

Signaling Friends and Head-Faking Enemies Simultaneously: Balancing Goal Obfuscation and Goal Legibility

Anagha Kulkarni and Siddharth Srivastava and Subbarao Kambhampati

CIDSE, Arizona State University, Tempe, AZ 85281 USA

{anaghak, siddharths, rao}@asu.edu

Abstract

In order to be useful in the real world, an AI agent needs to plan and act in the presence of other agents, who may be helpful or disruptive. In this paper, we consider the problem where an autonomous agent needs to act in a manner that clarifies its objectives to cooperative agents while preventing adversarial agents from inferring those objectives. This problem is solvable when the agent has access to the sensor models used by the cooperative and adversarial agents and access to their prior knowledge. We develop two new solution approaches: one provides an optimal solution to the problem given a fixed time horizon by using an integer programming solver, the other provides a satisficing solution using heuristic-guided forward search to achieve prespecified amount of obfuscation and legibility for adversarial and cooperative agents respectively. We show the feasibility and utility of our algorithms through extensive empirical evaluation on planning benchmarks.

1 Introduction

In a multi-agent environment, the activities performed by an agent may be observed by other agents. In such an environment, an agent should perform its tasks while taking into account the observers’ sensing capabilities and its relationship with the observers. Several prior works have explored the generation of legible agent behavior in the presence of cooperative observers [Dragan and Srinivasa, 2013; Zhang *et al.*, 2017; MacNally *et al.*, 2018] and obfuscating agent behavior in the presence of adversarial observers [Keren *et al.*, 2016b; Masters and Sardina, 2017b; Shekhar and Brafman, 2018; Kulkarni *et al.*, 2019]. Through legible behavior, an agent can convey the necessary information to a cooperative observer, while through obfuscating behavior it can hide sensitive information from an adversary. However, these prior works assume that the observers in the environment are either entirely cooperative or entirely adversarial.

In real-world settings of strategic importance, an agent might encounter both types of observers *simultaneously*. This would necessitate synthesizing a behavior that is simultaneously legible to friendly entities and obfuscatory to adversarial ones. For instance, in soccer, a player may perform feinting trick to confuse an opponent while signaling a teammate. Synthesizing a single behavior that is legible and obfuscatory

to different agents presents significant technical challenges. In particular, the agent may have to be deliberately less legible to its friends so that it can be effectively more obfuscatory to its adversaries. This problem gives rise to a novel optimization space that involves trading-off the amount of obfuscation desired for adversaries with the amount of legibility desired for friends.

In this paper, we present a novel problem framework called *mixed-observer controlled observability planning problem*, MO-COPP, that allows an agent to simultaneously control information yielded to both cooperative and adversarial observers while achieving its goal. Our framework models and exploits situations where different observers have differing sensing capabilities, which result in different “perceptions” of the same information. Typically, different observers in an environment may have different “sensors” (perception capabilities) due to differences in prior communication, background knowledge, and raw sensor hardware. For example, an agent may establish semaphore actions that are more meaningful to its allies than to its adversaries. Therefore, we assume that different types of observers have differing sensor models.

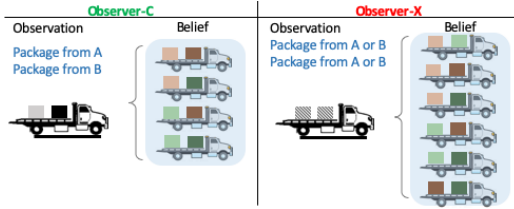
We propose two approaches to solve MO-COPP. **(1)** A novel integer programming (IP) encoding, which provides an optimal solution given a fixed time horizon. This involves maximizing the amount of trade-off between obfuscation achieved for adversarial observer and legibility achieved for cooperative observer with respect to the agent’s objectives. **(2)** A heuristic-guided search algorithm, which leverages a prior approach [Kulkarni *et al.*, 2019] that was designed to address entirely cooperative or entirely adversarial settings. Through theoretical and empirical analysis, we explore the strengths of the two proposed approaches. Additionally, we show that for mixed-observer setting, our solution approaches add significant value over approaches that consider either entirely cooperative or adversarial observers.

Motivating Example Consider the example in Figure 1, where a truck agent has to deliver two packages. As shown in Figure 1a, there are two packages at factory A and two at factory B. Let there be two observers who know the initial inventory at each factory, but do not know which two packages will be delivered by the agent. Each observer has different sensors (or sources of information) about the truck’s activities. Observer-C (cooperative observer) has access to a database of barcodes which allows it to determine the identity of the factory that the package was loaded from (but not the

(a) The actor's goal is to deliver two packages to the delivery area:



(b) Plan-1 - Actor delivers 1 package from factory A and 1 from B:



(c) Plan-2 - Actor delivers both the packages from factory A, helping observer-C and thwarting observer-X:

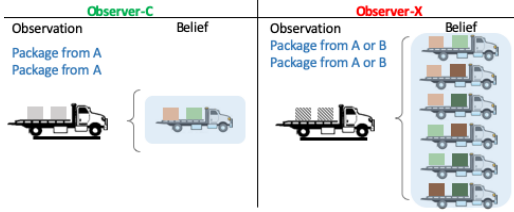


Figure 1: Different belief updates induced by the same plan for two observers; observer-C is cooperative, observer-X is adversarial.

identity of the individual packages), whereas observer-X (adversarial observer) does not have any additional information and cannot differentiate between any of the packages. The truck agent wants to convey its true goal to observer-C but obfuscate it to observer-X. When the agent loads one package each from A and B (Figure 1b), observer-C cannot distinguish the packages. However, when it loads both the packages from A (Figure 1c), observer-C can infer the identity of the loaded packages, while observer-X still does not know the identity of the packages. Therefore, plan-2 makes the true goal legible to observer-C while obfuscating the true goal with all the possible goals for observer-X.

2 MO-COPP

MO-COPP setting involves an actor (\mathbf{A}) and two observers, where one is adversarial observer (\mathbf{X}) while the other is cooperative (\mathbf{C}). We assume that the actor has full observability of its own activities and knows the sensor models used by the observers, and that the observers have different sensor models.¹ When the actor takes an action and reaches a new state, each observer receives an observation. After obtaining the observations, each observer updates its belief. The actor leverages the known limits in the observers' sensors to control the observability of multiple observers in the environment simultaneously. Given a set of candidate goals, the objective of the actor is to convey information about its goal to the cooperative observer and to hide it from the adversarial observer. This involves increasing the number of candidate goals possible in the adversary's belief, while decreasing the number of candidate goals possible in cooperative observer's belief.

¹If the actor does not have access to the sensor models of the observers or both the observers have the same sensor model, then there is of course no question of balancing obfuscation and legibility.

Definition 1. A *mixed-observer controlled observability planning problem* is a tuple, MO-COPP = $\langle \Lambda, \mathcal{P}, \mathcal{G}, \{\Omega_i\}_{i \in \Lambda}, \{\mathcal{O}_i\}_{i \in \Lambda}, \{\mathcal{B}_0^i\}_{i \in \{\mathbf{X}, \mathbf{C}\}} \rangle$, where,

- $\Lambda = \{\mathbf{A}, \mathbf{C}, \mathbf{X}\}$ is the set of agents.
- $\mathcal{P} = \langle \mathcal{F}, \mathcal{O}_p, \mathcal{I}, G_A \rangle$ is \mathbf{A} 's task captured as a classical planning problem [Geffner and Bonet, 2013], where \mathcal{F} is the set of fluents, \mathcal{O}_p is the set of actions, \mathcal{I} is the initial state and goal G_A is a subset of fluents. Also, for $a \in \mathcal{O}_p$, $pre(a)$, $add(a)$, $delete(a)$ are each a subset of fluents representing precondition, add effects and delete effects of a .
- $\mathcal{G} = \{G_1, G_2, \dots, G_{n-1}, G_A\}$ is the set of candidate goals, where G_A is the true goal of \mathbf{A} , which is not known to both \mathbf{C} and \mathbf{X} .
- Ω_i is the set of observation symbols for agent i , which are emitted when \mathbf{A} takes an action and reaches a new state. Further, $\Omega_A = \{o_{a,s}^A \mid a \in \mathcal{O}_p, s \in \mathcal{S}\}$.
- $\mathcal{O}_i : \mathcal{O}_p \times \mathcal{S} \rightarrow \Omega_i$ is agent i 's deterministic sensor model. \mathcal{S} is the set of states, where each state is instantiation of all fluents. Further, \mathcal{O}_A maps each action-state pair to a unique observation, $\forall a, a' \in \mathcal{O}_p, s, s' \in \mathcal{S}, a \neq a' \wedge s \neq s' : \mathcal{O}_A(a, s) \neq \mathcal{O}_A(a', s')$, while \mathcal{O}_X and \mathcal{O}_C are noisy sensor models that map multiple action-state pairs to the same observation symbol.
- \mathcal{B}_0^i is the initial belief of an observer, $i \in \{\mathbf{X}, \mathbf{C}\}$. The initial belief is a set of states inclusive of \mathcal{I} .

From the above definition, we see that, while \mathbf{X} and \mathbf{C} can have arbitrary sensor models with partial observability, \mathbf{A} has full observability, due to one-to-one mapping of the sensor model. The observers are aware of the planning domain of the actor and of the candidate goals, except they do not know which candidate goal is the actor's true goal, G_A .

Every time the actor acts, each $i \in \Lambda$ receives an observation consistent with its sensor model. The sensor model of an observer $i \in \{\mathbf{X}, \mathbf{C}\}$ supports many-to-one mapping of $\langle a, s \rangle$ pairs to observation symbols, i.e., $\exists a, a' \in \mathcal{O}_p, s, s' \in \mathcal{S}, a \neq a' \wedge s \neq s' : \mathcal{O}_i(a, s) = \mathcal{O}_i(a', s')$.² For an agent i , the inverse of sensor model gives the set of $\langle a, s \rangle$ pairs consistent with an observation symbol $o^i \in \Omega_i$, i.e., $\mathcal{O}_i^{-1}(o^i) = \{\langle a, s \rangle \mid \forall a \in \mathcal{O}_p, s \in \mathcal{S}, \mathcal{O}_i(a, s) = o^i\}$.

Each observer $i \in \{\mathbf{X}, \mathbf{C}\}$ maintains its own belief, which is a set of states. $\Gamma(\cdot)$ is a transition function, such that, $\Gamma(s, a) = \perp$ if $s \not\models pre(a)$; else $\Gamma(s, a) = s \cup add(a) \setminus delete(a)$. Now we can define the belief update: (1) at time step $t = 0$, the initial belief of observer i is given by \mathcal{B}_0^i , (2) at time step $t \in \{1, \dots, \mathcal{T}\}$, $\mathcal{B}_t^i = \{\hat{s} \mid \exists \hat{a}, \bar{s} \Gamma(\bar{s}, \hat{a}) = \hat{s}; \bar{s} \in \mathcal{B}_{t-1}^i; \mathcal{O}_i(\hat{a}, \bar{s}) = o_t^i\}$, where \mathcal{T} is the time horizon. That is, the belief is updated using the previous belief and the observation received. A plan π (or sequence of actions) is associated with a sequence of observations, $ObsSeq_i(\pi) = \{o_1^i, \dots, o_{\mathcal{T}}^i\}$ for each observer i . The belief of \mathbf{A} is a single state due to full observability.

MO-COPP Solutions A plan π solves a MO-COPP = $\langle \Lambda, \mathcal{P}, \mathcal{G}, \{\Omega_i\}_{i \in \Lambda}, \{\mathcal{O}_i\}_{i \in \Lambda}, \{\mathcal{B}_0^i\}_{i \in \{\mathbf{X}, \mathbf{C}\}} \rangle$, iff: $\Gamma(\mathcal{I}, \pi) \models G_A$.

²Our formulation supports sensor models that yield no observations. This can be modeled by using a dummy observation symbol that essentially maps to all the possible $\langle a, s \rangle$ pairs.

In other words, any solution to \mathcal{P} is a solution to MO-COPP. We measure the quality of a MO-COPP solution in terms of the extent to which the plan is legible to \mathbf{C} and obfuscating to \mathbf{X} . Given a solution plan, π , let $\mathcal{G}_{\mathbf{X}}^{\pi} = \{G \in \mathcal{G} : \exists s \in \mathcal{B}_{\mathcal{T}}^{\mathbf{X}}, s \models G\}$, denote the goals present in \mathbf{X} 's final belief and $\mathcal{G}_{\mathbf{C}}^{\pi} = \{G \in \mathcal{G} : \exists s \in \mathcal{B}_{\mathcal{T}}^{\mathbf{C}}, s \models G\}$, denote the goals present in \mathbf{C} 's final belief. Also, we can define the inverse of its observation sequence for observer i , as follows: $\forall i \in \{\mathbf{X}, \mathbf{C}\}, \text{ObsSeq}_i^{-1}(\pi) = \{\hat{\pi} | \exists s_0^i \in \mathcal{B}_0^i, G \in \mathcal{G}_i^{\pi}, \Gamma(s_0^i, \hat{\pi}) \models G, \text{ObsSeq}_i(\pi) = \text{ObsSeq}_i(\hat{\pi})\}$.

Here $|\mathcal{G}_{\mathbf{X}}^{\pi}|$ represents the amount of **goal obfuscation for \mathbf{X}** , and by increasing the number of goals in $\mathcal{G}_{\mathbf{X}}^{\pi}$, we can increase the goal obfuscation. Similarly, $|\mathcal{G}_{\mathbf{C}}^{\pi}|$ represents the amount of **goal legibility for \mathbf{C}** , and by decreasing the number of goals in $\mathcal{G}_{\mathbf{C}}^{\pi}$, we can increase the goal legibility. We now introduce a goal difference metric to estimate the quality of a solution plan to MO-COPP.

Definition 2. Given a solution plan π that solves MO-COPP, where $|\mathcal{G}| = n$, the **goal difference**, of π is given by: $GD(\pi) = \frac{|\mathcal{G}_{\mathbf{X}}^{\pi}| - |\mathcal{G}_{\mathbf{C}}^{\pi}|}{n-1}$ where the denominator represents the difference between ideal values of $\mathcal{G}_{\mathbf{X}}^{\pi}$ and $\mathcal{G}_{\mathbf{C}}^{\pi}$.

An optimal solution to MO-COPP maximizes the trade-off between amount of goal obfuscation and goal legibility. That is, it maximizes the difference between the number of goals in $\mathcal{G}_{\mathbf{X}}^{\pi}$ and $\mathcal{G}_{\mathbf{C}}^{\pi}$. Equivalently, closer the $GD(\pi)$ value to 1, better is the plan quality. A solution plan with $GD(\pi) = 1$ is an optimal plan. E.g., in Figure 1c, plan-2 is an optimal plan with $GD(\text{plan-2}) = \frac{6-1}{6-1} = 1$. The denominator is essential for comparing the quality of plans across different problems, with varying number of candidate goals.

Proposition 1. Given a solution plan, π , to MO-COPP, if $|\mathcal{G}_{\mathbf{C}}^{\pi}| = 1$, then $G_{\mathbf{A}} \in \mathcal{G}_{\mathbf{C}}^{\pi}$.

The above proposition states that when maximum goal legibility is achieved, only one goal $G_{\mathbf{A}}$ is present in \mathbf{C} 's final belief. This is because $\mathcal{I} \in \mathcal{B}_0^{\mathbf{C}}$, and while performing belief update at t , $\langle a_t, s_t \rangle \in \mathcal{O}_{\mathcal{T}}^{-1}(o_t^{\mathbf{C}})$. That is the true state is part of each belief update. Therefore, $G_{\mathbf{A}} \in \mathcal{G}_{\mathbf{C}}^{\pi}$.

3 Plan Computation

We now present two solution approaches. In the first approach, we formulate MO-COPP as a constraint optimization problem and provide an IP encoding to solve it in \mathcal{T} steps. The IP encoding provides an optimal solution for the given horizon by maximizing the trade-off between the amount of obfuscation and legibility. In the second approach, we use a heuristic-guided forward search to achieve preset levels of goal obfuscation and legibility. The search algorithm generates solutions that satisfy a prespecified lower bound on the amount of goal obfuscation and goal legibility.

3.1 MO-COPP as Integer Program

Variables We require the following binary variables for our encoding: $x_{a,t}$ is an indicator variable for action a at time t , $y_{s,t}$ is an indicator variable for state s at time t , $w_{o,t}^i$ is an indicator variable for observation o^i at time t , $b_{s,t}^i$ is an indicator variable for state s in belief \mathcal{B}^i at time t , $h_{s,a,t}^i$ is an indicator variable for action a being applicable in state s in

belief \mathcal{B}^i at time t and $g_{G,\mathcal{T}}^i$ is an indicator variable for a goal G present in belief $\mathcal{B}_{\mathcal{T}}^i$.

Objective Function The objective function is essentially the numerator of $GD(\cdot)$ metric, i.e., $\max \sum_{G \in \mathcal{G}} g_{G,\mathcal{T}}^{\mathbf{X}} - \sum_{G \in \mathcal{G}} g_{G,\mathcal{T}}^{\mathbf{C}}$. We skip the denominator of the GD metric, as it is a constant and does not contribute to the optimization. This provides a single solution that achieves the maximum difference between the number of goals possible for the two observers. Note that, it would make sense to get the Pareto optimal solutions if we wanted to explore all the combinations of goals achieved for the two observers. However, that is not our objective.

IP Constraints

$$\forall s \in \mathcal{S}, s = \mathcal{I} : y_{s,0} = 1; s \neq \mathcal{I} : y_{s,0} = 0; \sum_{G_{\mathbf{A}} \in s} y_{s,\mathcal{T}} = 1 \quad (1)$$

$$\forall i \in \{\mathbf{X}, \mathbf{C}\}, s \in \mathcal{S}, s \in \mathcal{B}_0^i : b_{s,0}^i = 1; s \notin \mathcal{B}_0^i : b_{s,0}^i = 0 \quad (2)$$

$$\forall i \in \{\mathbf{X}, \mathbf{C}\}, G \in \mathcal{G}, m > |\{s | G \in s\}| : m * g_{G,\mathcal{T}}^i - \sum_{G \in s} b_{s,\mathcal{T}}^i \geq 0 \quad (3)$$

$$\forall a \in \mathcal{O}_p, t \in \{1, \dots, \mathcal{T}\}, \text{pre}_a = \{s | \text{pre}(a) \in s\} : x_{a,t} \leq \sum_{s \in \text{pre}_a} y_{s,t-1} \quad (4)$$

$$\forall s, s' \in \mathcal{S}, t \in \{1, \dots, \mathcal{T}\}, \text{add}_{s'} = \{a | \text{pre}(a) \in s, \text{add}(a) \setminus \text{delete}(a) \in s'\}, \text{pre}_{s'} = \{s | \text{pre}(a) \in s \wedge \text{add}(a) \setminus \text{delete}(a) \in s'\} : \sum_{a \in \text{add}_{s'}} x_{a,t} + \sum_{s \in \text{pre}_{s'}} y_{s,t-1} - 2y_{s',t} \geq 0 \quad (5)$$

$$\forall a \in \mathcal{O}_p, t \in \{1, \dots, \mathcal{T}\}, \text{post}_a = \{s' | \text{add}(a) \setminus \text{delete}(a) \in s'\} : \sum_{s \in \text{pre}_a, s' \in \text{post}_a} y_{s,t-1} y_{s',t} = x_{a,t} \quad (6)$$

$$\forall i \in \{\mathbf{X}, \mathbf{C}\}, o \in \Omega_i, t \in \{1, \dots, \mathcal{T}\} : w_{o,t}^i = \sum_{a, s' \in \mathcal{O}_o^i} x_{a,t} y_{s',t} \quad (7)$$

$$\forall i \in \{\mathbf{X}, \mathbf{C}\}, s \in \mathcal{S}, t \in \{1, \dots, \mathcal{T}\}, a \in \text{add}_s, \text{add}_s = \{a | \text{pre}(a) \in s\} : b_{s,t-1}^i + w_{o,t}^i - h_{s,a,t}^i \leq 1 \quad (8)$$

$$\forall i \in \{\mathbf{X}, \mathbf{C}\}, s \in \mathcal{S}, o \in \Omega_i, t \in \{1, \dots, \mathcal{T}\}, a \in \text{add}_s, \text{add}_s = \{a | \text{pre}(a) \in s\} : h_{s,a,t}^i - b_{s,t-1}^i \leq 0 \quad (9)$$

$$\forall i \in \{\mathbf{X}, \mathbf{C}\}, s \in \mathcal{S}, t \in \{1, \dots, \mathcal{T}\}, a \in \text{add}_s, s' \in \text{post}_s, \text{add}_s = \{a | \text{pre}(a) \in s\}, \text{post}_s = \{s' | \text{add}(a) \setminus \text{delete}(a) \in s'\} : h_{s,a,t}^i - b_{s',t}^i \leq 0 \quad (10)$$

$$\forall i \in \{\mathbf{X}, \mathbf{C}\}, s \in \mathcal{S}, o \in \Omega_i, t \in \{1, \dots, \mathcal{T}\}, a \in \text{add}_s, \text{add}_s = \{a | \text{pre}(a) \in s\} : h_{s,a,t}^i - w_{o,t}^i \leq 0 \quad (11)$$

$$\forall i \in \{\mathbf{X}, \mathbf{C}\}, s, s' \in \mathcal{S}, t \in \{1, \dots, \mathcal{T}\}, \text{add}_{s'} = \{a | \text{pre}(a) \in s \wedge \text{add}(a) \setminus \text{delete}(a) \in s'\}, \text{pre}_{s'} = \{s | \text{pre}(a) \in s \wedge \text{add}(a) \setminus \text{delete}(a) \in s'\} : \sum_{s \in \text{pre}_{s'}, a \in \text{add}_{s'}} h_{s,a,t}^i - b_{s',t}^i \geq 0 \quad (12)$$

$$\forall t \in \{1, \dots, \mathcal{T}\} : \sum_{a \in \mathcal{O}_p} x_{a,t} \leq 1 \quad (13)$$

Constraint (1) sets the initial state and says that a state that satisfies the true goal should be achieved in the last time step for \mathbf{A} . Constraint (2) sets the initial belief for both the observers. Constraint (3) says that if a goal is satisfied in the final belief of an observer then the corresponding goal variable will be true. Constraint (4) through (6) enforce the transition function. Specifically, constraint (4) validates the applicability of an action in a state, constraint (5) states that for a resulting state to be true both the action and the state in which it is

applied should be true, and similarly constraint (6) validates an action with respect to its previous state and the resulting state. Constraint (7) enforces the corresponding observation symbol for each observer depending on the $\langle a, s' \rangle$ pair. Constraints (8) through (12) enforce a belief update. Specifically, constraint (8) states that an action is not applicable in a belief state if either the belief state or the observation is untrue. Constraint (9) states that an action cannot be applied in a belief state that is untrue. Constraint (10) states that an action cannot be true if the resulting belief state is untrue. Constraint (11) states that an action cannot be true if the corresponding observation is untrue. Constraint (12) states that a belief state is true if the sum of actions leading to it is at least 1. Constraint (13) ensures only one action is possible at each step.

Proposition 2. *The IP encoding listed above which takes time horizon \mathcal{T} as input, solves a MO-COPP = $\langle \Lambda, \mathcal{P}, \mathcal{G}, \{\Omega_i\}_{i \in \Lambda}, \{\mathcal{O}_i\}_{i \in \Lambda}, \{\mathcal{B}_0^i\}_{i \in \{\mathbf{X}, \mathbf{C}\}} \rangle$ such that, the following properties hold:*

- **Soundness:** A solution to the IP will solve MO-COPP.
- **Completeness:** If there exists a plan that solves the MO-COPP in \mathcal{T} time steps, then it will be a feasible solution for the IP encoding.
- **Optimality:** An optimal solution to the IP encoding will be a plan that solves the MO-COPP with an optimal value of GD given the time horizon, \mathcal{T} .

Since a solution to the IP, π_{IP} , satisfies (1), it solves MO-COPP. If there exists a plan $\pi_{\mathcal{T}}$ of \mathcal{T} time steps that solves MO-COPP, $\Gamma(\mathcal{I}, \pi_{\mathcal{T}}) \models G_A$, then $\pi_{\mathcal{T}}$ will satisfy the constraints (1), (4)-(6) and (13). The IP encoding uses the numerator of GD metric as its objective function. An optimal solution to the IP encoding, π_{IP}^* , is an optimal plan to MO-COPP given plans of \mathcal{T} time steps.

Goal Constraints The objective function trades off goal obfuscation with goal legibility for the observers. However, the actor can ensure a predefined level of goal obfuscation (say obfuscate with at least k candidate goals) by $\sum_{G \in \mathcal{G}} g_{G, \mathcal{T}}^{\mathbf{X}} \geq k$, *s.t.* $1 \leq k \leq |\mathcal{G}|$ and goal legibility (say legible with at most j goals) by $\sum_{G \in \mathcal{G}} g_{G, \mathcal{T}}^{\mathbf{C}} \leq j$, *s.t.* $1 \leq j \leq |\mathcal{G}|$. These constraints allow the actor to filter out solutions that do not satisfy minimum bound for goal obfuscation and goal legibility. The actor can improve the robustness of the plans generated by using these constraints.

3.2 Search Algorithm

In this section, we show that it is possible to leverage search techniques that address goal obfuscation and goal legibility in isolation to solve MO-COPP. We adapt Kulkarni *et al.* [2019] approach to address goal obfuscation and goal legibility simultaneously to two different observers. We specify bounds on the amount of goal obfuscation and goal legibility desired, similar to the ones seen in the IP: obfuscate with at least k goals, make it legible with at most j goals. These bounds, $\Phi = \langle \Phi_{\mathbf{X}}, \Phi_{\mathbf{C}} \rangle$, are given as input to the search algorithm.

Each search node maintains the associated beliefs for both observers. The *approx* function generates an approximate belief, b_{Δ}^i , of size Δ (i.e. cardinality of b_{Δ}^i is Δ). b_{Δ}^i is always inclusive of the true state of the actor, this is because the actor can only take actions that are consistent with its true state. If all such Δ -sized beliefs are explored then b_{Δ}^i of

Algorithm 1 Heuristic-Guided Search

```

1: Initialize open, closed and temp lists;  $\Delta = 1$ 
2:  $\langle b_{\Delta}^{\mathbf{X}}, b_{\Delta}^{\mathbf{C}} \rangle \leftarrow \text{approx}(\mathcal{I}, \mathcal{B}_0^{\mathbf{X}}, \mathcal{B}_0^{\mathbf{C}})$ 
3: open.push( $\mathcal{I}, \langle b_{\Delta}^{\mathbf{X}}, b_{\Delta}^{\mathbf{C}} \rangle, \langle \mathcal{B}_0^{\mathbf{X}}, \mathcal{B}_0^{\mathbf{C}} \rangle, \text{priority} = 0$ )
4: while  $\Delta \leq |\mathcal{S}|$  do
5:   while open  $\neq \emptyset$  do
6:      $s, \langle b_{\Delta}^{\mathbf{X}}, b_{\Delta}^{\mathbf{C}} \rangle, \langle \mathcal{B}^{\mathbf{X}}, \mathcal{B}^{\mathbf{C}} \rangle, h_{node} \leftarrow \text{open.pop}()$ 
7:     if  $|b_{\Delta}^{\mathbf{X}}| > \Delta$  or  $|b_{\Delta}^{\mathbf{C}}| > \Delta$  then
8:       temp.push( $s, \langle b_{\Delta}^{\mathbf{X}}, b_{\Delta}^{\mathbf{C}} \rangle, \langle \mathcal{B}^{\mathbf{X}}, \mathcal{B}^{\mathbf{C}} \rangle, h_{node}$ )
9:       continue
10:    end if
11:    add  $\langle b_{\Delta}^{\mathbf{X}}, b_{\Delta}^{\mathbf{C}} \rangle$  to closed
12:    if  $s \models G_A$  and  $\mathcal{B}^{\mathbf{X}} \models \Phi_{\mathbf{X}}$  and  $\mathcal{B}^{\mathbf{C}} \models \Phi_{\mathbf{C}}$  then
13:      return  $\pi, \text{ObsSeq}_{\mathbf{X}}(\pi), \text{ObsSeq}_{\mathbf{C}}(\pi)$ 
14:    end if
15:    for  $s' \in \text{successors}(s)$  do
16:       $o^{\mathbf{X}} \leftarrow \mathcal{O}_{\mathbf{X}}(a, s'); o^{\mathbf{C}} \leftarrow \mathcal{O}_{\mathbf{C}}(a, s')$ 
17:       $\widehat{\mathcal{B}}^{\mathbf{X}} = \text{Update}(\mathcal{B}^{\mathbf{X}}, o^{\mathbf{X}}); \widehat{\mathcal{B}}^{\mathbf{C}} = \text{Update}(\mathcal{B}^{\mathbf{C}}, o^{\mathbf{C}})$ 
18:       $\langle \widehat{b}_{\Delta}^{\mathbf{X}}, \widehat{b}_{\Delta}^{\mathbf{C}} \rangle \leftarrow \text{approx}(s', \widehat{\mathcal{B}}^{\mathbf{X}}, \widehat{\mathcal{B}}^{\mathbf{C}})$ 
19:       $h_{node} \leftarrow h_{G_A}(s') + h_{\mathcal{G}_{k-1}}(\widehat{\mathcal{B}}^{\mathbf{X}}) - h_{\mathcal{G}_{j-j}}(\widehat{\mathcal{B}}^{\mathbf{C}})$ 
20:      add new node to open if  $\langle \widehat{b}_{\Delta}^{\mathbf{X}}, \widehat{b}_{\Delta}^{\mathbf{C}} \rangle$  not in closed
21:    end for
22:  end while
23:  increment  $\Delta$ ; copy items from temp to open; empty temp
24: end while

```

$\Delta + 1$ size is computed, and this node gets put in the temporary list and is explored in the next outer iteration when Δ has been incremented. For each Δ , all Δ -sized unique combinations of belief (that include the actual state of the actor) are explored. This allows systematic and complete exploration of multiple paths to a given search node. The inner iteration performs heuristic guided forward search (we use greedy best first search) to find a plan while tracking at most Δ states in each b_{Δ}^i . In the inner loop, the node expansion is guided by (1) customized heuristic function, which computes value of the node based on true goal and belief constraints given by Φ for the observers, and (2) goal test, which checks for satisfaction of true goal and satisfaction of the belief constraints given by Φ . The algorithm stops either when a solution is found or when all the Δ iterations have been explored.

Proposition 3. *The search algorithm listed above which takes goal-constraints Φ as input, solves a MO-COPP = $\langle \Lambda, \mathcal{P}, \mathcal{G}, \{\Omega_i\}_{i \in \Lambda}, \{\mathcal{O}_i\}_{i \in \Lambda}, \{\mathcal{B}_0^i\}_{i \in \{\mathbf{X}, \mathbf{C}\}} \rangle$ such that, the following properties hold:*

- **Soundness** Any solution to the search algorithm is a plan that solves the MO-COPP.
- **Completeness** If there exists a plan that solves MO-COPP given goal-constraints Φ , it will be found by the search.

A solution to the search algorithm solves MO-COPP, since the goal test ensures the true state of \mathbf{A} satisfies G_A . The search algorithm necessarily terminates in $|\mathcal{S}|$ iterations of the Δ parameter. The Δ parameter allows systematic exploration of unique Δ -sized combinations of belief, starting with $\Delta = 1$ until a solution is found or the solution space is explored, $\Delta = |\mathcal{S}|$. The goal test checks for satisfaction of constraints in Φ . Hence, a plan that solves MO-COPP given Φ will be found by the search algorithm.

Property In both the solution approaches, we can assert a lower bound on the extent of goal obfuscation and goal legibility for a MO-COPP solution plan. In IP, we can specify the aforementioned goal constraints to assert this minimum

value, while in the search, the goal tests allow us to assert it. By setting k, j to desired values, we can eliminate solutions with low GD score. This affords the following guarantee:

Proposition 4. *Let \mathbf{X} and \mathbf{C} be perfectly rational adversarial and cooperative observers respectively. Given a MO-COPP $= \langle \Lambda, \mathcal{P}, \mathcal{G}, \{\Omega_i\}_{i \in \Lambda}, \{\mathcal{O}_i\}_{i \in \Lambda}, \{\mathcal{B}_0^i\}_{i \in \{\mathbf{X}, \mathbf{C}\}} \rangle$ with equally probably goals, $|\mathcal{G}| = n$, and goal constraints of at least k goal obfuscation for \mathbf{X} and at most j goal legibility for \mathbf{C} , then a solution plan, π , gives the following guarantees:*

1. \mathbf{X} can infer $G_{\mathbf{A}}$ with probability $\leq 1/k$,
2. \mathbf{C} can infer $G_{\mathbf{A}}$ with probability $\geq 1/j$, and
3. Goal difference metric, $GD(\pi) \geq \frac{k-j}{n-1}$

Given $|\mathcal{G}_{\mathbf{X}}^{\pi}| \geq k$, \mathbf{X} can infer $G_{\mathbf{A}}$ with probability $\leq 1/k$. Similarly, given $|\mathcal{G}_{\mathbf{C}}^{\pi}| \leq j$, \mathbf{C} can infer $G_{\mathbf{A}}$ with probability $\geq 1/j$. Also, similarly, $GD(\pi) \geq \frac{k-j}{n-1}$. The above proposition states that, based on the observation equivalence there is no additional information revealed about the actor’s true goal. Therefore, we can derive goal detection upper bound for \mathbf{X} and lower bound for \mathbf{C} . Also this allows us to derive a lower bound on the plan quality.

4 Empirical Evaluation

In this section, we evaluate the performance of our approaches against a baseline and discuss the relative strengths of both the approaches using 6 domains. We also compare the quality of the solutions generated by our approach against an approach that assumes either entirely adversarial or entirely cooperative observers.

Domains For the evaluation, we look at 6 domains: Gridworld, Box Pushing, Recycling Robot, Blocksworld, Logistics and Driverlog. In Gridworld, the actor can move in cardinal directions. $\mathcal{O}_{\mathbf{X}}$, consists of two observations, *vertical* for movement along N or S and *horizontal* otherwise. $\mathcal{O}_{\mathbf{C}}$ consists of symbols, *north-east* for N or E and *south-west* otherwise. In Box Pushing [Kube and Zhang, 1997], the actor’s goal is to face the box and push it to the topmost row. The actions available are: *move-forward*, *stay*, *turn-left*, *turn-right*. $\mathcal{O}_{\mathbf{X}}$ consists of two symbols, *turn* when the agent performs turning actions and *move*, otherwise. $\mathcal{O}_{\mathbf{C}}$ consists of *move-right* when turning right or moving and *leftwards* otherwise. In Recycling Robot [Sutton and Barto, 2018], the actor’s objective is to collect cans and put them in recycling bins. The actions available are: *N, S, E, W, pick-up, drop, charge, stay*. $\mathcal{O}_{\mathbf{X}}$ consists of four symbols, *horizontal* when moving E or W, *vertical* when moving N or S, *using-gripper* when picking or dropping and *charging* otherwise. $\mathcal{O}_{\mathbf{C}}$ consists of *north-east, south-west, charging-picking, and staying-dropping*. For the three IPC domains, we use the lifted action names as observations for adversarial observer and lifted action names with objects as observations for cooperative observer. For example, given Blocksworld with 4 blocks a, b, c, d , for \mathbf{C} , we can have *stack-a-b* to represent stacking of a or b on any other block and similarly *stack-c-d*. This is done for all the lifted actions. We produced similar observations for Logistics and Driverlog. In modeling the observations, we use the notion that in realistic scenarios, it’s likely that the actor may provide the ally with tools/information to infer observations more clearly.

We implemented our IP encoding using Gurobi optimizer [Gurobi Optimization, 2018]. We implemented the heuristic-guided search using the STRIPS planner Pyperplan [Alk-hazraji *et al.*, 2016] with hsa [Keyder and Geffner, 2008] heuristic. We used hsa heuristic because it gave better performance. For baseline planner, we used greedy best first search with hsa heuristic. We ran our experiments on 3.5 GHz Intel Core i7 processor with 16 GB RAM. We used Gridworld of size 8x8, Box Pushing of size 5x5 with a single box, Recycling Robot of size 4x4 with a can and battery of 5 levels, Blocksworld with 5 blocks, and Logistics and Driverlog had goals with 4 facts each. For each domain, we generated 30 problems with random initial state goals. For the first 3 domains, we created 3 goals per problem, and for the next 3 domains, 5 goals per problem.

Relative Strengths We report the averages and standard deviation of GD , plan length and time taken for baseline, IP and search algorithm in Figure 2. For the baseline, we computed plans considering only $G_{\mathbf{A}}$. For the IP, we set the plan horizon to 12 for Gridworld, 12 for Box-Pushing and 10 for Recycling-Robot. If solutions were not found for that horizon, we incremented the horizon by 2. The IP did not run for the complex IPC domains, so we only display results for the first three domains. For the search, we set the minimum number of goals for \mathbf{X} to 2 (at least 2 goals), and maximum number of goals for \mathbf{C} to 2 (at most 2 goals). For all three algorithms, the GD is calculated by counting the total number of goals present in the observers’ beliefs. We tested for statistical significance of the results by performing independent measures ANOVA to reject the null hypothesis that the three algorithms are the same and that the differences are due to any randomness in the experiments (e.g., the randomly chosen goals). For Gridworld, we found that the p-value is less than 0.00001 for GD , as well as for plan length and run time considering the results of all 3 algorithms. This is true for Box-Pushing and for Recycling-Robot as well. All the results are significant at $p < 0.05$.

The IP approach has several advantages. Firstly, it produces optimal solutions given a time horizon for the MO-COPP problem. Secondly, it provides a lot of flexibility: it automatically chooses the best candidate goals to be added to or removed from the final beliefs of the observers. Also, if we want to specifically add or remove a particular goal from the observer’s final belief, it is easy to add the necessary constraint. These advantages were evident in the results: IP has higher GD for all the 3 domains. On the other hand, the search algorithm is faster and generates satisficing solutions that meet the goal constraints: the average time for search is consistently lower than IP for all the domains. Also the search solutions are shorter in length than those of the IP. The baseline although fastest (satisficing solution to a single goal) produces worst plan quality (GD). Additionally, the search can run more complex problems. The average and standard deviation GD for the IPC domains is reported in Figure 3.

Comparison Against Other Approaches Here, we report the GD metric for IPC domains and show the relative performance of our algorithm against approaches that assume entirely adversarial or entirely cooperative observers using IPC domains. We used k -ambiguous and j -legible algorithms from [Kulkarni *et al.*, 2019] for achieving obfuscation and legibility in isolation as baseline. We compute the GD s for

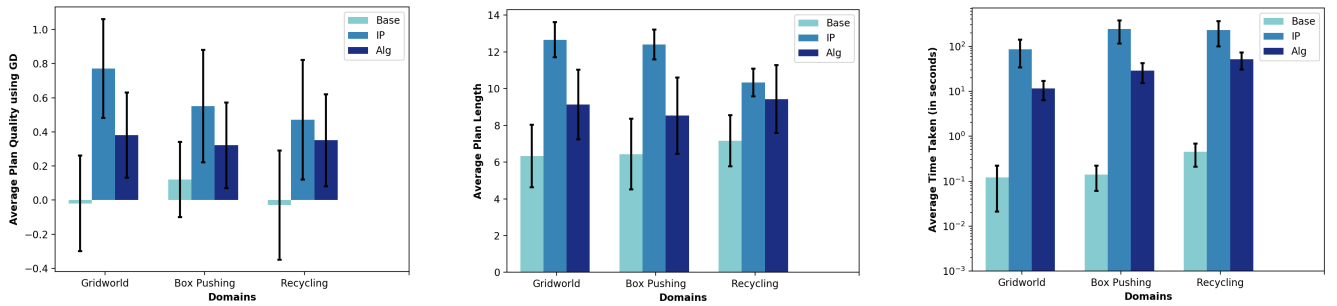


Figure 2: Comparison of average and standard deviation for goal difference (GD), plan length and run time using a baseline planner, IP planner and heuristic-guided planner over three domains.

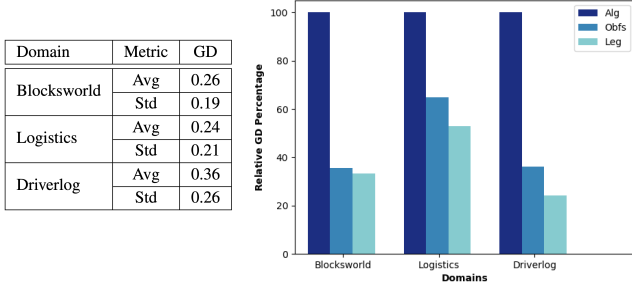


Figure 3: (a) Table shows the average and standard deviation GD for IPC domains. (b) Graph shows relative GD between our algorithm and approaches that achieve obfuscation/legibility in isolation.

the baseline goal obfuscation and goal legibility by allowing minimum constraint for the other case. That is, when running goal obfuscation algorithm, the minimum constraint is to be legible with respect to at most 5 goals. Similarly, when using goal legibility algorithm, the minimum constraint is to be obfuscating with respect to at least 1 goal. We set a timeout of 20 minutes, and altogether 7 problems timed out (3 from Logistics, 4 from Driverlog) out of 90 problems. Here we set k and j values to 3. In Figure 3, we report the relative GD percentages for solutions that achieve goal obfuscation and goal legibility in isolation with respect to GD of the solutions produced by our algorithm. The GD is computed by counting the number of goals in each observer’s belief.

From Figure 3, we can see that, our approach consistently outperforms the obfuscation and legibility algorithms with respect to the plan quality of the solutions (GD). This is because, as stated in Proposition 4, our approach makes sure that each solution achieves a minimum amount of GD . In this case, the minimum is 0 (since k and j values are set to 3). This ensures that our search algorithm does not output solutions with $GD < 0$, which was not the case for the other two approaches, as is seen from the relative GD percentages. This evaluation shows that the existing approaches that address obfuscation and legibility in isolation are not sufficient to produce good quality solutions to MO-COPP.

5 Related Work

This work is related to the plan recognition literature as the observers can use the observations to perform goal recognition. There are several prior works on goal/plan recognition [Ramirez and Geffner, 2009; Ramirez and Geffner, 2010; E-Martin *et al.*, 2015; Sohrabi *et al.*, 2016; Pereira *et al.*, 2017; Masters and Sardina, 2017a]. However, the following two

characteristics of our approach would not allow these goal recognition approaches to accurately rank goals: (1) our approach does not restrict the actor to optimal solutions (2) the observation equivalence in the many-to-one mapping of the observations does not provide additional information about $\langle a, s \rangle$ pairs. The aforementioned recognition systems assume 1-to-1 mapping of observations to actions or states. In our framework, given the observation sequence there exists a valid plan to each of the goals in observer’s final belief. This would throw off the goal recognition systems. In our framework, in the adversarial case, the set of possible goals in the observer’s final belief does not reveal information about the actor’s true goal, whereas, in the cooperative case, this set is indeed what the actor wants to convey to the observer.

Our framework not only accommodates both adversarial and cooperative agents but also tackles them simultaneously. However, several prior works have explored planning in adversarial environments [Keren *et al.*, 2016b; Keren *et al.*, 2016a; Masters and Sardina, 2017b; Shekhar and Brafman, 2018; Pozanco *et al.*, 2018; Kulkarni *et al.*, 2019] in isolation and also in cooperative environments [Keren *et al.*, 2014; Zhang *et al.*, 2017; Chakraborti *et al.*, 2017; Chakraborti *et al.*, 2016; MacNally *et al.*, 2018] in isolation. There are a few frameworks [Keren *et al.*, 2014; Kulkarni *et al.*, 2019] that are general enough to address both adversarial and cooperative settings in the same framework. However these still look at each setting in isolation in contrast to our framework. Additionally, we generalize the controlled observability planning problem [Kulkarni *et al.*, 2019] by manipulating the observations received by both adversarial and cooperative observers simultaneously. Figure 3 shows that our approaches significantly outperform Kulkarni *et al.* [2019] in settings with mixed adversarial and cooperative observers.

6 Conclusion

We present the MO-COPP formulation which is a more general framework for controlled observability planning problem. MO-COPP can tackle both adversarial and cooperative observers simultaneously. We provide two solution approaches: (1) we formulate the problem as a constraint optimization problem and show that the MO-COPP can be solved optimally given the time horizon by maximizing goal obfuscation and maximizing goal legibility simultaneously to separate observers, (2) we show that it is possible to leverage frameworks that tackle obfuscation and legibility in isolation to compute satisficing solutions to MO-COPP. We evaluate both of our approaches using 6 domains in total.

Acknowledgments

This research is supported in part by ONR grants N00014-16-1-2892, N00014-18-1-2442, and N00014-18-1-2840, AFOSR grant FA9550-18-1-0067, NASA grant NNX17AD06G and JP Morgan AI Faculty Research grant.

References

- [Alkhazraji *et al.*, 2016] Yusra Alkhazraji, Matthias Frorath, Markus Grützner, Thomas Liebetaut, Manuela Ortlieb, Jendrik Seipp, Tobias Springenberg, Philip Stahl, and Jan Wülfing. Pyperplan. <https://bitbucket.org/malte/pyperplan>, 2016.
- [Chakraborti *et al.*, 2016] Tathagata Chakraborti, Yu Zhang, David E Smith, and Subbarao Kambhampati. Planning with resource conflicts in human-robot cohabitation. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, 2016.
- [Chakraborti *et al.*, 2017] Tathagata Chakraborti, Sarath Sreedharan, Yu Zhang, and Subbarao Kambhampati. Plan explanations as model reconciliation: Moving beyond explanation as soliloquy. In *IJCAI*, 2017.
- [Dragan and Srinivasa, 2013] Anca Dragan and Siddhartha Srinivasa. Generating legible motion. In *Proceedings of Robotics: Science and Systems*, Berlin, Germany, June 2013.
- [E-Martin *et al.*, 2015] Yolanda E-Martin, Maria D R-Moreno, and David E Smith. A fast goal recognition technique based on interaction estimates. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.
- [Geffner and Bonet, 2013] Hector Geffner and Blai Bonet. A concise introduction to models and methods for automated planning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 8(1):1–141, 2013.
- [Gurobi Optimization, 2018] LLC Gurobi Optimization. Gurobi optimizer reference manual, 2018.
- [Keren *et al.*, 2014] Sarah Keren, Avigdor Gal, and Erez Karpas. Goal recognition design. In *ICAPS*, 2014.
- [Keren *et al.*, 2016a] Sarah Keren, Avigdor Gal, and Erez Karpas. Goal recognition design with non-observable actions. In *AAAI*, pages 3152–3158, 2016.
- [Keren *et al.*, 2016b] Sarah Keren, Avigdor Gal, and Erez Karpas. Privacy preserving plans in partially observable environments. In *IJCAI*, pages 3170–3176, 2016.
- [Keyder and Geffner, 2008] Emil Keyder and Héctor Geffner. Heuristics for planning with action costs revisited. In *ECAI*, pages 588–592, 2008.
- [Kube and Zhang, 1997] C Ronald Kube and Hong Zhang. Task modelling in collective robotics. *Autonomous Robots*, 4(1):53–72, 1997.
- [Kulkarni *et al.*, 2019] Anagha Kulkarni, Siddharth Srivastava, and Subbarao Kambhampati. A unified framework for planning in adversarial and cooperative environments. In *AAAI*, 2019.
- [MacNally *et al.*, 2018] Aleck M MacNally, Nir Lipovetzky, Miquel Ramirez, and Adrian R Pearce. Action selection for transparent planning. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pages 1327–1335. International Foundation for Autonomous Agents and Multiagent Systems, 2018.
- [Masters and Sardina, 2017a] Peta Masters and Sebastian Sardina. Cost-based goal recognition for path-planning. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, AAMAS ’17, pages 750–758, Richland, SC, 2017. International Foundation for Autonomous Agents and Multiagent Systems.
- [Masters and Sardina, 2017b] Peta Masters and Sebastian Sardina. Deceptive path-planning. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, *IJCAI-17*, pages 4368–4375, 2017.
- [Pereira *et al.*, 2017] Ramon Fraga Pereira, Nir Oren, and Felipe Meneguzzi. Landmark-based heuristics for goal recognition. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
- [Pozanco *et al.*, 2018] Alberto Pozanco, Yolanda E-Martín, Susana Fernández, and Daniel Borrajo. Counterplanning using goal recognition and landmarks. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, *IJCAI-18*, pages 4808–4814. International Joint Conferences on Artificial Intelligence Organization, 7 2018.
- [Ramirez and Geffner, 2009] Miquel Ramirez and Hector Geffner. Plan recognition as planning. In *Proceedings of the 21st international joint conference on Artificial intelligence*. Morgan Kaufmann Publishers Inc, pages 1778–1783, 2009.
- [Ramirez and Geffner, 2010] Miquel Ramirez and Hector Geffner. Probabilistic plan recognition using off-the-shelf classical planners. In *Proceedings of the Conference of the Association for the Advancement of Artificial Intelligence (AAAI 2010)*, 2010.
- [Shekhar and Brafman, 2018] Shashank Shekhar and Ronen I Brafman. Representing and planning with interacting actions and privacy. In *Twenty-Eighth International Conference on Automated Planning and Scheduling*, 2018.
- [Sohrabi *et al.*, 2016] Shirin Sohrabi, Anton V Riabov, and Octavian Udrea. Plan recognition as planning revisited. In *IJCAI*, pages 3258–3264, 2016.
- [Sutton and Barto, 2018] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.
- [Zhang *et al.*, 2017] Yu Zhang, Sarath Sreedharan, Anagha Kulkarni, Tathagata Chakraborti, Hankz H Zhuo, and Subbarao Kambhampati. Plan explicability and predictability for robot task planning. In *International Conference on Robotics and Automation (ICRA)*, 2017.