrightarrow}_{k+1}(M,w_2)$. For each $a \in \mathcal{A}$, there are at most $2^{F_k(M)}$ different possibilities for $\mathcal{R}_a(w)$. So $F_{k+1}(M) \leq 2^n \cdot (2^{F_k(M)})^m \leq 2^{mf(k)+n} = f(k+1)$. □

The next proposition states that k-bisimilarity between e-states is preserved under update by propositional actions.

Proposition 4.14. *Let $s = (M,w)$ and $s' = (M',w')$ be two e-states such that $s\underline{\leftrightarrow}_k s'$. Let $\alpha = (\mathcal{E}, e)$ be a propositional action. Then α is applicable in s iff α is applicable in s'; and if α is applicable in s, then $s \otimes \alpha \underline{\leftrightarrow}_k s' \otimes \alpha$.*

Proof. Let $M = (W,R,V)$, $M' = (W',R',V')$, and $\mathcal{E} = (E, \rightarrow, pre, post)$. Since $(M,w)\underline{\leftrightarrow}_k(M',w')$, $V(w) = V'(w')$. Since $pre(e)$ is propositional, $M,w \models pre(e)$ iff $M',w' \models pre(e)$. Hence α is applicable in s iff α is applicable in s'. Now suppose α is applicable in s. Let $M \otimes \mathcal{E} = (W^*, R^*, V^*)$ and $M' \otimes \mathcal{E} = (W'^*, R'^*, V'^*)$. We prove $s \otimes \alpha \underline{\leftrightarrow}_k s' \otimes \alpha$ by induction on k. Since $(M,w)\underline{\leftrightarrow}_0(M',w')$, $V(w) = V'(w')$. So $V^*((w,e)) = V'^*((w',e))$. So we have

proved the base case: $k = 0$. Induction step: Assume that the statement holds for k. We prove that it also holds for $k + 1$. We first prove the forth condition. Suppose $(w_1, e_1) \in W^*$ and $(w, e)R_a^*(w_1, e_1)$. Then $M, w_1 \models pre(e_1)$, wR_aw_1 and $e \to_a e_1$. Since $(M, w)\underline{\leftrightarrow}_{k+1}(M', w')$, there exists w_1' such that $w'R_aw_1'$ and $(M, w_1)\underline{\leftrightarrow}_k(M', w_1')$. By induction hypothesis, $M \otimes \mathcal{E}, (w_1, e_1)\underline{\leftrightarrow}_k M' \otimes \mathcal{E}, (w_1', e_1)$. Meanwhile, $(w', e)R_a'^*(w_1', e_1)$. The back condition can be similarly proved. Thus, $M \otimes \mathcal{E}, (w, e)\underline{\leftrightarrow}_{k+1}M' \otimes \mathcal{E}, (w', e)$. □

It is important to note that without the restriction to propositional actions, Proposition 4.14 does not hold. We illustrate this with the following example:

Example 5. Let $\mathcal{A} = \{a\}$ and $\mathcal{P} = \{p\}$. Figure 4 illustrates two e-states s_1 and s_2, action α, and the two e-states $s_1 \otimes \alpha$ and $s_2 \otimes \alpha$. Here α has a single event e with $pre(e) = K_ap$. $s_2 \otimes \alpha$ has a single state (v, e), since e is not executable in v_1 or v_2. Clearly, $s_1\underline{\leftrightarrow}_1 s_2$, but not $s_1 \otimes \alpha\underline{\leftrightarrow}_1 s_2 \otimes \alpha$.

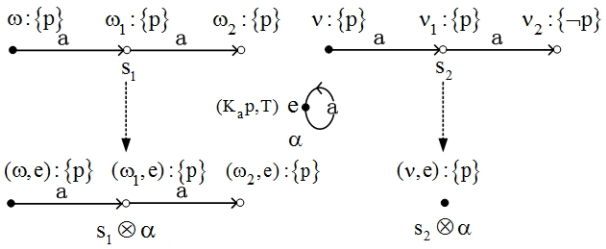

Figure 4: The update of epistemic states in Example 5

We are now ready to present the main result of this section:

Theorem 4.15. *Explanatory diagnosing is decidable when observations are in $\mathcal{L}_K$ and all actions are propositional.*

Proof. Let $P = (\Delta, s_0, \phi_o)$ be such a diagnosis problem. Let $k = md(\phi_o)$. By Propositions 4.10 and 4.12, any e-state s and its k-bisimulation contraction are k-bisimilar and hence agree on ϕ_o. By Proposition 4.14, k-bisimularity is preserved under update by propositional actions. Thus, when we perform search of the e-state space, we can replace each e-sate by its k-bisimulation contraction. By Propositions 4.13 and 4.1, there are only finitely many non-isomorphic k-bisimulation contractions. So we get a finite search space. □

5 A decidability result in the presence of common knowledge

In this section, we present a decidability result of the explanatory diagnosing problem in the presence of common knowledge where we allow only propositional epistemic actions. We will use PE to abbreviate for "propositional epistemic". Our result is inspired by the work of Löwe *et al.* [2011], which we introduce first.

Löwe *et al.* identified two important properties of event models, defined as follows:

Definition 5.1. *Two actions α_1 and α_2 commute if for all e-states s, we have $(s \otimes \alpha_1) \otimes \alpha_2\underline{\leftrightarrow}(s \otimes \alpha_2) \otimes \alpha_1$.*

Let α be an action, and $n \geq 1$. We use α^n to denote the sequence consisting of exactly n α's.

Definition 5.2. An action α is called self-absorbing if for all e-states s, $s \otimes \alpha^2\underline{\leftrightarrow}s \otimes \alpha$.

If all actions are self-absorbing and commute, then multi-agent epistemic planning is decidable, since we can reduce the search space of action sequences to the finite space of action sequences where each action appears at most once. They showed that PE actions commute and identified a special type of PE actions which are self-absorbing.

Proposition 5.3. *Propositional epistemic actions commute.*

Definition 5.4. An epistemic event model $\mathcal{E} = (E, \to, pre)$ is almost-mutex if $e_\top \in E$, $e_\top \to_a e_\top$ for all $a \in \mathcal{A}$, and the formulas $pre(e)$ with $e \neq e_\top$ are pairwise inconsistent.

Definition 5.5. A propositional epistemic event model is called almost-mutex transitive if it is almost-mutex and all accessibility relations are transitive.

Proposition 5.6. *Almost-mutex transitive propositional epistemic actions are self-absorbing.*

In the following, we will identify a wide variety of special types of PE actions, and show that any action α of these types is self-absorbing in the general sense, *i.e.*, there exists a number n which depends only on α, such that for any e-state s, we have $s\otimes\alpha^{n+1}\underline{\leftrightarrow}s\otimes\alpha^n$. Thus we will be able to show that explanatory diagnosing is decidable when only these types of PE actions are allowed.

We first state a simple property of PE actions:

Proposition 5.7. *Let $s = (M, w_0)$ be an e-state where $M = (W, R, V)$, and $\alpha = (\mathcal{E}, e_0)$ a propositional epistemic action where $\mathcal{E} = (E, \to, pre)$. Suppose that α^n is applicable in s. Then for any $w \in W$, any $\tau = (e_1, \ldots, e_n) \in E^n$, (w, τ) is a world of $s \otimes \alpha^n$ iff $w \models pre(\tau)$, where $pre(\tau)$ denotes $pre(e_1) \wedge \ldots \wedge pre(e_n)$.*

It is easy to prove the following:

Proposition 5.8. *Public propositional epistemic actions are self-absorbing.*

Definition 5.9. An epistemic event model $\mathcal{E} = (E, \to, pre)$ is possibly oblivious if $e_\top \in E$, and for all $e \in E$ and all $a \in \mathcal{A}$, $e \to_a e_\top$.

Intuitively, a possibly oblivious event model is one where when any event happens, any agent thinks it is possible that nothing happens. It might be strange that the performer of the event also thinks it is possible that nothing happens. But this is the case for fallible events.

Proposition 5.10. *Possibly oblivious propositional epistemic actions are self-absorbing.*

Proof. Let $\alpha = (\mathcal{E}, e_0)$ be a possibly oblivious PE action where $\mathcal{E} = (E, \to, pre)$. Let $s = (M, w_0)$ be an e-state where $M = (W, R, V)$. We show that $s \otimes \alpha^2\underline{\leftrightarrow}s \otimes \alpha$. Let $M \otimes \mathcal{E} = M' = (W', R', V')$ and $M' \otimes \mathcal{E} = M'' = (W'', R'', V'')$. We define $\rho \subseteq W'' \times W'$ as follows: $\rho = \{((w, e_1, e_2), (w, e)) \mid e = e_1 \text{ or } e = e_2\}$. We show that ρ is a bisimulation between $s \otimes \alpha^2$ and $s \otimes \alpha$. First, $(w_0, e_0, e_0)\rho(w_0, e_0)$. Now suppose $((w, e_1, e_2), (w, e)) \in$

ρ. Since $\mathcal{E}$ is epistemic, we have $V'((w,e)) = V(w) = V''((w,e_1,e_2))$. Suppose $(w,e_1,e_2)R''_a(w',e'_1,e'_2)$. Since $(w',e'_1,e'_2) \in W''$, we have $w' \models pre(e'_1) \wedge pre(e'_2)$. If $e = e_1$, let $e' = e'_1$, else let $e' = e'_2$. Then (w',e') is the forth witness. Suppose $(w,e)R'_a(w',e')$. Since $\mathcal{E}$ is possibly oblivious, $e_\top \in E$, and for all $e \in E$, $e \to_a e_\top$. If $e = e_1$, let $e'_1 = e'$ and $e'_2 = e_\top$, else let $e'_1 = e_\top$ and $e'_2 = e'$. Then (w',e'_1,e'_2) is the back witness. □

Definition 5.11. A binary relation R is functional dependent if whenever xRy and xRz, we have $y = z$. An event model is called functional dependent if all accessibility relations are.

Intuitively, a functional dependent event model is one where when any event happens, any agent is certain but may be mistaken about the current event. An example of such event models is secret communication.

Definition 5.12. An event model is binary if it has exactly two events.

An example of binary event models is binary sensing.

Definition 5.13. A binary relation R is divergent if every element is related to at least two distinct elements. An event model is triple dichotomous if it has exactly three events and every accessibility relation is either functional dependent or divergent.

Intuitively, a triple dichotomous event model is one where there are only two types of agents: the first is certain about the current event when any event happens, and the second is uncertain about the current event when any event happens.

Proposition 5.14. *Let $\alpha = (\mathcal{E}, e_0)$ be a propositional epistemic action which is functional dependent, or binary, or triple dichotomous. Let n be the number of events in $\mathcal{E}$. Then for all e-states s, $s \otimes \alpha^{n+1} \underline{\leftrightarrow} s \otimes \alpha^n$.*

Proof. Let $\mathcal{E} = (E, \to, pre)$. Let $s = (M, w_0)$ be an e-state where $M = (W, R, V)$. Let $M \otimes \mathcal{E}^n = (W', R', V')$ and $M \otimes \mathcal{E}^{n+1} = (W'', R'', V'')$. We define $\rho \subseteq W'' \times W'$ as follows: $\rho = \{((w,\tau_1),(w,\tau_2)) \mid set(\tau_1) = set(\tau_2)\}$, where $set(\tau)$ is the set of elements occurring in τ. We show that ρ is a bisimulation between $s \otimes \alpha^{n+1}$ and $s \otimes \alpha^n$. First, $(w_0, e_0^{n+1})\rho(w_0, e_0^n)$. Now suppose $((w,\tau_1),(w,\tau_2)) \in \rho$. Since $\mathcal{E}$ is epistemic, $V''((w,\tau_1)) = V(w) = V'((w,\tau_2))$. Let $\tau_1 = (e_1, \ldots, e_n, e_{n+1})$. Since $|E| = n$ and $set(\tau_1) = set(\tau_2)$, there exist i, j s.t. $1 \le i \le j \le n+1$, $e_i = e_j$, and τ_2 is obtained from τ_1 by removing e_i and possibly permuting the other elements. Since $\mathcal{E}$ is propositional epistemic, the order of events does not matter. Without loss of generality, assume that $e_n = e_{n+1}$ and $\tau_2 = (e_1, \ldots, e_n)$. Suppose $(w,\tau_2)R'_a(w',\tau'_2)$ where $\tau'_2 = (e'_1, \ldots, e'_n)$. Let $\tau'_1 = (e'_1, \ldots, e'_n, e'_n)$. Since $(w',\tau'_2) \in W'$, $w' \models pre(\tau'_2)$, hence $w' \models pre(\tau'_1)$, so $(w',\tau'_1) \in W''$. Then (w',τ'_1) is the back witness. Finally, suppose $(w,\tau_1)R''_a(w',\tau'_1)$ where $\tau'_1 = (e'_1, \ldots, e'_n, e'_{n+1})$. We show there is τ'_2 so that (w',τ'_2) is the forth witness. We consider three cases: (1) $\mathcal{E}$ is functional dependent. Since $e_n = e_{n+1}$, $e_n \to_a e'_n$, and $e_{n+1} \to_a e'_{n+1}$, we have $e'_n = e'_{n+1}$. We let $\tau'_2 = (e'_1, \ldots, e'_n)$.
(2) $\mathcal{E}$ is binary. So $n = 2$. If $e'_2 = e'_3$, the proof is the same as in (1). Otherwise, $e'_2 \neq e'_3$. Since $n = 2$, $e'_1 = e'_2$ or $e'_1 = e'_3$. If $e'_1 = e'_2$, let $\tau'_2 = (e'_1, e'_3)$, else let $\tau'_2 = (e'_1, e'_2)$.
(3) $\mathcal{E}$ is triple dichotomous, so $n = 3$. If $e'_3 = e'_4$, the proof is the same as in (1). Otherwise, $e'_3 \neq e'_4$, so $\to_a$ is divergent. Since $n = 3$, $e_2 \to_a e'_3$ or $e_2 \to_a e'_4$. If $e_2 \to_a e'_3$, we let $\tau'_2 = (e'_1, e'_3, e'_4)$, otherwise let $\tau'_2 = (e'_1, e'_4, e'_3)$. □

We now present the main result of this section.

Theorem 5.15. *Explanatory diagnosing is decidable when only the following types of propositional epistemic actions are allowed: public, almost-mutex transitive, possibly oblivious, functional dependent, binary, and triple dichotomous.*

Proof. Give a diagnosis problem, we perform a search in the space of action sequences. Note that bisimilar e-states agree on all formulas of $\mathcal{L}_{KC}$. By Proposition 5.3, PE actions commute. So we can restrict the search space to sequences of powers of different actions, *i.e.*, sequences of the form: $\alpha_1^{k_1} \ldots \alpha_n^{k_n}$, where for any $i \neq j$, $\alpha_i \neq \alpha_j$. By Propositions 5.6, 5.8, 5.10, and 5.14, we can further restrict the search space so that each power k_i is bounded by 1 if the action α_i is almost-mutex transitive or possibly oblivious, or by n where n is the number of events in the action. Since there are only finitely many actions, we get a finite search space. □

So we have shown that for a wide range of propositional epistemic actions, explanatory diagnosing is decidable in the presence of common knowledge.

6 Conclusions

Explanatory diagnosis is a common task in real life and an essential ability of intelligent agents. In this paper, we have formally defined multi-agent epistemic explanatory diagnosis in the framework of dynamic epistemic logic. Since explanatory diagnosing is undecidable in general, we identify important decidable fragments of it. Our first result is that when observations do not contain common knowledge and all actions are propositional, explanatory diagnosing is decidable. Our second result is that in the presence of common knowledge, for a wide range of propositional epistemic actions, explanatory diagnosing is decidable. We would like to remark that these results carry over to multi-agent epistemic planning. There are also two results inherited from the literature: the first is that explanatory diagnosing is decidable when all actions are globally deterministic; the other is that single-agent explanatory diagnosing is decidable when the frames are all S5, all KD45, or all K45. For the future, we would like to identify decidable fragments of explanatory diagnosing that go beyond propositional actions, investigate the computational complexity of explanatory diagnosing for the decidable fragments we have identified, implement multi-agent explanatory diagnosis solver, and apply it to multi-agent high-level program execution.

Acknowledgments

We thank the anonymous reviewers for helpful comments. This work was supported by the Natural Science Foundation of China under Grant No. 61073053. The first author was also supported by the Natural Science Foundation of Guizhou Province, China under Grant No. [2012]2310.

References

[Blackburn *et al.*, 2002] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, 2002.

[Bolander and Andersen, 2011] T. Bolander and M. B. Andersen. Epistemic planning for single and multi-agent systems. *Journal of Applied Non-Classical Logics*, 21(1):9–34, 2011.

[Gerbrandy, 1999] J. Gerbrandy. *Bisimulations on Planet Kripke*. PhD thesis, ILLC, Amsterdam, 1999.

[Grastien *et al.*, 2007] A. Grastien, Anbulagan, J. Rintanen, and E. Kelareva. Diagnosis of discrete-event systems using satisfiability algorithms. In *AAAI*, pages 305–310, 2007.

[Löwe *et al.*, 2011] Benedikt Löwe, Eric Pacuit, and Andreas Witzel. Del planning and some tractable cases. In *Proc. of 3rd Intl. Workshop on Logic, Rationality and Interaction (LORI-III)*, pages 179–192, 2011.

[McIlraith, 1998] S. A. McIlraith. Explanatory diagnosis: Conjecturing actions to explain observations. In *KR*, pages 167–179, 1998.

[Pencolé and Cordier, 2005] Y. Pencolé and M. Cordier. A formal framework for the decentralised diagnosis of large scale discrete event systems and its application to telecommunication networks. *Artif. Intell.*, 164(1-2):121–170, 2005.

[Reiter, 1987] R. Reiter. A theory of diagnosis from first principles. *Artif. Intell.*, 32(1):57–95, 1987.

[Rintanen and Grastien, 2007] J. Rintanen and A. Grastien. Diagnosability testing with satisfiability algorithms. In *IJCAI*, pages 532–537, 2007.

[Sampath *et al.*, 1995] M. Sampath, R. Sengupta, S. Lafortune, K. Sinnamohideen, and D. Teneketzis. Diagnosability of discrete-event systems. *IEEE Transactions on Automatic Control*, 40(9):1555–1575, 1995.

[Sohrabi *et al.*, 2010] S. Sohrabi, J. A. Baier, and S. A. McIlraith. Diagnosis as planning revisited. In *KR*, 2010.

[Sohrabi *et al.*, 2011] S. Sohrabi, J. A. Baier, and S. A. McIlraith. Preferred explanations: Theory and generation via planning. In *AAAI*, 2011.

[van Ditmarsch and Kooi, 2008] H. van Ditmarsch and B. P. Kooi. Semantic results for ontic and epistemic change. In G. Bonanno, W. van der Hoek, and M. Wooldridge, editors, *Logic and the Foundation of Game and Decision Theory (LOFT 7), Texts in Logic and Games 3*, pages 87–117. Amsterdam University Press, 2008.

[van Ditmarsch *et al.*, 2007] H. van Ditmarsch, W. van der Hoek, and B. P. Kooi. *Dynamic epistemic logic*. Springer, 2007.

[Visser, 1996] A. Visser. Uniform interpolation and layered bisimulation. In *Gödel 96 (Brno,1996)*, volume 6 of Lecture Notes Logic. Springer, 1996.