Planning with Epistemic Preferences

Toryn Q. Klassen^{1,2,3} , **Christian Muise**^{2,4} , **Sheila A. McIlraith**^{1,2,3}
^{1}Department of Computer Science, University of Toronto, Toronto, Canada
^{2}Vector Institute for Artificial Intelligence, Toronto, Canada
^{3}Schwartz Reisman Institute for Technology and Society, Toronto, Canada
^{4}School of Computing, Queen's University, Kingston, Canada
toryn@cs.toronto.edu, christian.muise@queensu.ca, sheila@cs.toronto.edu

Abstract

Within the field of automated planning, two areas of study are planning with preferences and epistemic planning. Planning with preferences involves generating plans that optimize for properties of the plan instead of, or in addition to, trying to reach a fixed goal. Epistemic planning allows for planning over the knowledge or belief states of one or more agents for the purpose of achieving epistemic goals (where agents have particular states of knowledge or belief). In this paper we motivate and explore the task of planning with epistemic preferences, proposing a method by which existing automated planning techniques can be combined for this purpose.

1 Introduction

Given a description of the current state of the world and some desired goal, a (symbolic) automated planning system generates a plan, typically a sequence of actions that, when executed beginning in the initial state, will achieve the goal. In many cases, we may wish to reason about the beliefs of agents and to devise plans to change their beliefs in some way. Such planning is referred to as *epistemic planning* and has been the subject of a number of recent computational advances that have seen the development of epistemic planners that can generate plans to achieve epistemic goals as well as goals in the physical world (so-called *ontic* goals) (e.g., (Bolander and Andersen 2011; Baier, Mombourquette, and McIlraith 2014; Kominis and Geffner 2015; Engesser et al. 2017; Baral et al. 2017; Le et al. 2018; Fabiano et al. 2020; Cooper et al. 2021; Fabiano et al. 2021; Wan, Fang, and Liu 2021; Muise et al. 2022; Hu, Miller, and Lipovetzky 2022)).

Our interest here is in planning with epistemic *preferences*: beliefs of agents that are desirable to achieve but not mandatory. For example, an agent might want to move an object from the living room to the kitchen and prefer that other agents in the environment *know* the new location, or prefer that Bob knows the new location but not Alice. Alternatively, an agent may have the goal of changing the password on their bank account but prefer that other agents *not know* the new password, or not know that it was changed.

The examples above are domain-specific illustrations of epistemic preferences, but sometimes it may be useful to generate more generic epistemic preferences. Consider the AI safety problem described by Amodei et al. (2016) of negative *side effects* – undesirable changes that an AI system might make (e.g., a robot breaking a vase in its path) because its explicitly given objective did not preclude them. While Amodei et al. were considering the context of reinforcement learning, it is also the case that when planning without preferences, any plan that achieves the explicit goal may seem equally good to the planner (modulo action costs, if any). Klassen et al. (2022) proposed an automated planning approach to avoid some types of side effects, which was essentially this: create, for each possibly negative side effect, a preference for its negation. This establishes a connection between side effects and preferences. The side effects considered there were physical, but (in the context of reinforcement learning) Klassen, Alamdari, and McIlraith (2022; 2023) suggested that some negative side effects could be epistemic, such as when a household robot rearranges items in kitchen cupboards, causing a human to not *know* where things are. In this paper we identify a number of generic sorts of preferences, such as to maximize true beliefs among agents, some of which may have safety applications.

While planning with ontic preferences has received much study (e.g., (Baier and McIlraith 2008)), to the best of our knowledge, this is the first work to study and realize a planner that plans with epistemic preferences. The contributions of this paper include proving the correctness of an encoding of planning with epistemic preferences as a traditional (non-epistemic) planning problem, and demonstrating the feasibility of this approach through experimentation.

2 Background

We will review epistemic logic (see, e.g., (Fagin et al. 1995)) and planning, leading up to RP-MEP (restricted perspectival multi-agent epistemic planning) (Muise et al. 2022), the epistemic planning formalism that we use in this paper.

2.1 Epistemic Logic

Given a finite set of propositional symbols $\mathcal{P}$ and a finite set of agents Ag, the language of epistemic logic is given by

$$\psi ::= p \mid \neg\psi \mid \psi \wedge \psi \mid \Box_i \psi \mid \top \mid \bot$$

where $p \in \mathcal{P}$ and $i \in Ag$. We read $\Box_i \psi$ as saying that agent i believes ψ. We use the abbreviation $\Diamond_i \psi \overset{\text{def}}{=} \neg \Box_i \neg\psi$.

Other operators like implication ($\psi_1 \supset \psi_2 \overset{\text{def}}{=} \neg(\psi_1 \wedge \neg\psi_2)$) can also be defined as abbreviations in the usual way. A *literal* is either a proposition p or its negation, and $\bar{\ell}$ is the complement of ℓ (i.e., $\bar{\ell} = \neg p$ if $\ell = p$ and vice versa).

The semantics are given with a model $\mathfrak{M} = \langle W, \tau, \{R_i : i \in Ag\}\rangle$ where W is a set of worlds, $\tau : W \to 2^{\mathcal{P}}$ identifies which propositions are true at each world, and each $R_i \subseteq W \times W$ is a relation indicating what worlds agent i considers possible. The truth of a formula is defined relative to a model $\mathfrak{M}$ and a world $w \in W$:

- $\mathfrak{M}, w \models p$ iff $p \in \tau(w)$
- $\mathfrak{M}, w \models \neg\psi$ iff $\mathfrak{M}, w \not\models \psi$
- $\mathfrak{M}, w \models \psi_1 \wedge \psi_2$ iff $\mathfrak{M}, w \models \psi_1$ and $\mathfrak{M}, w \models \psi_2$
- $\mathfrak{M}, w \models \Box_i\, \psi$ iff $\mathfrak{M}, w' \models \psi$ for all w' s.t. $R_i(w, w')$
- $\mathfrak{M}, w \models \top$ and $\mathfrak{M}, w \not\models \bot$

Entailment can be defined in the usual way: $\psi_1 \models \psi_2$ if whenever $\mathfrak{M}, w \models \psi_1$, it is also the case that $\mathfrak{M}, w \models \psi_2$.

It is common to restrict the set of models to only those which obey particular axioms, like the KD45$_n$ axioms, which RP-MEP uses, and which can be written as follows (for each agent i):

K: $(\Box_i\, \psi_1 \wedge \Box_i(\psi_1 \supset \psi_2)) \supset \Box_i\, \psi_2$
D: $\Box_i\, \psi \supset \Diamond_i\, \psi$
4: $\Box_i\, \psi \supset \Box_i \Box_i\, \psi$ 5: $\Diamond_i\, \psi \supset \Box_i \Diamond_i\, \psi$

Proper Epistemic Knowledge Bases As we will see later, RP-MEP represents planning states with proper epistemic knowledge bases (PEKBs), a concept originally introduced by Lakemeyer and Lespérance (2012). A *proper epistemic knowledge base (PEKB)* is a set (or conjunction) of *restricted modal literals (RMLs)*, which are formulas of epistemic logic given by this grammar:

$$\varphi ::= p \mid \neg p \mid \Box_i\, \varphi \mid \Diamond_i\, \varphi \mid \top \mid \bot$$

The absence of disjunction simplifies some computations. The *depth* of an RML φ is the number of belief operators in it; e.g., p has depth 0 and $\Box_i \Diamond_j \Box_i \neg p$ has depth 3.

2.2 Classical Planning

We now review non-epistemic automated planning.

Definition 1. A classical$^+$ planning problem[1] is a tuple $\langle F, I, G, O\rangle$ where

- F is a set of *fluents*,
- $I \subseteq F$ is the initial state and $G \subseteq F$ is the goal,
- and O is the set of operators, where an operator $o \in O$ is a tuple $\langle Pre_o, \mathit{Eff}_o^+, \mathit{Eff}_o^-\rangle$ where
  - $Pre_o \subseteq F$ is the *precondition* of o,
  - and each of Eff_o^+ and Eff_o^- are sets of conditional effects of the form $\langle C, f\rangle$ where $C \in (2^F)^2$ and $f \in F$.

As we saw with the initial state, a state s is represented by a subset of F (the fluents in s are understood to be true, and those not are false). An operator o is applicable in a state s if its precondition is true there, i.e. if $Pre_o \subseteq s$. Eff_o^+ is the set of positive conditional effects, that can make fluents true, while Eff_o^- is the set of negative conditional effects, that can make fluents false. A condition $C = \langle C^+, C^-\rangle$ of a conditional effect *fires* in a state s if all the fluents in C^+ are true in s and all those in C^- are false in s. Applying an applicable operator o to a state s transforms it into a new state s' by removing the fluents that are effects of negative conditional effects that fired in s, and then adding the fluents that are effects of positive conditional effects that fired in s. A sequence of operators is a plan for a classical$^+$ planning problem if each is applicable in turn starting from I, and in the resulting end state s, the goal is true (i.e., $G \subseteq s$).

Preferences Various ways of defining and aggregating preferences over plans are reviewed by Baier and McIlraith (2008). We will follow the simple approach of extending the definition of a planning problem $\langle F, I, G, O\rangle$ to have another entry, X, where X is a set of weighted preferences – pairs $\langle G_i, r_i\rangle$ where $G_i \subseteq F$ and $r_i \in \mathbb{R}$. A plan is optimal for $\langle F, I, G, O, X\rangle$ if in the end state s, not only is the goal true, but the weighted sum of satisfied preferences, $\sum_{\langle G_i, r_i\rangle \in X : G_i \subseteq s} r_i$, is maximized. One way to compute an optimal plan is to compile the preferences into operator costs as described by Keyder and Geffner (2009) and then use standard planning tools to find a minimum cost plan.

2.3 RP-MEP

We now are almost ready to describe RP-MEP. First, following Muise et al. (2022) we define general MEP problems.

Definition 2. A multi-agent epistemic planning (MEP) problem is a tuple of the form $\langle \mathcal{P}, \mathcal{A}, Ag, \mathcal{I}, \mathcal{G}\rangle$ where

- $\mathcal{P}$ is the set of propositions;
- $\mathcal{A}$ is a finite set of actions, where an action a is a pair $\langle Pre_a, \mathit{Eff}_a\rangle$ (representing the preconditions and effects), in which
  - Pre_a is a PEKB and
  - Eff_a is a set of conditional effects of the form $\langle \gamma_i, \varphi_i\rangle$ where the PEKB γ_i is the condition and the RML φ_i is the effect;
- Ag is a finite set of agents;
- $\mathcal{I}$ is a PEKB representing the initial state;
- and $\mathcal{G}$ is another PEKB, representing the goal.

RP-MEP problems can then be defined in terms of restrictions on the components of MEP problems. Intuitively, an RP-MEP problem considers things from the perspective of one agent (hence the “perspectival” in the name).

Definition 3. An *RP-MEP problem with depth bound* d is a MEP problem with the restriction that there is some agent $\bigstar \in Ag$ (the *root agent*) such that any RML in the initial state, goal, or an action precondition is of the form $\Box_\bigstar\, \varphi$, any RML in the condition or effect of a conditional effect is either of the form $\Box_\bigstar\, \varphi$ or $\Diamond_\bigstar\, \varphi$, and any RML anywhere in the problem has depth at most $d+1$ (i.e., has at most $d+1$ modal operators, including the initial one for the root agent).

[1] We use the term “classical$^+$” because sometimes classical planning is defined to not have conditional effects.

We now briefly review how actions work in RP-MEP. An action a is applicable in a PEKB P if $P \models \mathit{Pre}_a$. Applying a *progresses* P into another PEKB, $P' = \mathit{Progress}(P, a)$. The transition is such that for each conditional effect $\langle \gamma_i, \varphi_i \rangle$ of the action such that $P \models \gamma_i$, $P' \models \varphi_i$. Additionally, there may be other changes to ensure that P' still follows the $\mathrm{KD45}_n$ axioms. For the full definition of progression in terms of belief revision and update, see the paper by Muise et al. (2022, p. 7).[2] Note that agents can have false beliefs and retract previously held beliefs. For a sequence of actions $\vec{a} = a_1, \ldots, a_k$, the progression $\mathit{Progress}(P, \vec{a})$ can be defined by progressing by each of the actions in turn. $\vec{a}$ is a *plan* for an RP-MEP problem if each action is applicable in turn starting from $\mathcal{I}$ and $\mathit{Progress}(\mathcal{I}, \vec{a}) \models \mathcal{G}$.

Muise et al. (2022) proposed a way of compiling an RP-MEP problem $R = \langle \mathcal{P}, \mathcal{A}, Ag, \mathcal{I}, \mathcal{G} \rangle$ with depth bound d into a corresponding classical$^+$ planning problem $\mathtt{C}(R) = \langle F, I, G, O \rangle$. The set of fluents $F = \mathtt{C}_{\mathtt{fluent}}(\mathcal{P}, Ag, d)$ includes a fluent to represent each RML of the form $\Box_{\bigstar}\, \varphi$ or $\Diamond_{\bigstar}\, \varphi$ of depth $\leq d + 1$. O contains an operator $o = \mathtt{C}_{\mathtt{op}}(a, F)$ corresponding to each action $a \in \mathcal{A}$. The initial state $I = \mathtt{C}_{\mathtt{init}}(\mathcal{I}, F)$ and goal $G = \mathtt{C}_{\mathtt{goal}}(\mathcal{G}, F)$ are also constructed based on their counterparts in the RP-MEP problem. Note that the encoding is fairly modular; e.g., changing the goal of R will change only the goal in $\mathtt{C}(R)$, and not other parts of the encoding. The paper by Muise et al. (2022, pp. 8–10) has the full description of the encoding. The encoding is sound and complete in the following sense.

Theorem 1. *Let R be an RP-MEP problem $\langle \mathcal{P}, \mathcal{A}, Ag, \mathcal{I}, \mathcal{G} \rangle$ and $\mathtt{C}(R) = \langle F, I, G, O \rangle$ its classical$^+$ encoding. Then an action sequence $a_1, \ldots, a_k$ (for any k) is a plan for R' just in case $\mathtt{C}_{op}(a_1, F), \ldots, \mathtt{C}_{op}(a_k, F)$ is a plan for $\mathtt{C}(R)$.*

Proof. It follows from Muise et al.'s (2022) Theorem 2. □

3 Formalizing Epistemic Preferences

We can easily extend RP-MEP problems to incorporate (weighted) preferences.

Definition 4. An RP-MEP problem with preferences (with depth bound d) is a tuple $\langle \mathcal{P}, \mathcal{A}, Ag, \mathcal{I}, \mathcal{G}, \mathit{Prefs} \rangle$ where $\mathcal{P}, \mathcal{A}, Ag, \mathcal{I}, \mathcal{G}$ are as in Definition 3, and Prefs is the set of preferences with associated weights, represented as pairs of the form $\langle \psi_i, r_i \rangle$ where ψ_i is a PEKB containing only RMLs of the form $\Box_{\bigstar}\, \varphi$ of depth $\leq d + 1$, and $r_i \in \mathbb{R}$.

Similarly to classical$^+$ planning with preferences, a plan π is optimal for an RP-MEP problem with preferences if it maximizes the sum of the weights of the preferences satisfied, that is, $\sum_{\langle \psi_i, r_i \rangle \in \mathit{Prefs} : \mathit{Progress}(\mathcal{I}, \pi) \models \psi_i} r_i$.

To illustrate some sorts of epistemic preferences that one might want to be (not) satisfied, consider the following categories (where ℓ is an arbitrary literal). (Note that because of the "perspectival" nature of RP-MEP, any preference has to be expressed in terms of the root agent's beliefs, but the preferences we consider here involve beliefs beyond just that.)

truth (the root agent believes that) agent i correctly believes that the literal ℓ is true: $(\Box_{\bigstar}\, \ell \land \Box_{\bigstar}\, \Box_i\, \ell)$. Note that if both $(\Box_{\bigstar}\, \ell \land \Box_{\bigstar}\, \Box_i\, \ell)$ and $(\Box_{\bigstar}\, \bar{\ell} \land \Box_{\bigstar}\, \Box_i\, \bar{\ell})$ are preferences with same weight, that amounts to a preference about whether (the root agent believes that) agent i has the correct belief as to whether ℓ is true.

misconception (the root agent believes that) agent i incorrectly believes that ℓ is true: $(\Box_{\bigstar}\, \bar{\ell} \land \Box_{\bigstar}\, \Box_i\, \ell)$

oblivious (the root agent believes that) agent i considers ℓ possible (doesn't believe it is false): $\Box_{\bigstar}\, \Diamond_i\, \ell$:

conscious (the root agent believes that) agent i believes ℓ: $\Box_{\bigstar}\, \Box_i\, \ell$

What preferences one would have would often be domain specific (e.g., particular instances of the above four types). However, it is also possible to automatically generate generic preferences (e.g., that every agent in some subset of Ag has the correct beliefs about every literal) which could be useful for, e.g., safety purposes. This is somewhat analagous to Wang et al.'s (2020) various generic sorts of belief-dependent reward functions for POMDPs, such as a "human-certainty" reward function that rewarded the agent for reducing the human's (probabilistic) uncertainty.

The RP-MEP formalism also allows for expressing preferences over nested beliefs. For example, it could be preferred that (the root agent believes that) Alice believes that Bob believes that Alice did not eat his chocolate cake. While solving RP-MEP problems with the classical$^+$ encoding does not scale very well with the depth of nested belief (Muise et al. 2022), we believe that many interesting epistemic preferences have low depth, like the examples we have mentioned.

3.1 Computation

Recall that Muise et al. (2022) showed how to encode an RP-MEP problem R (without preferences) as a classical$^+$ problem $\mathtt{C}(R)$. We can extend that to also encode preferences in a straight-forward way, by encoding each preference formula in the same manner as the goal:

Definition 5. Let R be an RP-MEP problem $\langle \mathcal{P}, \mathcal{A}, Ag, \mathcal{I}, \mathcal{G} \rangle$ with depth bound d and R' that problem extended with preferences: $R' = \langle \mathcal{P}, \mathcal{A}, Ag, \mathcal{I}, \mathcal{G}, \mathit{Prefs} \rangle$ (such that the preferences' RMLs also have depth $\leq d + 1$). We define the encoding of R' as a classical$^+$ problem with preferences as $\mathtt{C}(R') = \langle F, I, G, O, X \rangle$, where $\langle F, I, G, O \rangle = \mathtt{C}(R)$ and $X = \{ \langle \mathtt{C}_{\mathtt{goal}}(\psi_i, F), r_i \rangle : \langle \psi_i, r_i \rangle \in \mathit{Prefs} \}$.

The following theorem shows that an optimal plan for the classical$^+$ problem with preferences will yield an optimal plan for the original RP-MEP problem.

Theorem 2. *Given an RP-MEP problem with preferences $R = \langle \mathcal{P}, \mathcal{A}, Ag, \mathcal{I}, \mathcal{G}, \mathit{Prefs} \rangle$, an action sequence $\vec{a} = a_1, \ldots, a_k$ is an optimal plan for R just in case the operator sequence $\mathtt{C}_{op}(\vec{a}, F) = \mathtt{C}_{op}(a_1, F), \ldots, \mathtt{C}_{op}(a_k, F)$ is an optimal plan for the encoding of R as a classical$^+$ problem with preferences $\mathtt{C}(R) = \langle F, I, G, O, X \rangle$.*

Proof. That $\vec{a}$ is a plan for R just in case $\mathtt{C}_{\mathtt{op}}(\vec{a}, F)$ is a plan for $\mathtt{C}(R)$ follows from Theorem 1. We want to additionally show that $\vec{a}$ satisfies a preference ψ_i in R just in case $\mathtt{C}_{\mathtt{op}}(\vec{a}, F)$ satisfies the preference $\mathtt{C}_{\mathtt{goal}}(\psi_i, F)$ in $\mathtt{C}(R)$. Observe that $\vec{a}$ satisfies the preference ψ_i in R just

[2] We are using deterministic actions only.

Problem	Preference Type **none**		**truth**			**misconception**			**oblivious**			**conscious**		
	$\lvert\pi\rvert$	time	$\lvert\pi\rvert$	time	prefs	$\lvert\pi\rvert$	time	prefs	$\lvert\pi\rvert$	time	prefs	$\lvert\pi\rvert$	time	prefs
Corridor-3	5	0.36	6	0.35	**2**/6	6	0.36	**1**/6	6	0.36	**3**/6	6	0.37	**3**/6
Corridor-5	5	0.40	8	0.63	**4**/10	6	0.64	**2**/10	6	0.43	**6**/10	6	0.46	**5**/10
Corridor-7	5	3.72	8	14.76	**5**/14	9	12.76	**4**/14	6	5.04	**8**/14	7	2.18	**7**/14
Grapevine-4-2	4	4.66	11	37.45	**15**/32	7	51.42	**11**/32	5	33.34	**26**/32	8	33.84	**16**/32
Grapevine-4-4	6	2.91	15	33.32	**14**/32	10	33.30	**10**/32	8	49.93	**24**/32	9	32.60	**16**/32
Grapevine-4-8	11	45.43	14	37.37	**12**/32	13	33.46	**8**/32	13	32.73	**20**/32	12	34.67	**16**/32
Grapevine-8-2	4	31.13	19	61.22	**63**/128	16	57.01	$51^\dagger$/128	6	59.55	118/128	9	61.53	$40^\dagger$/128
Grapevine-8-4	5	24.63	17	59.78	$37^\dagger$/128	15	61.16	$28^\dagger$/128	15	60.71	116/128	13	59.36	$26^\dagger$/128
Grapevine-8-8	7	32.82	32	61.13	60/128	27	60.73	52/128	13	59.93	112/128	18	61.00	$28^\dagger$/128

Table 1: Experimental results. Only problems with depth bound $d = 1$ were used. Corridor-n is a problem in the Corridor domain with n (non-root) agents, and Grapevine-n-g is a problem in the Grapevine domain with n (non-root) agents and g RMLs in its goal. Each "$|\pi|$" column show the length of the found plan. For the **none** preference type, the planner considered only the goal, while for the others there also were automatically generated preferences of the given type. An entry x/y in a "prefs" column indicates that the problem had y preferences, of which x were satisfied by the found plan. As noted below, for problems with preferences of the type **truth**, **misconception**, or **conscious**, it is impossible to satisfy more than 50% of the preferences. If x is bold, that solution is known be optimal; the † annotation indicates that the solution is known to be suboptimal. Times (in seconds) are the times taken by LAMA, the planner, on the encoded classical$^+$ problem with operator costs (encoding times are not included). LAMA was run with a search time limit of 30 seconds (the overall time can be longer).

in case $\vec{a}$ is a plan for the RP-MEP problem (without preferences) $R_i = \langle \mathcal{P}, \mathcal{A}, Ag, \mathcal{I}, \psi_i \rangle$. By Theorem 1, $\vec{a}$ is a plan for R_i just in case $\mathtt{C}_{\mathtt{op}}(\vec{a}, F)$ is a plan for $\mathtt{C}(R_i)$. Since $\mathtt{C}(R_i)$ is identical to $\mathtt{C}(R)$ except for the goal and preferences, $\mathtt{C}_{\mathtt{op}}(\vec{a}, F)$ is a plan for $\mathtt{C}(R_i)$ just in case $\mathtt{C}_{\mathtt{op}}(\vec{a}, F)$ satisfies $\mathtt{C}_{\mathtt{goal}}(\psi_i, F)$ in $\mathtt{C}(R)$. □

4 Experiments

To demonstrate finding plans with epistemic preferences, as a proof of concept we take some epistemic planning problems from the literature and add preferences generated from the **truth**, **misconception**, **oblivious**, and **conscious** categories described in Section 3. In each experiment, all preferences from the relevant category (e.g., **truth**) that involve any non-root agent i and any literal ℓ are generated (excluding some literals that agents always know the truth value of). All the preferences are given weight 1.

The domains we use are the following:

Corridor (Muise et al. 2022, Section 7.2): An agent who has a secret can walk around and make (possibly false) announcements that are believed by other nearby agents.

Grapevine (Muise et al. 2022, Section 7.3): All (non-root) agents can move and make (possibly false) announcements; each starts with its own secret. Agents only believe announcements that don't contradict their own beliefs.

Both domains were slightly modified to make the root agent initially believe the secrets (the **truth** and **misconception** preference types cannot be achieved if the root agent does not have a belief about whether the literal in question is true).

To compute plans for each problem, we first compile the problem into a classical$^+$ planning problem, using the RP-MEP program from Muise et al. (2022). This RP-MEP compilation only has to be done once per problem, since the encoding of the non-preference parts of the problem do not change when preferences are added. Furthermore, since our preferences are automatically generated, we are able to generate them directly in the classical$^+$ encoding rather than having to compile them as a separate step. Each classical$^+$ planning problem with preferences is then transformed by applying essentially the established compilation from preferences into costs from Keyder and Geffner (2009). The resulting planing problem with costs is solved using LAMA (Richter and Westphal 2010), a configuration of Fast Downward (v22.12) (Helmert 2006).

The results[3] are shown in Table 1. In each case the planner was able to find plans which satisfy many of the preferences. Note that for problems with preferences of the **truth**, **misconception**, and **conscious** types, it would not be possible to satisfy more than half the preferences (an agent cannot believe both ℓ and $\bar{\ell}$). To determine whether solutions were optimal we ran additional experiments in which LAMA had no time limit (in some cases memory limitations precluded deciding that question). Compilation times for RP-MEP (0.49–5.51 seconds) and the Keyder and Geffner encoding (0.03–0.10 seconds) are omitted. For more details and the code, see https://github.com/tqk/epistemic-preferences.

5 Conclusion

We have considered planning with *epistemic* preferences, i.e., over knowledge or beliefs. Our approach to computing plans for such problems makes use of two existing compilations – the RP-MEP encoding of epistemic planning problems as classical$^+$ problems, and Keyder and Geffner's (2009) compilation of preferences into costs. An advantage of this is that further developments in improving the efficiency of traditional planning would also help our approach. Future work may further explore applications of planning with epistemic preferences in areas such as AI safety.

[3] from a Linux system with two Intel Xeon E5-2667 v4 processors and 32 GB of RAM

Acknowledgements

We gratefully acknowledge funding from the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canada CIFAR AI Chairs Program, and Microsoft Research. The first author also received funding from Open Philanthropy.

References

Amodei, D.; Olah, C.; Steinhardt, J.; Christiano, P. F.; Schulman, J.; and Mané, D. 2016. Concrete problems in AI safety. *arXiv preprint arXiv:1606.06565*.

Baier, J. A., and McIlraith, S. A. 2008. Planning with preferences. *AI Magazine* 29(4):25–36.

Baier, J. A.; Mombourquette, B.; and McIlraith, S. A. 2014. Diagnostic problem solving via planning with ontic and epistemic goals. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference, KR 2014*, 388–397. AAAI Press.

Baral, C.; Bolander, T.; van Ditmarsch, H.; and McIlraith, S. 2017. Epistemic Planning (Dagstuhl Seminar 17231). *Dagstuhl Reports* 7(6):1–47.

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.

Cooper, M. C.; Herzig, A.; Maffre, F.; Maris, F.; Perrotin, E.; and Régnier, P. 2021. A lightweight epistemic logic and its application to planning. *Artificial Intelligence* 298:103437.

Engesser, T.; Bolander, T.; Mattmüller, R.; and Nebel, B. 2017. Cooperative epistemic multi-agent planning for implicit coordination. In *Proceedings of the Ninth Workshop on Methods for Modalities, M4M@ICLA*, volume 243 of *EPTCS*, 75–90.

Fabiano, F.; Burigana, A.; Dovier, A.; and Pontelli, E. 2020. EFP 2.0: A multi-agent epistemic solver with multiple e-state representations. In *Proceedings of the Thirtieth International Conference on Automated Planning and Scheduling*, 101–109. AAAI Press.

Fabiano, F.; Burigana, A.; Dovier, A.; Pontelli, E.; and Son, T. C. 2021. Multi-agent epistemic planning with inconsistent beliefs, trust and lies. In *PRICAI 2021: Trends in Artificial Intelligence*, volume 13031 of *Lecture Notes in Computer Science*, 586–597. Springer.

Fagin, R.; Halpern, J. Y.; Moses, Y.; and Vardi, M. Y. 1995. *Reasoning About Knowledge*. MIT Press.

Helmert, M. 2006. The Fast Downward planning system. *Journal of Artificial Intelligence Research* 26:191–246.

Hu, G.; Miller, T.; and Lipovetzky, N. 2022. Planning with perspectives – decomposing epistemic planning using functional STRIPS. *Journal of Artificial Intelligence Research* 75:489–539.

Keyder, E., and Geffner, H. 2009. Soft goals can be compiled away. *Journal of Artificial Intelligence Research* 36:547–556.

Klassen, T. Q.; Alamdari, P. A.; and McIlraith, S. A. 2022. Epistemic side effects & avoiding them (sometimes). In *2022 NeurIPS ML Safety Workshop*.

Klassen, T. Q.; Alamdari, P. A.; and McIlraith, S. A. 2023. Epistemic side effects: An AI safety problem. In *Proceedings of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023) (Blue Sky Ideas Track)*, 1797–1801.

Klassen, T. Q.; McIlraith, S. A.; Muise, C.; and Xu, J. 2022. Planning to avoid side effects. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI 2022)*, 9830–9839.

Kominis, F., and Geffner, H. 2015. Beliefs in multiagent planning: From one agent to many. In *Proceedings of the Twenty-Fifth International Conference on Automated Planning and Scheduling, ICAPS 2015,*, 147–155. AAAI Press.

Lakemeyer, G., and Lespérance, Y. 2012. Efficient reasoning in multiagent epistemic logics. In *ECAI 2012 - 20th European Conference on Artificial Intelligence*, 498–503. IOS Press.

Le, T.; Fabiano, F.; Son, T. C.; and Pontelli, E. 2018. EFP and PG-EFP: epistemic forward search planners in multiagent domains. In *Proceedings of the Twenty-Eighth International Conference on Automated Planning and Scheduling, ICAPS 2018*, 161–170. AAAI Press.

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

Richter, S., and Westphal, M. 2010. The LAMA planner: Guiding cost-based anytime planning with landmarks. *Journal of Artificial Intelligence Research* 39:127–177.

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

Wang, A.; Chitnis, R.; Li, M.; Kaelbling, L. P.; and Lozano-Pérez, T. 2020. A unifying framework for social motivation in human-robot interaction. In *The AAAI 2020 Workshop on Plan, Activity, and Intent Recognition (PAIR 2020).*