

# Implicit Coordination Using FOND Planning

**Thorsten Engesser**

Department of Computer Science  
University of Freiburg, Germany  
engesser@cs.uni-freiburg.de

**Tim Miller**

School of Computing and Information Systems  
University of Melbourne, Australia  
tmiller@unimelb.edu.au

## Abstract

Epistemic planning can be used to achieve implicit coordination in cooperative multi-agent settings where knowledge and capabilities are distributed between the agents. In these scenarios, agents plan and act on their own without having to agree on a common plan or protocol beforehand. However, epistemic planning is undecidable in general. In this paper, we show how implicit coordination can be achieved in a simpler, propositional setting by using nondeterminism as a means to allow the agents to take the other agents' perspectives. We identify a decidable fragment of epistemic planning that allows for arbitrary initial state uncertainty and nondeterminism, but where actions can never increase the uncertainty of the agents. We show that in this fragment, planning for implicit coordination can be reduced to a version of fully observable nondeterministic (FOND) planning and that it thus has the same computational complexity as FOND planning. We provide a small case study, modeling the problem of multi-agent path finding with destination uncertainty in FOND, to show that our approach can be successfully applied in practice.

## Introduction

Epistemic planning has gained increasing interest in recent years (Baral et al. 2017). One of the main features of epistemic planning is the support of *knowledge goals*. For example, epistemic planning is well-suited to model problems in which information is to be confidentially passed between agents. The assumption is usually that the agents act with respect to a centralized plan. However, recent work has shown that epistemic planning can also be used to achieve *implicit coordination* in a setting where multiple agents plan and act for themselves towards a joint goal (Engesser et al. 2017). The idea is that the explicit modeling of the agents' knowledge can be exploited as a means to enforce coordination via *perspective taking*. In particular, by putting themselves into the shoes of the others, agents can account for possible contributions of other agents in their own plans. Bolander et al. (2018) showed under which conditions such plans are guaranteed to be successful.

One example for the application of implicit coordination is the *multi-agent path finding with destination uncertainty* (MAPF/DU) problem, which was first described by Bolander et al. (2018) and analyzed more thoroughly by Nebel et al. (2019). It is a generalization of the multi-agent path finding problem, relaxing the assumption that the agents' destinations are commonly known. Instead, it is assumed that each agent has a set of *potential destinations* which is commonly known and which is disjoint to each of the other agents' potential destination sets. Also, each agent knows its own actual destination. As their final action, agents are allowed to reveal that they have arrived at their actual destination. Using implicit coordination, agents will clear the way for one agent to reach any of its potential destinations, if they know that after the agent has moved to any of these destination candidates and announces success, the remaining problem can be solved analogously.

Another interesting application for implicit coordination is the cooperative card game Hanabi, which has gained some attention recently (Bard et al. 2019). The objective is to score points by putting down the cards on the table in an ascending order. However, players only see the other players' cards. They have to use actions with which they can hint the colors and values of cards to the others. While the concept of implicit coordination by epistemic perspective taking has been defined only for achievement goals so far, we think that it can be generalized to optimization problems like maximizing the expected score in Hanabi (Reifsteck et al. 2019).

The problem of planning for implicit coordination was originally formalized as a variant of contingent planning in the space of epistemic states (i.e., Kripke models), with actions represented by the action models from *Dynamic Epistemic Logic* (DEL) (van Ditmarsch, van der Hoek, and Kooi 2007). The formalization is similar to the one of Bolander and Andersen (2011), which produces sequential plans. Bolander and Andersen have shown that this type of epistemic planning is undecidable in general. However, some decidable fragments have been identified that rely on restricting the structure of action models and the form of allowed preconditions (Aucher and Bolander 2013; Bolander, Jensen, and Schwarzenrüber 2015; Charrier, Maubert, and Schwarzenrüber 2016). On the practical side, Kominis and

Geffner (2015) and Muise et al. (2015) have identified fragments of epistemic planning that can be solved by compilation to classical planning. Besides cooperative problems, the DEL formalism has also been used to model adversarial problems (Maubert, Pinchinat, and Schwarzenrüber 2019).

In this paper, we look at epistemic planning for implicit coordination in a propositional setting. The key insight is that we can use nondeterminism to simulate perspective taking. We characterize a fragment of DEL for which the problem of planning for implicit coordination corresponds, in terms of computational complexity, to FOND planning. Finally, we provide a small case study modeling and solving instances of multi-agent path finding with destination uncertainty in FOND.

## Theoretical Background

We will first recapitulate the DEL planning framework using the conventions of Bolander and Andersen (2011), but including conditional effects in the style of van Benthem, van Eijck, and Kooi (2006). We will then review strong fully observable nondeterministic planning (Cimatti et al. 2003; Ghallab, Nau, and Traverso 2004) and planning for implicit coordination (Engesser et al. 2017; Bolander et al. 2018).

### The DEL Planning Framework

For a fixed set of agents  $\mathcal{A}$  and a fixed set of atomic propositions  $P$  (which we both assume to be finite), the epistemic language  $\mathcal{L}_{KC}$  is given by the BNF

$$\phi ::= p \mid \neg\phi \mid \phi \wedge \phi \mid K_i\phi \mid C\phi, \text{ where } p \in P \text{ and } i \in \mathcal{A}.$$

We read  $K_i\phi$  as “agent  $i$  knows  $\phi$ ” and  $C\phi$  as “it is common knowledge between all agents that  $\phi$ ”. The additional connectives  $\vee$ ,  $\leftarrow$ ,  $\rightarrow$ ,  $\leftrightarrow$  can be defined as abbreviations, analogously to their definition in propositional logic.

We evaluate such formulas over *epistemic models*. An epistemic model is a tuple  $\mathcal{M} = \langle W, (R_i)_{i \in \mathcal{A}}, V \rangle$ , where  $W$  is a non-empty, finite set of worlds (the *domain* of  $\mathcal{M}$ ),  $R_i \subseteq W \times W$  is an equivalence relation for each agent  $i \in \mathcal{A}$  (the *indistinguishability relation* of  $i$ ), and  $V : P \rightarrow 2^W$  (the *valuation function*). We write  $R^+$  for the transitive closure of  $\bigcup_{i \in \mathcal{A}} R_i$ . The truth of a formula  $\phi \in \mathcal{L}_{KC}$  in a world  $w$  of a model  $\mathcal{M}$  is then given as follows, where the propositional cases are standard and hence left out:

$$\begin{aligned} \mathcal{M}, w \models p & \quad \text{iff } w \in V(p) \\ \mathcal{M}, w \models K_i\phi & \quad \text{iff } \mathcal{M}, w' \models \phi \text{ for all } wR_iw' \\ \mathcal{M}, w \models C\phi & \quad \text{iff } \mathcal{M}, w' \models \phi \text{ for all } wR^+w' \end{aligned}$$

We depict epistemic models as graphs where nodes correspond to the worlds in the model and are additionally labeled with the atomic propositions that are true in that particular world. The indistinguishability relations are given as labeled edges between the worlds. For readability, we will omit reflexive edges as well as edges that are implied by transitivity. Consider the following epistemic model:

$$\mathcal{M}_0 = \begin{array}{ccc} & \bullet & \xrightarrow{1,2} & \bullet \\ & w_1 : p & & w_2 : \end{array}$$

In our example, both agents 1 and 2 do not know whether or not  $p$  is true (which is the case in  $w_1$ ) or false (which is the

case in  $w_2$ ). Also, it is common knowledge between the two agents that they do not know. We will now define example actions for agent 1, first to sense the value of  $p$  and then to announce it to agent 2.

To define actions, we use *event models*. These can change the facts about the world as well as the knowledge of the agents. Analogous to epistemic models, an event model is a tuple  $\mathcal{E} = \langle E, (Q_i)_{i \in \mathcal{A}}, \text{pre}, \text{eff} \rangle$ , where  $E$  is a non-empty, finite set of events (the *domain* of  $\mathcal{E}$ ) and  $Q_i \subseteq E \times E$  is an equivalence relation for each agent  $i \in \mathcal{A}$  (the *indistinguishability relation* of  $i$ ). Instead of a valuation function, we have two functions  $\text{pre} : E \rightarrow \mathcal{L}_{KC}$  and  $\text{eff} : E \rightarrow (P \rightarrow \mathcal{L}_{KC})$ , assigning a *precondition* and conditional *effects* to each event. While the precondition specifies the condition under which an effect is applicable in some world of a model, the conditional effects tell us whether or not an atomic proposition will be true in the successor world.

We depict event models analogously to epistemic models with the difference that nodes now correspond to events, which are additionally labeled with their respective preconditions and effects. Consider the following event model:

$$\mathcal{E}_{\text{sense}} = \begin{array}{ccc} & \bullet & \xrightarrow{2} & \bullet \\ e_1 : \langle p, \{p \mapsto p\} \rangle & & & e_2 : \langle \neg p, \{p \mapsto p\} \rangle \end{array}$$

An event model *updates* an epistemic model by pairing up every world with every applicable event (i.e., of which the precondition is satisfied). Two updated worlds are indistinguishable for an agent if both the original worlds and the events are indistinguishable for that agent. Furthermore, a proposition is true in an updated world if and only if the event’s conditional effect concerning that proposition evaluates to true in the original world.

For example,  $\mathcal{E}_{\text{sense}}$  consists of two events with preconditions  $p$  and  $\neg p$ , respectively. For both events, the effect is  $\{p \mapsto p\}$ , meaning  $p$  will be true if  $p$  was true before (from now on, we will omit these trivial effects that preserve the value of an atomic proposition in our depiction of event models). Since the events are distinguishable for agent 1, the agent will, after the execution of the action, be able to distinguish worlds in which  $p$  is true from worlds in which  $p$  is false.

Formally, we define the *product update*  $\mathcal{M} \otimes \mathcal{E}$  of model  $\mathcal{M} = \langle W, (R_i)_{i \in \mathcal{A}}, V \rangle$  with respect to an event model  $\mathcal{E} = \langle E, (Q_i)_{i \in \mathcal{A}}, \text{pre}, \text{eff} \rangle$  as model  $\langle W', (R'_i)_{i \in \mathcal{A}}, V' \rangle$  where

- $W' = \{(w, e) \in W \times E \mid \mathcal{M}, w \models \text{pre}(e)\}$ ,
- $R'_i = \{((w, e), (w', e')) \in W' \times W' \mid wR_iw', eQ_i e'\}$ ,
- $V'(p) = \{(w, e) \in W' \mid \mathcal{M}, w \models \text{eff}(e, p)\}$ .

In particular, if we apply  $\mathcal{E}_{\text{sense}}$  in  $\mathcal{M}_0$ , we obtain the following epistemic model:

$$\mathcal{M}_0 \otimes \mathcal{E}_{\text{sense}} = \begin{array}{ccc} & \bullet & \xrightarrow{2} & \bullet \\ & (w_1, e_1) : p & & (w_2, e_2) : \end{array}$$

As intended, agent 1 knows now whether or not  $p$  is true. Note that additionally agent 2 is aware of this. The event model  $\mathcal{E}_{\text{sense}}$  represents *semi-private sensing*, meaning that even though the result of the sensing will only be known to agent 1, agent 2 will know that the sensing has taken place.

For planning, we usually consider pointed models  $(\mathcal{M}, w)$ , i.e., where one world  $w$  from the domain of  $\mathcal{M}$  is designated as the actual world. In contrast, we model *epistemic actions* as multi-pointed event models  $(\mathcal{E}, E_d)$  where  $E_d$  is a subset of the domain of  $\mathcal{E}$ . This is necessary, since sometimes we want the events to be deliberately chosen by the acting agents and sometimes by the environment. E.g., our semi-private sensing action should be defined as  $(\mathcal{E}_{sense}, \{e_1, e_2\})$ . Since both events are designated, it can be applied regardless of whether  $p$  is true or false. Applied in  $(\mathcal{M}_0, w_1)$ , the action results in the pointed model  $(\mathcal{M}_0 \otimes \mathcal{E}_{sense}, (w_1, e_1))$  and applied in  $(\mathcal{M}_0, w_2)$  it results in  $(\mathcal{M}_0 \otimes \mathcal{E}_{sense}, (w_2, e_2))$ . The similar action  $(\mathcal{E}_{sense}, \{e_1\})$  is only applicable in the case where  $p$  is true. It can, e.g., be used to model the action of a third agent, who already knows that  $p$  is true, semi-privately informing agent 1 of this.

Formally, an epistemic action  $(\mathcal{E}, E_d)$  is applicable in  $(\mathcal{M}, w)$  if there is an applicable event  $e \in E_d$ , meaning that  $\mathcal{M}, w \models \text{pre}(e)$ . The application of  $(\mathcal{E}, E_d)$  in  $(\mathcal{M}, w)$  then nondeterministically leads to a pointed model  $(\mathcal{M} \otimes \mathcal{E}, (w, e))$  such that  $\mathcal{M}, w \models \text{pre}(e)$ .

Note that any *epistemic state* represented by a pointed model  $(\mathcal{M}, w)$ , has infinitely many epistemically equivalent representations (i.e., other pointed models that satisfy the exact same set of formulas). It is a central theorem of modal logic that finite models are epistemically equivalent if and only if they are bisimilar (Blackburn, de Rijke, and Venema 2001). In the following, when using pointed models as states in a transition system, we think of them as representatives of their whole equivalence class. I.e., we consider two epistemic states  $(\mathcal{M}, w)$  and  $(\mathcal{M}', w')$  as identical if they are epistemically equivalent. We say that two epistemic states  $(\mathcal{M}, w)$  and  $(\mathcal{M}', w')$  are indistinguishable for an agent  $i$  if there is a world  $w''$  in  $\mathcal{M}$  that is indistinguishable to  $w$  for agent  $i$  such that  $(\mathcal{M}, w'')$  and  $(\mathcal{M}', w')$  are identical. An initial epistemic state together with a set of epistemic actions thus induces a nondeterministic transition system where all states are epistemically different from each other.

## FOND Planning

Our definition of FOND planning loosely follows the conventions of Ghallab, Nau, and Traverso (2004). In particular, our actions consist of one common precondition and a set of possible effects, from which one will always be chosen nondeterministically. However, since we want to start out with a formalization that is as close as possible to our DEL formalism, we allow arbitrary propositional formulas as action preconditions and goals. We also use *conditional effects* which we restrict to *effect normal form*, which is a special case of Rintanen’s unary conditionality normal form (Rintanen 2003).

We define a FOND planning task as a tuple  $\langle \mathcal{F}, \mathcal{I}, \gamma, Act \rangle$  where  $\mathcal{F}$  is a finite set of fluents (atomic propositions),  $\mathcal{I} \subseteq \mathcal{F}$  is the initial state,  $\gamma$  is a propositional goal formula over  $\mathcal{F}$  and  $Act$  is a finite set of actions. Each action  $a = \langle \text{pre}_a, \text{effs}_a \rangle \in Act$  consists of a propositional formula  $\text{pre}_a$  over  $\mathcal{F}$  (the precondition) and a set  $\text{effs}_a$  (the conditional effects). Each conditional effect  $e \in \text{effs}_a$  is of the form  $\bigwedge_{f \in \mathcal{F}} (\chi_f^e \triangleright f) \wedge (\chi_{\neg f}^e \triangleright \neg f)$ , where  $\chi_f^e$  and  $\chi_{\neg f}^e$  are

mutually inconsistent propositional formulas over  $\mathcal{F}$  (i.e., their conjunction is unsatisfiable). They can be interpreted as “effect  $e$  makes  $f$  true under the condition  $\chi_f^e$  and false under the condition  $\chi_{\neg f}^e$ ”. Such a FOND task induces a finite transition system starting with the initial state  $\mathcal{I}$  and connecting two states  $\mathcal{S}$  and  $\mathcal{S}'$  via action  $a$  iff  $\mathcal{S} \models \text{pre}_a$  and there is an effect  $e \in \text{effs}_a$  such that the conditional effects in  $e$  transform  $\mathcal{S}$  to  $\mathcal{S}'$ .

This gives us a trivial compilation from FOND to DEL. I.e., we compile the initial state into an epistemic state with exactly one world  $w_0$  where  $V(p) = \{w_0\}$  iff  $p \in \mathcal{I}$ , or  $V(p) = \emptyset$  otherwise. For each action  $a \in Act$ , we construct an epistemic action with one event for each nondeterministic effect  $e \in \text{effs}_a$ , with precondition  $\text{pre}_a$  and effect  $\{f \mapsto \chi_f^e \vee (f \wedge \neg \chi_{\neg f}^e) \mid f \in \mathcal{F}\}$ . All events are designated and pairwise distinguishable for all agents. The transition system that we get from our compiled DEL state and actions is isomorphic to the FOND transition system and identified states share the same propositional valuation.

One solution to FOND planning tasks are *strong policies*. A strong policy  $\pi$  is a partial mapping from states to actions which satisfies the following properties (Cimatti et al. 2003):

- For every state  $\mathcal{S}$  that is reachable via  $\pi$  from  $\mathcal{I}$ , there is some state  $\mathcal{S}'$  that is reachable from  $\mathcal{S}$  via  $\pi$ , s.t.  $\mathcal{S}' \models \gamma$ .
- There are no cycles, i.e. states  $\mathcal{S} \neq \mathcal{S}'$  such that both  $\mathcal{S}'$  is reachable from  $\mathcal{S}$  via  $\pi$  and  $\mathcal{S}$  is reachable from  $\mathcal{S}'$  via  $\pi$ .

Since the transition system is finite, following a strong policy always leads to a goal state in finitely many steps. It seems reasonable to assume that the concept of strong policies is also useful for contingent planning over epistemic states.

## Implicit Coordination

We define an epistemic planning task as a tuple  $\langle s_0, A, \omega, \gamma \rangle$  where  $s_0$  is an epistemic state (the *initial state*),  $A$  is a finite set of epistemic actions (the *action library*),  $\omega : A \rightarrow \mathcal{A}$  is a function mapping each action to its owner (the *owner function*), and  $\gamma \in \mathcal{L}_{KC}$  is the *goal formula*.

For example, consider the epistemic planning task with  $s_0 = (\mathcal{M}_0, w_1)$ ,  $A = \{\text{sense}, \text{ann}_p, \text{ann}_{\neg p}\}$  and with the actions  $\text{sense} = (\mathcal{E}_{sense}, \{e_1, e_2\})$ ,  $\text{ann}_p = (\mathcal{E}_{\text{ann}_p}, e_1)$  and  $\text{ann}_{\neg p} = (\mathcal{E}_{\text{ann}_{\neg p}}, e_1)$ . The actions  $\text{ann}_p$  and  $\text{ann}_{\neg p}$  are public announcement actions for agent 1, announcing that  $p$  is true or false, respectively. That is, the event models  $\mathcal{E}_{\text{ann}_{\neg p}}$  and  $\mathcal{E}_{\text{ann}_p}$  are given as follows:

$$\mathcal{E}_{\text{ann}_p} = e_1 : \langle p, \emptyset \rangle \quad \mathcal{E}_{\text{ann}_{\neg p}} = e_1 : \langle \neg p, \emptyset \rangle$$

We assume that all actions are owned by agent 1, i.e.,  $\omega = \{\text{sense} \mapsto 1, \text{ann}_p \mapsto 1, \text{ann}_{\neg p} \mapsto 1\}$ . The goal is for agent 2 to know whether or not  $p$  is true, i.e.,  $\gamma = K_2 p \vee K_2 \neg p$ .

A strong policy in the sense of Cimatti et al. (2003) would be to just apply the action  $\text{ann}_p$  in  $s_0$ . This is because the action is applicable in  $(\mathcal{M}_0, w_1)$  and its application would lead to a successor state consisting of only one world  $(w_1, e_1)$  in which  $p$  (and  $K_2 p$ ) is true. We argue that from the perspective of the agents (who initially do not know whether  $p$  is

true or false), this is not a reasonable solution. If we want agent 1 to be able to come up with the plan for himself, we must consider his incomplete knowledge about the situation. Intuitively, a good plan for agent 1 is to first apply the sensing action and then, depending on the sensing result, apply the action  $\text{ann}_p$  or  $\text{ann}_{\neg p}$ . This plan works for both states  $(\mathcal{M}, w_1)$  and  $(\mathcal{M}, w_2)$ , which agent 1 considers possible.

To capture this, we require *uniform* policies. A uniform policy is a partial function  $\pi$  from epistemic states to sets of epistemic actions, satisfying the following constraints:

- **Applicability:** for each state  $s$  and action  $a \in \pi(s)$ , the action  $a$  is applicable in state  $s$ .
- **Uniformity:** for each state  $s$ , action  $a \in \pi(s)$ , and states  $s'$  that are indistinguishable to  $s$  for the owner  $\omega(a)$  of the action, also  $a \in \pi(s')$ .

This definition ensures that the agents can always infer from their own knowledge whether or not and how the policy wants them to act. This also implies that an action is only applicable by an agent if the agent *knows* that the action is applicable. Note that because of the uniformity constraint, it is necessary to allow policies to assign multiple actions per state. E.g., sometimes we want a policy to assign an action  $a$  of agent 1 to some state  $s$  and an action  $b$  of agent 2 to some state  $s'$ . Then by uniformity, if there is a state  $s''$  that is indistinguishable to  $s$  for agent 1 and to  $s'$  for agent 2, we have to assign both  $a$  and  $b$  to  $s''$ .

We then say that a uniform policy is *subjectively strong* if it is finite and satisfies the exact properties of strong plans, but based on *subjective reachability*: a state  $s'$  is a subjective successor of  $s$ , given an action  $a$ , if there is a state  $s''$  that is indistinguishable to  $s$  for agent  $\omega(a)$  and  $s'$  is a successor of  $s''$  and  $a$ . In our example, the subjective successors of  $(\mathcal{M}_0, w_1)$  and  $(\mathcal{E}_{\text{sense}}, \{e_1, e_2\})$  are exactly the states  $(\mathcal{M}_0 \otimes \mathcal{E}_{\text{sense}}, (w_1, e_1))$  and  $(\mathcal{M}_0 \otimes \mathcal{E}_{\text{sense}}, (w_2, e_2))$ . A state  $s'$  is then subjectively reachable from  $s$  if either  $s'$  is identical to  $s$  or  $s'$  is subjectively reachable from a subjective successor of  $s$ . In particular, a policy  $\pi$  that is subjectively strong for an epistemic planning task  $\langle s_0, A, \omega, \gamma \rangle$  guarantees for each subjectively reachable state  $s$  and action  $a \in \pi(s)$ , that  $\pi$  is also subjectively strong for  $\langle s, A, \omega, \gamma \rangle$ , as well as for all planning tasks  $\langle s', A, \omega, \gamma \rangle$  with an initial state  $s'$  that is indistinguishable to  $s$  for  $\omega(a)$ . Thus the acting agent always knows that the action is applicable and goal-directed. Subjectively strong policies are our main solution concept for implicit coordination.

## Implicit Coordination in FOND

In the following, we will show how implicit coordination can be achieved in a much simpler, propositional setting. The assumption is that both the epistemic initial state and the epistemic actions of the agents can be represented using propositional planning states and actions.

Since we want to compute subjectively strong policies, we have to consider subjective instead of objective reachability. We do this by introducing additional, nondeterministic *perspective shifting* actions. These are actions that are not ultimately executed by the agents, but are applied during planning before each action to shift to the perspective of

the acting agent. We can then use a variant of strong FOND planning to compute subjectively strong policies.

## Propositional Representation of Epistemic States

We use the approach by Kominis and Geffner (2015) to represent epistemic states as classical states. The idea is that we generate fluents directly from the worlds and indistinguishability relation of an initial state  $s_0$ , such that we can use them to encode the valuation functions and indistinguishability relations of arbitrary states reachable from  $s_0$ .

Given the initial state  $s_0 = (\langle W, (R_i), V \rangle, w_0)$ , for each proposition  $p \in P$  and world  $w \in W$ , we introduce a fluent  $p^w \in \mathcal{F}$  (read: “ $p$  is true in world  $w$ ”). Similarly, for each agent  $i \in \mathcal{A}$  and any two worlds  $w_1, w_2 \in W$  we introduce a fluent  $D_i^{\{w_1, w_2\}} \in \mathcal{F}$  (read: “ $w_1$  is distinguishable to  $w_2$  for agent  $i$ ”). Finally, for each world  $w \in W$  we introduce the fluent  $w^* \in \mathcal{F}$  (read: “ $w$  is the designated world”). This is different to the approach by Kominis and Geffner, where marking a designated world is not necessary. We make use of these fluents in our perspective shifting actions.

A propositional state  $\mathcal{S} \subseteq \mathcal{F}$  then represents an epistemic state  $(\langle W, (R'_i), V' \rangle, w')$  where (1)  $w \in V'(p)$  iff  $p^w \in \mathcal{S}$ , (2)  $w_1 R'_i w_2$  iff  $D_i^{\{w_1, w_2\}} \notin \mathcal{S}$ , and (3)  $w' = w$  iff  $w^* \in \mathcal{S}$ .

For example, if the initial state is  $s_0 = (\mathcal{M}_0, w_1)$ , then  $\mathcal{F} = \{p^{w_1}, p^{w_2}, D_1^{\{w_1, w_2\}}, D_2^{\{w_1, w_2\}}, w_1^*, w_2^*\}$ . The state  $s_0$  is then translated to the FOND state  $\mathcal{I} = \{p^{w_1}, w_1^*\}$ .

## Evaluation of Epistemic Formulas

Checking whether a propositional formula  $\phi$  is true in world  $w$  of an epistemic state that is represented by a classical state  $\mathcal{S} \subseteq \mathcal{F}$  is simple. We replace the occurrences of each proposition  $p$  in  $\phi$  by  $p^w$  and check the resulting formula in  $\mathcal{S}$ .

Checking formulas with knowledge operators is slightly more complicated. Kominis and Geffner (2015) use *axioms* to compile away all knowledge subformulas into *derived variables*, the values of which can be inferred in polynomial time. In this paper, we assume that all of this is given and that we can thus express each epistemic formula  $\phi$  as a formula  $\phi^w$  that evaluates to true in a classical state representing an epistemic state  $(\mathcal{M}, w_0)$  iff  $\mathcal{M}, w \models \phi$ .

For evaluating a formula  $\phi$  (e.g., the goal formula) in the designated world of a state, we use the formula  $\phi^*$ , which we define as  $\bigwedge_{w \in W} (w^* \rightarrow \phi^w)$ .

## Epistemic Actions and Perspective Shifts

Using our propositional representation of epistemic states, we can now model certain epistemic actions as classical planning actions. E.g., we can represent an agent  $i$  sensing a formula  $\phi$  using one conditional effect for each pair  $w, w'$  of worlds. If  $\phi$  evaluates differently in the two worlds, they will be made distinguishable, i.e.  $((\phi^w \wedge \neg \phi^{w'}) \vee (\neg \phi^w \wedge \phi^{w'})) \triangleright D_i^{\{w, w'\}}$ . In general, it is possible to compile arbitrary DEL actions with mutually inconsistent preconditions to classical planning actions in a similar way. This is because such actions never increase the uncertainties of the agents, and only the valuations and indistinguishabilities of the worlds have to be updated. Also, such actions are always deterministic.

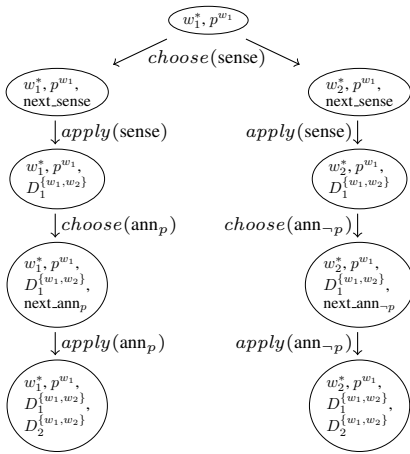


Figure 1: FOND version of a subjectively strong policy.

All actions that can be expressed by Kominis and Geffner’s action language fall into this category.

For now, we assume that such a translation is given and that we can compile each action from the action library of an epistemic planning task into a semantically equivalent FOND action. We later generalize this by characterizing a larger fragment of DEL which also supports some limited kind of nondeterminism and by defining a systematic translation of actions from that fragment to FOND actions.

To use strong FOND planning to compute subjectively strong policies, we simulate subjective instead of objective transitions. We do this by further splitting up each FOND action  $a$  into two: One auxiliary action  $choose(a)$  for *choosing* the action that we want to apply in a state and one action  $apply(a)$  that actually *applies* the effects (and has the preconditions) of the previously selected action. In each choice action, we additionally simulate a *perspective shift*: we change the designated world nondeterministically to any of the worlds that are indistinguishable for the owner agent of the action. This can be achieved using conditional nondeterministic effects. The action  $apply(a)$  then must be applied in all successor states. Thus, subjective successors of an action in the original problem become objective successors of the successive application of both new actions.

Figure 1 shows a strong policy for the FOND version of our example planning task. As we can see, the policy branches after the first agent chooses its sensing action. This is because the actual world is  $w_1$  and for agent 1,  $w_1$  is indistinguishable from  $w_2$  in  $\{w_1^*, p^{w_1}\}$ . After the sensing,  $w_1$  and  $w_2$  become distinguishable, so when agent 1 chooses to announce  $p$  or  $\neg p$ , there is no branching.

### Subjectively Strong Policies in FOND

To show that the approach is sound and complete, we have to show that for every subjectively strong policy in the epistemic planning task, there is a strong plan for the FOND version of the task with auxiliary actions, and vice versa.

We first show that for each subjectively strong policy for an epistemic planning task, there is a strong policy for its

FOND version. This is because each transition for an action  $a$  in the original task corresponds to transitions via the actions  $choose(a)$  and  $apply(a)$  in the FOND version. A strong FOND policy can be constructed iteratively by rolling out the policy starting from the initial state, using the actions that are assigned by the epistemic policy to the corresponding epistemic states. Note that it is possible that the epistemic policy assigns multiple actions to the same state. Thus we also have to allow this for the FOND policy.

Given a strong policy in the FOND version, generating a subjectively strong policy for the epistemic planning task is not as straight-forward and works only under certain conditions. This is because FOND policies can contain multiple propositional states representing epistemically equivalent states (e.g., consider the states  $\{p^{w_1}, w_1^*\}$  and  $\{p^{w_2}, w_2^*\}$ ). If we roll out the epistemic policy from the initial state using the actions from the FOND policy, the resulting epistemic policy can become cyclic. In fact, there are epistemic planning tasks for which there are no subjectively strong policies, but for which there are strong policies in the FOND version because epistemic duplicates are not identified.

We deal with this issue by considering a *modified version* of FOND planning, where classical states are identified if and only if the epistemic states they represent are equivalent. We thus get a direct correspondence between the induced transition systems, and policies can be translated in both directions simply by replacing epistemic states by their FOND representations and vice versa.

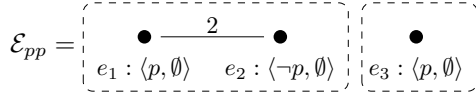
For this variant of FOND, existing planning algorithms and implementations have to be modified only slightly, replacing duplicate checking by checks based on bisimilarity. This can be done efficiently and in polynomial time (Dovier, Piazza, and Policriti 2004). Thus the computational complexity remains the same as for regular FOND planning. We believe that in practice, it would often be clear from the description of the actions that no epistemic duplicates can be created. In these cases, we can also use regular FOND planning. This is the case for our case study later in which we model multi-agent path finding with destination uncertainty.

### The FOND Fragment of DEL

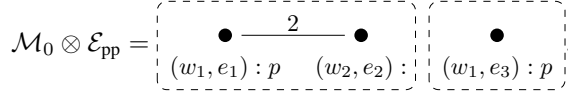
We have not yet defined a formal compilation from DEL to FOND actions. And we have only looked at examples with purely deterministic actions. This setting is sufficient for tasks with initial state uncertainty like multi-agent path finding with destination uncertainty. Even in Hanabi, the whole uncertainty of the agents can be regarded as initial state uncertainty about the order of cards on the stack. However, the number of possible deals in Hanabi is much too big for these epistemic states to be represented explicitly. A better solution is to instead use nondeterminism to model drawing cards from the stack.

Therefore, we now want to define a fragment of DEL planning that combines the best of both worlds: actions can be nondeterministic but there may be no outcomes where the total uncertainty (i.e., the number of worlds) increases. We achieve this by requiring our event models to be partitioned into disjoint connected components with mutually inconsistent preconditions. This allows us to think of each

of the components as a single nondeterministic effect. Consider the following event model:



For example, the action  $(\mathcal{E}_{pp}, \{e_1, e_3\})$  could model an agent 3 trying to semi-privately announce  $p$  to agent 1 (e.g., by whispering). However, there is the possibility that the confidentiality of the announcement is compromised and  $p$  is thus effectively publicly announced. If we apply this action in  $(\mathcal{M}_0, w_1)$ , it results either in  $(\mathcal{M}_0 \otimes \mathcal{E}_{pp}, (w_1, e_1))$  or  $(\mathcal{M}_0 \otimes \mathcal{E}_{pp}, (w_1, e_3))$  where



Formally, for actions  $(\langle E, (Q_i), \text{pre}, \text{eff} \rangle, E_d)$  from our action set, we require that the domain  $E$  can be partitioned into disjoint subsets  $E_1, \dots, E_k$  such that (1) for each pair of events  $e, e' \in E_j$  from the same component  $j \in \{1, \dots, k\}$  with  $e \neq e'$ , the preconditions  $\text{pre}(e)$  and  $\text{pre}(e')$  are mutually inconsistent, and (2) two events  $e, e' \in E$  are only allowed to be indistinguishable for an arbitrary agent  $i \in \mathcal{A}$ , i.e.  $e Q_i e'$ , if they belong to the same component, i.e. if there exists a  $j \in \{1, \dots, k\}$  such that  $e, e' \in E_j$ .

We can see that if we apply such an action to an arbitrary epistemic state, due to condition (1), each world will be paired with at most one of the events of each component. Furthermore, due to condition (2), two worlds are only distinguishable for an agent if the events they were generated with are from the same component. Thus, each state resulting from an action application will consist of at most  $k$  connected components, which consist each of at most as many worlds as the original state. Since we can throw away all components that do not contain the designated world, we obtain a state that can be represented by at most as many worlds as the original state. We call this fragment the *FOND fragment of DEL* and the special case where  $k = 1$  for all actions the *deterministic fragment of DEL*.

In Hanabi, agents can both gain new knowledge (e.g., by receiving hints) and new uncertainty (by drawing a new card from the stack). While it is difficult to model such actions, it still can be done using modeling tricks. These are based on generating uncertainty at the same time as new information is gained. This works because the overall uncertainty never becomes greater than the uncertainty in the initial state.

### Compilation of Epistemic Actions

Based on the compilation of epistemic states and formulas from before, we now show how an action  $a = (\langle E, (Q_i), \text{pre}, \text{eff} \rangle, E_d)$  that can be partitioned into distinct components  $E_1, \dots, E_k$  accordingly to our fragment, can be compiled into a FOND action  $a^{\text{FOND}} = \langle \text{pre}_a, \text{eff}_a \rangle$ .

We know that an action is applicable in a state  $(\mathcal{M}, w)$  if there is some event  $e \in E_d$  such that  $\mathcal{M}, w \models \text{pre}(e)$ . We directly translate this to  $\text{pre}_a = \bigvee_{e \in E_d} \text{pre}(e)^*$ . We can then translate each component  $E_j$  of our event model

into a nondeterministic effect  $\text{eff}_{a,j}$ , i.e., we get  $\text{eff}_a = \{\text{eff}_{a,j} \mid j = 1, \dots, k\}$ . These nondeterministic effects can make propositions true or false, as well as make worlds distinguishable or completely inaccessible. We construct each nondeterministic effect  $\text{eff}_{a,j}$  as follows:

$$\text{eff}_{a,j} = \text{eff}_{a,j}^{P+} \wedge \text{eff}_{a,j}^{P-} \wedge \text{eff}_{a,j}^{D+} \wedge \text{eff}_{a,j}^{\times}$$

First, each fluent  $p^w$  is made true or false according to the effects of the event  $e \in E_j$  that is applied in  $w$ :

$$\text{eff}_{a,j}^{P+} = \bigwedge_{p \in P} \bigwedge_{w \in W} \left( \bigvee_{e \in E_j} (\text{pre}(e)^w \wedge \text{eff}(e, p)^w) \triangleright p^w \right)$$

$$\text{eff}_{a,j}^{P-} = \bigwedge_{p \in P} \bigwedge_{w \in W} \left( \bigvee_{e \in E_j} (\text{pre}(e)^w \wedge \neg \text{eff}(e, p)^w) \triangleright \neg p^w \right)$$

Two worlds  $w$  and  $w'$  become distinguishable if the events  $e$  and  $e'$  they were updated with are distinguishable:

$$\text{eff}_{a,j}^{D+} = \bigwedge_{\substack{w, w' \in W \\ i \in \mathcal{A}, w \neq w'}} \left( \bigvee_{\substack{e, e' \in E_j \\ \neg e Q_i e'}} (\text{pre}(e)^w \wedge \text{pre}(e')^{w'}) \triangleright D_i^{\{w, w'\}} \right)$$

If in some world  $w$ , none of the events from  $E_j$  are applicable, the world should not have a successor. We simulate this by making  $w$  distinguishable from all other worlds.

$$\text{eff}_{a,j}^{\times} = \bigwedge_{i \in \mathcal{A}, w \neq w'} \left( \bigwedge_{e \in E_j} \neg \text{pre}(e)^w \right) \triangleright D_i^{\{w, w'\}}$$

### Compilation of Epistemic Planning Tasks

We define the FOND compilation of an epistemic planning task  $\langle s_0, A, \omega, \gamma \rangle$  as  $\langle \mathcal{F}, \mathcal{I}, \gamma^*, \text{Act} \rangle$  where  $\mathcal{F}$  and  $\mathcal{I}$  are derived from  $s_0$  as defined above and  $\text{Act} = \{a^{\text{FOND}} \mid a \in A\}$ . Since by construction, an action  $a^{\text{FOND}} \in \text{Act}$  changes a FOND state  $\mathcal{S}$  analogously to how the DEL action  $a$  changes the DEL counterpart  $s$  of  $\mathcal{S}$ , it is clear that the successor states of  $\mathcal{S}$  and  $s$  will be again counterparts. Thus a strong policy for an epistemic planning task corresponds to a strong policy in its FOND compilation and vice versa. The policy for an epistemic planning task can be extracted from the FOND policy as described in the section about implicit coordination in FOND. We thus obtain the following result:

**Proposition 1.** *For any epistemic planning task in the FOND fragment of DEL, there is a strong policy if and only if there is a strong policy for its FOND compilation.*

The following proposition follows, given the EXPTIME-completeness of the plan existence problem for strong planning in FOND (Rintanen 2004) and given the fact that both compilations from DEL to FOND and from FOND to DEL (as defined in the background section) are polynomial:

**Proposition 2.** *In the FOND fragment of DEL, the problem of deciding whether there exists a strong policy for a given planning task is EXPTIME-complete.*

Since we can find subjectively strong policies by introducing perspective shifting actions, and with only a polynomial blowup, we finally obtain the following result:

**Proposition 3.** *In the FOND fragment of DEL, the problem of deciding whether there exists a subjectively strong policy for a given planning task is EXPTIME-complete.*

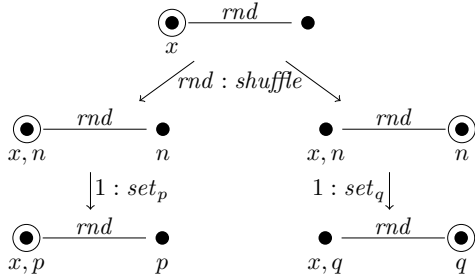


Figure 2: Using subjective reachability to simulate a nondeterministic action in the deterministic fragment of DEL.

### Expressiveness of the FOND Fragment

We now compare the FOND fragment of DEL to the deterministic fragment of DEL. In particular, we want to know whether allowing multiple nondeterministic components makes the FOND fragment more expressive than the deterministic fragment or whether it is just syntactic sugar. We distinguish between the problems of finding strong and subjectively strong policies, starting with the former.

In the deterministic fragment of DEL, each action  $a$  consists of only one connected component. Thus each action  $a^{\text{FOND}}$  consists of only one nondeterministic effect, making it a deterministic action. Thus it is possible to check for the existence of a strong policy by classical planning (for which plan existence is PSPACE-complete). Since we can also express arbitrary classical planning tasks in the deterministic fragment of DEL, we obtain the following result:

**Proposition 4.** *In the deterministic fragment of DEL, the problem of deciding whether there exists a strong policy (or a sequential plan) for a planning task is PSPACE-complete.*

For subjectively strong policies this argument does not hold. As it turns out, we can use subjective reachability to model (objective) nondeterminism. Consider the example depicted in Figure 2. We simulate the nondeterministic action  $\langle \top, \{p, q\} \rangle$  using two agents and the deterministic DEL actions  $shuffle$ ,  $set_p$  and  $set_q$ , which are defined below:

$$\begin{aligned}
 shuffle &= \begin{array}{c} \bullet \\ \text{---} \\ \langle \neg n, \{n \mapsto \top\} \rangle \end{array} \\
 set_p &= \begin{array}{c} \bullet \text{---} rnd \text{---} \bullet \\ \langle n \wedge x, \{p \mapsto \top, n \mapsto \perp\} \rangle \langle n \wedge \neg x, \{p \mapsto \top, n \mapsto \perp\} \rangle \end{array} \\
 set_q &= \begin{array}{c} \bullet \text{---} rnd \text{---} \bullet \\ \langle n \wedge x, \{q \mapsto \top, n \mapsto \perp\} \rangle \langle n \wedge \neg x, \{q \mapsto \top, n \mapsto \perp\} \rangle \end{array}
 \end{aligned}$$

We use the uncertainty of the agent  $rnd$ , who performs the action  $shuffle$ , to generate both a subjective successor where  $set_p$  and one where  $set_q$  has to be applied next by agent 1 (which we enforce using the auxiliary propositions  $n$  and  $x$ ), making either  $p$  or  $q$  globally true. This can be generalized to actions with arbitrary many nondeterministic effects by using a number of additional worlds, events and auxiliary variables linear in the number of desired nondeterministic effects. Since any FOND task can be polynomially

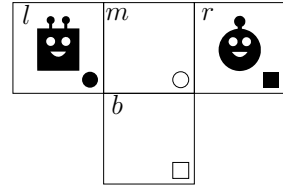


Figure 3: A MAPF/DU instance.

compiled to DEL and determinized in this way, we obtain the following complexity result:

**Proposition 5.** *In the deterministic fragment of DEL, the problem of deciding whether there exists a subjectively strong policy for a planning task is EXPTIME-complete.*

### Case Study: Modeling MAPF/DU in PDDL

We demonstrate our compilation approach by modeling instances of multi-agent path finding with destination uncertainty. Figure 3 shows an example of a MAPF/DU instance with two agents. The potential destinations are  $r$  and  $b$  for the square agent, and  $l$  and  $m$  for the circle agent. Neither agent knows the other agent's actual destination, and they block each others' path. One subjectively strong policy for both agents to get to their true destination is for the square agent to first go to  $b$  and to let the circle agent move to  $l$ , independently of the actual destinations of the agents. From there, the square agent can go to its true destination (which, depending on the designated world, will be either  $r$  or  $b$ ) and announce success. Afterwards, the circle agent can go to his true destination (which will be either  $l$  or  $m$ ). Note that after the initial movements of the square agent, the policy has to consider all four possible destination combinations. This is because the square agent does not know the actual destination of the circle agent and the circle agent will not know the actual destination of the square agent.

We model this task directly in PDDL (McDermott 1998). This has the advantage that we do not need to model single DEL event models and then compile them one by one. We will use the types  $agt$  for agents,  $pos$  for positions, and  $wld$  for worlds. We introduce fluents  $(at \ ?a \ ?p)$  to denote that agent  $?a$  is at position  $?p$ ,  $(adj \ ?p \ ?q)$  to denote that an agent can step from position  $?p$  to position  $?q$ , and  $(announced \ ?a)$  to denote that the agent  $?a$  has already announced success and will no longer act. We use  $(dest \ ?w \ ?a \ ?p)$  to denote that the actual destination of agent  $?a$  in world  $?w$  is position  $?p$ . Note that we make use of the fact that there is only uncertainty about the destinations but not about the positions of the agents. Since we will use one world for each combination of destinations and destinations never change as the result of an action application, there will never be states that are propositionally different but epistemically equivalent.

To denote indistinguishability of two worlds  $?w1$  and  $?w2$  for agent  $?a$ , we use the fluent  $(ind \ ?a \ ?w1 \ ?w2)$ . We mark the designated world  $?w$  using the fluent  $(dsg \ ?w)$ . Additionally, we use auxiliary predicates  $(next-choose)$ ,  $(next-move \ ?a \ ?p1 \ ?p2)$  and

(next-announce ?a) to enforce the alternation of auxiliary perspective-shifting actions and actual actions.

We now show how to split the movement actions into the actions choose-move and apply-move. The action (choose-move ?a ?w ?p ?q) simulates a perspective shift to agent ?a by nondeterministically switching to an arbitrary world that is indistinguishable from the designated world for agent ?a. By setting the fluent (next-move ?a ?p ?q) to true, it then enforces a movement action for agent ?a from ?p to ?q in the successor state.

```
(:action choose-move
:parameters (?a - agt ?w - wld
             ?p ?q - pos)
:precondition (and (dsg ?w) (next-choose))
:effect (and
  (not (next-choose))
  (next-move ?a ?p ?q)
  (oneof ; simulate perspective shift
    (when (and (ind ?a ?w w1) (not (= ?w w1))
              (and (not (dsg ?w)) (dsg w1))))
    (when (and (ind ?a ?w w2) (not (= ?w w2))
              (and (not (dsg ?w)) (dsg w2))))
    ...)))
```

Unfortunately, we have to enumerate all possible worlds to simulate the perspective shift. This forces us to include the worlds as constants into the domain definition. It would be more convenient if there was a parametrized version of oneof to automatically generate nondeterministic effects by explicitly quantifying over objects (in our case, worlds).

The apply-move action, which must be applied afterwards, performs the actual change of the agent’s position. This action also contains the actual precondition for movements: the destination vertex must be adjacent and empty. Also, it prescribes the next action to again be a choose action by setting the fluent next-choose to true.

```
(:action apply-move
:parameters (?a - agt ?w - wld
             ?p ?q - pos)
:precondition (and
  (dsg ?w) (next-move ?a ?p ?q)
  (at ?a ?p) (adj ?p ?q)
  (not (announced ?a))
  (not (exists (?a2 - agt) (at ?a2 ?q))))
:effect (and
  (not (at ?a ?p)) (at ?a ?q)
  (not (next-move ?a ?p ?q))
  (next-choose)))
```

Checking the original action’s precondition in the apply component of the action is the systematically correct version to split up actions into a choice and an application component. However, in this particular case we can move most of the preconditions from the apply action into the choose action since they are not dependent on the designated world ?w. This optimization speeds up the search significantly.

The announcing actions choose-ann and apply-ann can be defined similarly. Announcing works by making all worlds where the agent has a different destination than its current position distinguishable from any other world for all agents. Our example instance from Figure 3 can then be defined using the following initial state and goal descriptions:

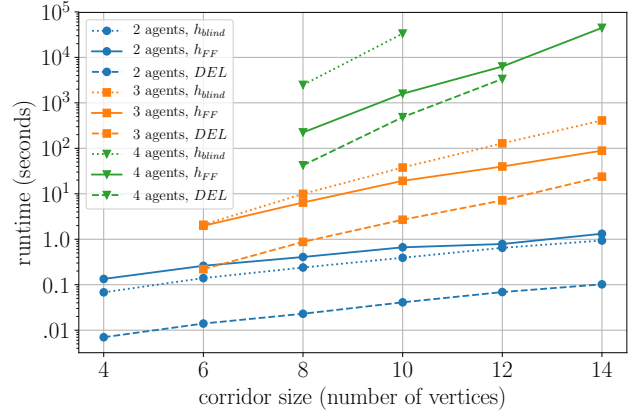


Figure 4: Running time for different MAPF/DU instances.

```
(:objects a1 a2 - agt l m r b - pos)
(:init (adj l m) (adj m l) (adj m r) ...
  (ind a1 w1 w2) (ind a1 w2 w1) ...
  (ind a2 w1 w3) (ind a2 w3 w1) ...
  (dest w1 a1 r) (dest w1 a2 l)
  ...; destinations for w2, w3, w4
  (dsg w1) (next-choose))
(:goal (and
  (imply (dsg w1) (and (at a1 r)
                       (at a2 l)))
  (imply (dsg w2) ...) ...))
```

We evaluated our MAPF/DU planning domain using the myND planner of Mattmüller et al. (2010), which is, to the best of our knowledge, the only publicly available FOND planner that supports both strong (acyclic) plans as well as conditional effects. It also supports axioms, although we did not need them for our example. Figure 4 shows the running time of the planner on the example instance from Figure 3 as well as on versions with additional agents and vertices in the graph, comparing it to the DEL-based planner of Engesser et al. (2017)<sup>1</sup>. The time limit for each instance was one day.

We can see that both the DEL-based and the compilation-based approach work quite well for small instances with only two agents but take substantially longer for three or four agents. This is not surprising. Nebel et al. (2019) showed that the plan existence problem for MAPF/DU is NP-complete and that the plan existence problem with a given bound for the policy depth is PSPACE-complete. The naive algorithm they proposed for finding optimal plans has a runtime complexity of  $O(n^{a^2+a})$  where  $n$  is the number of graph vertices and  $a$  is the number of agents. We can also see that while the FOND planner with a blind heuristic performs systematically worse than the DEL-based approach, using the FF heuristic (Hoffmann and Nebel 2001) significantly speeds up the search on larger instances. Our largest instance was not solved by the DEL-based planner since it runs out of memory (32GB), whereas it was still solved by myND using the FF heuristic. The possibility of using avail-

<sup>1</sup>The PDDL files and experiment logs can be downloaded at <https://gkigit.informatik.uni-freiburg.de/tengesser/mapfdu-fond>.



able heuristic and optimization techniques is one of the benefits of using existing FOND planners.

## Conclusion

We have demonstrated how FOND planning can be used to generate subjectively strong policies. We have also introduced a decidable fragment of DEL where finding strong and subjectively strong policies has the same complexity as strong planning in FOND. These are the first known complexity results for finding subjectively strong policies in a general setting, besides the undecidability of the unrestricted problem. For future work it is worth noticing that DEL is well-suited for modeling games. In particular, there is a translation from a large fragment of the game description language GDL-III to DEL (Engesser et al. 2018). We would like to generalize the solution concept of subjectively strong policies to optimization problems and to scenarios that are at least partially competitive. It will be interesting to investigate how the idea of simulating perspective taking via nondeterminism can be used with techniques such as Monte Carlo tree search or model-based reinforcement learning (e.g., value iteration in fully observable MDPs).

## Acknowledgments

This research was supported under the Australia-Germany Joint Research Cooperation Scheme 2017/18, project 57319997. We thank Robert Mattmüller and the anonymous reviewers for their helpful and insightful comments.

## References

- Aucher, G., and Bolander, T. 2013. Undecidability in epistemic planning. In *Proceedings of IJCAI 2013*, 27–33.
- Baral, C.; Bolander, T.; van Ditmarsch, H.; and McIlraith, S. A. 2017. Epistemic planning (Dagstuhl seminar 17231). *Dagstuhl Reports* 7(6):1–47.
- Bard, N.; Foerster, J. N.; Chandar, S.; Burch, N.; Lanctot, M.; Song, H. F.; Parisotto, E.; Dumoulin, V.; Moitra, S.; Hughes, E.; Dunning, I.; Mourad, S.; Larochelle, H.; Bellemare, M. G.; and Bowling, M. 2019. The hanabi challenge: A new frontier for AI research. *CoRR* abs/1902.00506.
- Blackburn, P.; de Rijke, M.; and Venema, Y. 2001. *Modal Logic*, volume 53 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press.
- Bolander, T., and Andersen, M. B. 2011. Epistemic planning for single and multi-agent systems. *Journal of Applied Non-Classical Logics* 21(1):9–34.
- Bolander, T.; Engesser, T.; Mattmüller, R.; and Nebel, B. 2018. Better eager than lazy? How agent types impact the successfulness of implicit coordination. In *Proceedings of KR 2018*, 445–453.
- Bolander, T.; Jensen, M. H.; and Schwarzenrüber, F. 2015. Complexity results in epistemic planning. In *Proceedings of IJCAI 2015*, 2791–2797.
- Charrier, T.; Maubert, B.; and Schwarzenrüber, F. 2016. On the impact of modal depth in epistemic planning. In *Proceedings of IJCAI 2016*, 1030–1036.
- Cimatti, A.; Pistore, M.; Roveri, M.; and Traverso, P. 2003. Weak, strong, and strong cyclic planning via symbolic model checking. *Artificial Intelligence* 147(1–2):35–84.
- Dovier, A.; Piazza, C.; and Policriti, A. 2004. An efficient algorithm for computing bisimulation equivalence. *Theor. Comput. Sci.* 311(1–3):221–256.
- Engesser, T.; Bolander, T.; Mattmüller, R.; and Nebel, B. 2017. Cooperative epistemic multi-agent planning for implicit coordination. In *Proceedings of M4M 2017*, 75–90.
- Engesser, T.; Mattmüller, R.; Nebel, B.; and Thielscher, M. 2018. Game description language and dynamic epistemic logic compared. In *Proceedings of IJCAI 2018*, 1795–1802.
- Ghallab, M.; Nau, D. S.; and Traverso, P. 2004. *Automated planning - theory and practice*. Elsevier.
- Hoffmann, J., and Nebel, B. 2001. The FF planning system: Fast plan generation through heuristic search. *J. Artif. Intell. Res.* 14:253–302.
- Kominis, F., and Geffner, H. 2015. Beliefs in multiagent planning: From one agent to many. In *Proceedings of ICAPS 2015*, 147–155.
- Mattmüller, R.; Ortlieb, M.; Helmert, M.; and Bercher, P. 2010. Pattern database heuristics for fully observable non-deterministic planning. In *Proceedings of ICAPS 2010*, 105–112.
- Maubert, B.; Pinchinat, S.; and Schwarzenrüber, F. 2019. Reachability games in dynamic epistemic logic. In *Proceedings of IJCAI 2019*, 499–505.
- McDermott, D. 1998. PDDL – The Planning Domain Definition Language, Version 1.2. Technical Report CVC TR-98-003/DCS TR-1165, Yale Center for Computational Vision and Control.
- Muise, C.; Belle, V.; Felli, P.; McIlraith, S.; Miller, T.; Pearce, A. R.; and Sonenberg, L. 2015. Planning over multi-agent epistemic states: A classical planning approach. In *Proceedings of AAI 2015*, 3327–3334.
- Nebel, B.; Bolander, T.; Engesser, T.; and Mattmüller, R. 2019. Implicitly coordinated multi-agent path finding under destination uncertainty: Success guarantees and computational complexity. *J. Artif. Intell. Res.* 64:497–527.
- Reifsteck, D.; Engesser, T.; Mattmüller, R.; and Nebel, B. 2019. Epistemic multi-agent planning using monte-carlo tree search. In *Proceedings of KI 2019*, 277–289.
- Rintanen, J. 2003. Expressive equivalence of formalisms for planning with sensing. In *Proceedings of ICAPS 2003*, 185–194.
- Rintanen, J. 2004. Complexity of planning with partial observability. In *Proceedings of ICAPS 2004*, 345–354.
- van Benthem, J.; van Eijck, J.; and Kooi, B. 2006. Logics of communication and change. *Information and Computation* 204(11):1620–1662.
- van Ditmarsch, H. P.; van der Hoek, W.; and Kooi, B. 2007. *Dynamic Epistemic Logic*. Springer Heidelberg.