The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)

Modeling Probabilistic Commitments for Maintenance Is Inherently Harder than for Achievement

Qi Zhang, Edmund H. Durfee, Satinder Singh
Computer Science and Engineering, University of Michigan
{qizhg, durfee, baveja}@umich.edu

Abstract

Most research on probabilistic commitments focuses on commitments to achieve enabling preconditions for other agents. Our work reveals that probabilistic commitments to instead maintain preconditions for others are surprisingly harder to use well than their achievement counterparts, despite strong semantic similarities. We isolate the key difference as being not in how the commitment provider is constrained, but rather in how the commitment recipient can locally use the commitment specification to approximately model the provider's effects on the preconditions of interest. Our theoretic analyses show that we can more tightly bound the potential suboptimality due to approximate modeling for achievement than for maintenance commitments. We empirically evaluate alternative approximate modeling strategies, confirming that probabilistic maintenance commitments are qualitatively more challenging for the recipient to model well, and indicating the need for more detailed specifications that can sacrifice some of the agents' autonomy.

Introduction

In multiagent systems, agents are often interdependent in that what one agent does can help or hinder another. In a cooperative system, agents can mutually benefit from helping each other. Specifically, we focus on interdependency where an agent (the commitment *provider*) makes a social commitment (Singh 1999; Kalia, Zhang, and Singh 2014) to another (the commitment *recipient*). In essence, a commitment abstracts the effect that the provider's behavior has on the recipient's local environment, simplifying the coordination between the agents. When stochasticity is inherent in the environment, the provider cannot guarantee to bring about the outcomes the recipient wants, and in fact could discover after committing that its plan to pursue the outcomes is more costly or risky than it had previously realized. Under such circumstances, commitments are conditional (Singh 2008). And when the conditions are expensive for the provider to enumerate and/or the recipient to observe, the likelihood of their holding can be summarized numerically, leading to a *probabilistic* commitment.

Prior work has focused on semantics and mechanisms for the provider to follow to faithfully pursue its commitments despite uncertainty (Jennings 1993; Xing and Singh 2001; Winikoff 2006; Durfee and Singh 2016). The previous work held that a probabilistic commitment should be considered fulfilled if the provider's actions would have brought about the desired outcome by the promised time with at least the promised probability, even if in a particular instance the desired outcome was not realized. In this vein, the focus was largely on the provider's pursuit of *achievement* commitments (Xuan and Lesser 1999; Maheswaran et al. 2008; Witwicki and Durfee 2009; Zhang et al. 2016), where the provider commits to changing some features of the state in a way desired by the recipient with some probability by some time. For example, the recipient plans to take an action (e.g., move from one room to another) with a precondition (e.g., the door separating rooms is open) that the provider has promised to likely enable by some deadline.

This paper also considers another form of commitment, a *maintenance* commitment, where the provider instead commits to courses of action that, up until a promised time, are sufficiently unlikely to change features that are already the way the recipient wants them maintained. After that time, the provider can freely change the features. For example, a door the recipient wants open might initially be so, but the provider wants to close it to clean behind it during housekeeping tasks. The provider could postpone closing it (clean elsewhere first), but by changing other doors while cleaning elsewhere it might accidentally introduce a draft that could prematurely close the door the recipient wants left open.

Even though decision-theoretic formulations of, and reasoning methods for, achievement and maintenance commitments are nearly identical, prior work has found it much harder to successfully coordinate for maintenance than achievement (Clement and Schaffer 2008; Goldman et al. 2008; Hiatt 2009). In the past, it has been assumed that the difficulty lies on the provider's side—that it might be inherently harder for a provider to find good policies that maintain a feature than to change it. However, in this paper we claim (and justify) that instead the challenge actually lies on the recipient's side: that *a maintenance commitment is fundamentally harder for the recipient to model safely than an*

achievement commitment is.

We substantiate this claim theoretically and empirically. We begin by analyzing a straightforward strategy, adopted in previous work, where the recipient models an achievement commitment pessimistically by assuming the feature will not (probabilistically) attain its desired value any earlier than the commitment's promised time. We show analytically that the worst-case suboptimality induced by such pessimism can be bounded fairly tightly. For the maintenance counterpart, however, we show that no comparable pessimistic model, and hence no bound on suboptimality, exists. We also empirically measure suboptimality for several alternative modeling strategies, and the results show that there is no model the recipient can adopt for maintenance commitments that safely limits the suboptimality of coordination with the provider. Our results suggest that successful maintenance commitments will generally require that the provider's and recipient's plans need to be more tightly coupled than for achievement commitments.

Related Work. The literature on protocols for the commitment lifecycle focuses on (awareness of) the progression of agents' joint commitment's status, including whether some have been abandoned to pursue more valuable goals (Desai, Narendra, and Singh 2008; Günay, Liu, and Zhang 2016; Pereira, Oren, and Meneguzzi 2017). We focus just on the "detached" stage where an agreed-upon commitment is being actively pursued, and the pursuit requires a sequence of actions, where some might not have desired outcomes, or an agent's priorities could change in the midst of executing the sequence.

We adopt the probabilistic commitment framework (Xuan and Lesser 1999; Witwicki and Durfee 2007; Bannazadeh and Leon-Garcia 2010) that summarizes the likelihood the commitment will be successfully discharged by a given time, versus violated due to bad luck or a better option appearing. Probabilities let a decision-theoretic recipient optimally hedge for violations while waiting for the provider. Others have adopted alternative frameworks, such as conditional commitments (Singh 2012; Vokrínek, Komenda, and Pechoucek 2009) and contracting frameworks (Sandholm and Lesser 2001), for managing the uncertainty when the commitment is being pursued.

Preliminaries

In this section, we describe the decision-theoretic setting we adopt for analyzing probabilistic commitments for the recipient and the provider, including both achievement commitments and maintenance commitments.

The **recipient**'s environment is modeled as a Markov Decision Process (MDP) defined by the tuple $M = (\mathcal{S}, \mathcal{A}, P, R, H, s_0)$ where $\mathcal{S}$ is the finite state space, $\mathcal{A}$ is the finite action space, $P : \mathcal{S} \times \mathcal{A} \to \Delta(\mathcal{S})$ ($\Delta(\mathcal{S})$ denotes the set of all probability distributions over $\mathcal{S}$) is the transition function, $R : \mathcal{S} \to \mathbb{R}$ is the reward function, H is the finite horizon, and s_0 is the initial state. The state space is partitioned into disjoint sets by the time step, $\mathcal{S} = \bigcup_{h=0}^{H} \mathcal{S}_h$, where states in $\mathcal{S}_h$ only transition to states in $\mathcal{S}_{h+1}$. The MDP starts in s_0 and terminates in $\mathcal{S}_H$. Given a policy $\pi : \mathcal{S} \to \mathcal{A}$ and starting in the initial state, a random sequence of transitions $\{(s_h, a_h, r_{h+1}, s_{h+1})\}_{h=0}^{H-1}$ is generated by $a_h = \pi(s_h), s_{h+1} \sim P(s_h, a_h), r_{h+1} = R(s_{h+1})$. The value function of π is $V_M^\pi(s) = \mathbb{E}[\sum_{h'=h+1}^{H} r_{h'} | \pi, s_h = s]$ where h is such that $s \in \mathcal{S}_h$. The optimal policy for M, denoted as π_M^*, maximizes V_M^π for all $s \in \mathcal{S}$, and its value function $V_M^{\pi_M^*}$ is abbreviated as V_M^*. The value of the initial state is abbreviated as $v_M^\pi := V_M^\pi(s_0)$.

Similarly, the **provider**'s environment is modeled as another MDP with a finite state space, a finite action space, and a finite horizon. As one way to model the interaction between the provider and the recipient (Witwicki and Durfee 2010; Zhang et al. 2016), we assume that both the recipient's state and the provider's state can be factored into state features. The recipient's state is factored as $s = (l, u)$, where l is the set of all the recipient's state features locally controlled by the recipient, and u is the set of the state features shared with the provider. The provider's state features, including u, are all locally controlled by the provider. The provider and the recipient are weakly coupled in the sense that the shared state features u are only controllable by the provider. Formally, the dynamics of the recipient's state can be factored as

$$\begin{aligned} P(s_{h+1}|s_h, a_h) =& P\left((l_{h+1}, u_{h+1})|(l_h, u_h), a_h\right) \\ =& P_u(u_{h+1}|u_h) P_l\left(l_{h+1}|(l_h, u_h), a_h\right). \end{aligned}$$

We refer to P_u as the true *influence* that the provider exerts on the recipient's environment dynamics (Witwicki and Durfee 2010; Oliehoek, Witwicki, and Kaelbling 2012; Oliehoek, Spaan, and Witwicki 2015), which is the transition function of u that is fully determined by the provider's policy (it is not a function of a_h). We assume that the recipient's reward function is dependent on l but not on u, $R(s_h) = R((l_h, u_h)) = R(l_h)$, such that the cumulative reward of an episode is determined by the trajectory of l, $(l_1, ..., l_H)$. Note that though the value of u_h does not directly affect the reward for time step h, it can enable action choices that affect the value of l_{h+1} at the next time step.

Commitment Semantics

A commitment is concerned with state features u that are shared by both agents but only controllable by the provider. Intuitively, a commitment provides partial information about P_u from which the recipient can plan accordingly. In this paper, we focus on the setting where u contains a single state feature that takes binary value, letting u^+, as opposed to u^-, be the value of u that is desirable for the recipient. Intuitively, $u^+(u^-)$ stands for an enabled (disabled) precondition needed by the recipient. We will refer to u as the commitment feature. Further, we assume that u can be toggled at most once (Hindriks and van Riemsdijk 2007; Witwicki and Durfee 2009; Zhang et al. 2016). In transactional settings, a feature (e.g., possession of goods) changing only once is common, as it is in multiagent planning domains where one agent enables a precondition needed by an action of another. Some cooperative agent work requires agents to return changed features to prior values (e.g., shutting the door after opening and passing through it), and in

extreme cases where toggling reliably repeats (e.g., a traffic light) there may be no need for explicit commitments. While, in general, toggling more than once can be modeled by a series of alternating achievement and maintenance commitments, the fundamental differences between these commitment types are most readily revealed and understood without such complications, and so in what follows we consider the two types separately.

Achievement Commitments. Let the initial state be factored as $s_0 = (l_0, u_0)$. For achievement commitments, the initial value of the commitment feature is u^-, i.e. $u_0 = u^-$. The provider commits to pursuing a course of action that can bring about the commitment feature desirable to the recipient with some minimum probability. Formally, an achievement commitment is defined by tuple $c_a = (T_a, p_a)$, where T_a is the achievement commitment time, and p_a is the achievement commitment probability (Witwicki and Durfee 2009; Zhang et al. 2016). The commitment semantics is that the provider is to follow a policy that sets u to u^+ by time step T_a with *at least* probability p_a, i.e.

$$\Pr(u_{T_a} = u^+ | u_0 = u^-) \geq p_a. \quad (1)$$

When planning with the achievement commitment, the provider finds an optimal policy (one that maximizes its local value) that respects the commitment's semantics. A straightforward way of doing so adopted in prior work solves the provider's planning problem using linear programming (LP) (Altman 1999), where the commitment semantics are captured simply by adding the above inequality as an additional constraint to the LP (Witwicki and Durfee 2007; Steinmetz, Hoffmann, and Buffet 2016).

Maintenance Commitments. As a reminder, a maintenance commitment is appropriate in scenarios where the initial value of state feature u is desirable to the recipient, who wants it to maintain its initial value for some interval of time (e.g., (Hindriks and van Riemsdijk 2007; Duff, Thangarajah, and Harland 2014)), but where the provider might want to take actions that could change it. Formally, a maintenance commitment is defined by tuple $c_m = (T_m, p_m)$, where T_m is the maintenance commitment time, and p_m is the maintenance commitment probability. Given such a maintenance commitment, the provider is constrained to follow a policy that keeps u unchanged for the first T_m time steps with *at least* probability p_m. Since u can be toggled at most once, this is equivalent to probabilistically guaranteeing that u is still u^+ at the commitment time T_m, i.e.

$$\Pr(u_{T_m} = u_0 | u_0 = u^+) \geq p_m. \quad (2)$$

As with an achievement commitment, the provider with a maintenance commitment finds a policy that optimizes its local value while respecting the commitment semantics, again by including the commitment constraint in its LP. Hence, from the provider's perspective, achievement and maintenance commitments are treated essentially identically.

The Approximate Influence

The similarity in how achievement and maintenance commitments can be captured in the provider's reasoning, combined with the intuition that the provider's reasoning is what is challenging with commitments (since the provider is constrained by the commitment, while the recipient is not), suggests that coordination using the two types of commitments can be done similarly with similar effectiveness. But experience indicates otherwise (Clement and Schaffer 2008; Goldman et al. 2008; Hiatt 2009). We now explain how this is because achievement and maintenance commitments differ fundamentally from the *recipient's* perspective.

As we have seen, the commitment specification and semantics constrain the provider's policy based on a single future timestep: at that timestep, the value of u will (still) be u^+ with at least the promised probability. By not committing to the probabilities at intervening (and subsequent) timesteps, the provider retains flexibility to revise its policy on the fly (for example, if its reward function changes because of a new goal). Our prior work has shown the value to the provider of having such flexibility (Zhang et al. 2016; Zhang, Singh, and Durfee 2017).

The commitment specification is also the *only* information that the recipient has about P_u, and while information about only a single future timestep might give the provider flexibility, it imposes uncertainty on the recipient. That is, while the recipient knows something about P_u at the commitment's timestep (that the probability of u^+ is at least the given value), and that the probability changes monotonically (due to u toggling at most once), it can only guess at the values of influence at other timesteps. We notate the approximate influence that it uses for its planning as $\widehat{P}_u$.

We are specifically interested in the quality of the recipient's plan computed from approximate influence $\widehat{P}_u$ when evaluated in (true) influence P_u. Formally, given $\widehat{P}_u$, let $\widehat{M} = (\mathcal{S}, \mathcal{A}, \widehat{P}, R, H, s_0)$ be the approximate model that only differs from M in terms of the dynamics of u, i.e. $\widehat{P} = (P_l, \widehat{P}_u)$. The quality of $\widehat{P}_u$ is evaluated using the difference between the value of the optimal policy for $\widehat{M}$ and the value of the optimal policy for M when both policies are evaluated in M starting in s_0, i.e.

$$\textit{Suboptimality} : v^*_M - v^{\pi^*_{\widehat{M}}}_M.$$

Note that when the support of P_u is not fully contained in the support of $\widehat{P}_u$, the recipient's policy $\pi^*_{\widehat{M}}$ can associate zero occupancy (hence plan no action) for certain states when executed in M, which makes $V^{\pi^*_{\widehat{M}}}_M$ ill-defined. In this paper, we resolve this by re-planning: during execution of $\pi^*_{\widehat{M}}$ in M, the recipient re-plans from any zero occupancy state that it happens to reach.

Previous work chooses an intuitive and straightforward approximate influence for achievement commitments that models a single branch, *at the commitment time*, for when u^- probabilistically toggles to u^+ (Witwicki and Durfee 2010; Zhang et al. 2016). Modelling the commitment with a single branch for toggling to u^+ at the latest possible time ignores possibilities of being enabled earlier than the deadline and of being enabled serendipitously after the deadline. Such an approximate influence models the achievement commitment *pessimistically*, in the sense that it minimizes the ex-

pected duration of u being enabled over all influences that respect the achievement commitment semantics (Eq. (1)):

$$\min_{P_u \sim (1)} \mathbb{E}_{P_u}\left[\sum_{h=0}^{H} 1_{\{u_h = u^+\}}\right]$$

where $P_u \sim$ (1) means influence P_u satisfies Eq. (1), and 1_E is the indicator function that takes value one if event E occurs and zero otherwise. We refer to this model as the *minimal enable duration* influence, as formalized in Definition 1.

Definition 1. Given achievement commitment $c_a = (T_a, p_a)$, its minimal enable duration influence $\widehat{P}^{\min+}_{u,c_a}$ toggles u in the transition from time step $h = T_a - 1$ to $h = T_a$ with probability p_a, and does not toggle u at any other time step.

For maintenance commitments, the counterpart minimizes the expected enablement duration over all influences that respect the maintenance commitment semantics (Eq. (2)):

$$\min_{P_u \sim (2)} \mathbb{E}_{P_u}\left[\sum_{h=0}^{H} 1_{\{u_h = u^+\}}\right].$$

The minimizer models a probabilistic toggling to u^- at the earliest possible time, and a deterministic toggling to u^- (if it had not toggled earlier) after the commitment time, as formalized in Definition 2.

Definition 2. Given maintenance commitment $c_m = (T_m, p_m)$, its minimal enable duration influence $\widehat{P}^{\min+}_{u,c_m}$ toggles u in the transition from time step $h = 0$ to $h = 1$ with probability $1 - p_m$, and (unless already toggled) from $h = T_m$ to $h = T_m + 1$ with probability one. It does not toggle u at any other time step.

Theoretical Analysis

In this section, we derive bounds on the suboptimality of the minimal enable duration influence. Our analyses make the following two assumptions. Assumption 1 intuitively says that u^+ establishes a precondition for an action that would be irrational to take when u^- holds. For example, if u^+ is a door being open, then the action of moving into the doorway could be part of an optimal plan, but taking that action if the door is closed (u^-) never is. Assumption 2 is a simplifying assumption for our analyses stating the true influence agrees with the minimal enable duration influence after the commitment time, so that any suboptimality is caused by the imperfect modeling up until the commitment time.

Assumption 1. Let $s^- = (l, u^-)$ and $s^+ = (l, u^+)$ be a pair of states that only differ in u. For any M with arbitrary influence P_u, we have

$$P_l\left(\cdot|s^-, \pi^*_M(s^-)\right) = P_l\left(\cdot|s^+, \pi^*_M(s^-)\right).$$

Assumption 2. $P_u(u_{h+1}|u_h)$ agrees with the minimal enable duration influence for $h \geq T$, where T is the commitment time.

To derive bounds on achievement and maintenance commitments, we will make use of the following lemma, where M^+ (M^-) is defined as the recipient's MDP identical to M except that u is always set to $u^+(u^-)$. Lemma 1 directly follows from Assumption 1, stating that the value of M^- is no more than that of M^+ and the value of any M is between the two.

Lemma 1. For any M with arbitrary influence P_u and initial value of u, we have $v^*_{M^-} \leq v^*_M \leq v^*_{M^+}$.

Proof. Let's first consider the case in which P_u toggles u only at a single time step. We show $v^*_{M^-} \leq v^*_M$ by constructing a policy in M for which the value is $v^*_{M^-}$ by mimicking $\pi^*_{M^-}$. Whether u is initially u^- and later toggled to u^+ or *vice versa*, we can construct a policy π_M that chooses the same actions as $\pi^*_{M^-}$ assuming $u = u^-$ throughout the episode. Formally, for any $s^- = (l, u^-)$, letting $s^+ = (l, u^+)$,

$$\pi_M(s^+) = \pi_M(s^-) = \pi^*_{M^-}(s^-).$$

By Assumption 1, π_M in M yields the same distribution over the trajectory of l as $\pi^*_{M^-}$ in M^-, and therefore $v^{\pi_M}_M = v^*_{M^-}$ since the cumulative reward only depends on the trajectory of l.

Similarly, we show $v^*_M \leq v^*_{M^+}$ by constructing a policy π_{M^+} in M^+ for which the value is v^*_M by mimicking π^*_M. Formally, for time steps when $u = u^-$ in M, let $\pi_{M^+}(s^+) = \pi^*_M(s^-)$. For time steps when $u = u^+$ in M, let $\pi_{M^+}(s^+) = \pi^*_M(s^+)$, where $s^- = (l, u^-)$, $s^+ = (l, u^+)$.

When P_u toggles u at $K > 1$ time steps, we can decompose the value function for P_u as the weighted average of K value functions corresponding to the K influences that toggle u at a single time step, and the weights of the average are the toggling probabilities of P_u at these K time steps. □

Bounding Suboptimality for Achievement

Here, we derive Theorem 1 that bounds the suboptimality for achievement commitments as the difference between $v^*_{M^-}$ and $v^*_{M^+}$. We use Assumptions 1 and 2, and Lemma 2 which states that, for achievement commitments, the possible ways the true influence differs from the minimal enable duration influence can only improve the expected value.

Lemma 2. Given achievement commitment $c_a = (T_a, p_a)$, let $\widehat{P}_u = \widehat{P}^{\min+}_{u,c_a}$, then we have $v^{\pi^*_{\widehat{M}}}_M \geq v^{\pi^*_{\widehat{M}}}_{\widehat{M}}$ where influence P_u in M respects the commitment semantics of c_a.

Proof. For achievement commitments, the initial value of u is u^-. Let $P_u(t)$ be the probability that u is enabled to u^+ at t in influence P_u, and $\overline{v}^{\pi}_t$ be the initial state's value under π when u is enabled from u^- to u^+ at t with probability one. By Assumption 2, $v^{\pi^*_{\widehat{M}}}_M$ and $v^{\pi^*_{\widehat{M}}}_{\widehat{M}}$ can be decomposed as

$$v^{\pi^*_{\widehat{M}}}_M = \sum_{t=1}^{T_a} P_u(t)\overline{v}^{\pi^*_{\widehat{M}}}_t + (1 - p_a)v^{\pi^*_{\widehat{M}}}_{M^-},$$
$$v^{\pi^*_{\widehat{M}}}_{\widehat{M}} = p_a\overline{v}^{\pi^*_{\widehat{M}}}_{T_a} + (1 - p_a)v^{\pi^*_{\widehat{M}}}_{M^-}.$$

When u is enabled at t in M, $\pi^*_{\widehat{M}}$ can be executed as if u is not enabled, by Assumption 1, yielding identical trajectory distribution of l (therefore value) as in $\widehat{M}$. Therefore, the

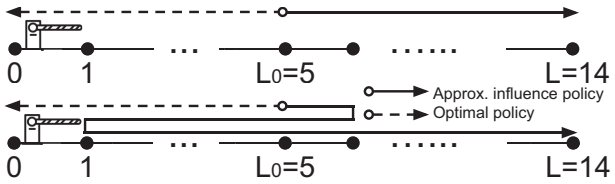


Figure 1: 1D Walk. *Top:* Example in the proof of Theorem 1. *Bottom:* Example in the proof of Theorem 2.

recipient's re-planning at t when $u = u^+$ will derive a better policy if possible. Therefore, the value of executing $\pi^*_{\widehat{M}}$ in M is no less than that in $\widehat{M}$, i.e. $\overline{v}_t^{\pi^*_{\widehat{M}}} \geq \overline{v}_{T_a}^{\pi^*_{\widehat{M}}}$. Therefore,

$$\begin{aligned} v_M^{\pi^*_{\widehat{M}}} &= \textstyle\sum_{t=1}^{T_a} P_u(t)\overline{v}_t^{\pi^*_{\widehat{M}}} + (1-p_a)v_{M^-}^{\pi^*_{\widehat{M}}} \\ &\geq \textstyle\sum_{t=1}^{T_a} P_u(t)\overline{v}_{T_a}^{\pi^*_{\widehat{M}}} + (1-p_a)v_{M^-}^{\pi^*_{\widehat{M}}} \\ &\geq p_a\overline{v}_{T_a}^{\pi^*_{\widehat{M}}} + (1-p_a)v_{M^-}^{\pi^*_{\widehat{M}}} \quad \text{(commitment semantics)} \\ &= v_{\widehat{M}}^{\pi^*_{\widehat{M}}}. \end{aligned}$$

□

Theorem 1. Given achievement commitment c_a, let $\widehat{P}_u = \widehat{P}^{\min+}_{u,c_a}$. The suboptimality can be bounded as

$$v_M^* - v_M^{\pi^*_{\widehat{M}}} \leq v_{M^+}^* - v_{M^-}^* \tag{3}$$

where influence P_u in M respects the commitment semantics of c_a. Further, there exists an achievement commitment for which the equality is attained.

Proof. The derivation of the bound in Eq. (3) is straightforward from Lemma 2:

$$v_M^* - v_M^{\pi^*_{\widehat{M}}} \leq v_{M^+}^* - v_{\widehat{M}}^{\pi^*_{\widehat{M}}} \leq v_{M^+}^* - v_{M^-}^*.$$

Next, we use a simple illustrative example to give an achievement commitment for which the equality is attained.

Example: An Achievement Commitment in 1D Walk. Consider the example of a 1D walk on $[0, L]$, as shown in Figure 1(top), where the recipient starts at L_0 and can move right, left, or stay still. There is a gate between 0 and 1 for which u^+ denotes the state of open and u^- closed. The provider toggles the gate stochastically according to P_u. For each time step the recipient stays at either end, it gets a reward of $+1$. Hence, the optimal policy is to reach either end as soon as possible in expectation. We assume $1 \leq L_0 < L/2$ to avoid the uninteresting case of $v_{M^-}^* = v_{M^+}^*$. A negative reward of -1 is incurred when bumping into the closed gate, which makes Assumption 1 hold.

Here, we derive an achievement commitment for which the bound in Theorem 1 is attained. Consider $L = 14, L_0 = 5, H = 15$, achievement commitment ($T_a = L - L_0 = 9, p_a = 1$), and the true influence P_u in M that toggles the gate to open at $t = 4$ with probability $p_a = 1$. The optimal policy in M is to move left to 0. Therefore, $v_M^* = v_{M^+}^* = H - L_0 = 10$. Given the minimal enable duration influence, moving right to L (arriving at time 9) is faster than waiting for the gate to toggle at $T_a = 9$ and then reaching location 0 at time 10. Had the recipient known the gate would toggle at time 4, it would have moved left, but by the time it toggles at time 4 the recipient is at location 9, and continuing on to L is the faster choice. Therefore $v_M^{\pi^*_{\widehat{M}}} = v_{M^-}^* = H - (L - L_0) = 6$, and the bound in Theorem 1 is attained. □

Bounding Suboptimality for Maintenance

We next ask if the bound in Eq. (3) on suboptimality in achievement commitments also holds for maintenance commitments. Unfortunately, as stated in Theorem 2, the optimal policy of the minimal enable duration influence for maintenance commitments can be arbitrarily bad when evaluated in the true influence, incurring a suboptimality exceeding the bound in Eq. (3). We give an example for an existence proof.

Theorem 2. Consider $\widehat{P}_u = \widehat{P}^{\min+}_{u,c_m}$ to be the approximate influence when modelling the maintenance commitment in $\widehat{M}$. There exists an MDP M and a maintenance commitment c_m, such that the true influence P_u in M respects the commitment semantics of c_m, $v_M^* = v_{M^+}^*$, $v_{M^-}^* > v_M^{\pi^*_{\widehat{M}}} = 0$, and therefore the suboptimality

$$v_M^* - v_M^{\pi^*_{\widehat{M}}} = v_{M^+}^* \tag{4}$$

exceeds the bound in Eq. (3).

Proof. As an existence proof, we give an example of a maintenance commitment in 1D Walk for which $v_M^* = v_{M^+}^*$ and $v_{M^-}^* > v_M^{\pi^*_{\widehat{M}}} = 0$.

Consider 1D Walk with the same $L = 14, L_0 = 5, H = 15$ as in the example for Theorem 1. Consider maintenance commitment ($T_m = 7, p_m = 0$), and P_u toggles the gate to closed at $t = 6$ with probability $1 - p_m = 1$. As shown in Figure 1(bottom), the optimal policy should take 5 steps to move directly to 0, for which the value is $v_M^* = v_{M^+}^*$. We have computed for Theorem 1 that $v_{M^-}^* = 6$. With probability $1 - p_m = 1$, the gate is closed at $t = 6$, and $\pi^*_{\widehat{M}}$ takes $19 > H$ steps to reach $L = 14$. Thus, $v_{M^-}^* > v_M^{\pi^*_{\widehat{M}}} = 0$. □

In the example used in the existence proof above, the maximum suboptimality is incurred with maintenance commitment probability $p_m = 0$ (a no-guarantee commitment), because this is when the recipient is most uncertain about the influence and will be most negatively affected by the uncertainty. Note that for achievement, a no-guarantee commitment still falls within the Theorem 1 bound.

Comparing the bound Eq. (3) in Theorem 1 with the bound Eq. (4) in Theorem 2 reveals a fundamental difference between achievement and maintenance commitments: maintenance commitments are inherently less tolerant to an unexpected change in the commitment feature. For achievement commitments, the easily-constructed minimal enable duration influence has the property of being pessimistic, in that any unexpected changes to the feature, if they impact the recipient at all, can only improve the expected value. Thus, if despite its minimal enable duration influence approximation, a recipient has chosen to follow a policy that

exploits the commitment, it can never experience a true influence that would lead it to regret having done so. The same cannot be said for maintenance commitments. There, the easily-constructed minimal enable duration influence is *not* pessimistic—it does not guarantee that any deviations from the influence can only improve the expected value. As our theoretical results show, the minimal enable duration influence assuming toggling from u^+ to u^- right away can still lead to negative surprises, since if the toggling does not immediately occur the influence suggests that it is safe to assume no toggling until T_m, but that is not true since toggling could happen sooner, after the recipient has incurred cost for a policy that would need to be abandoned. In the example for Theorem 2, the worst time for toggling to u^- is not right away, but right before the precondition would be used (the gate shutting just as the recipient was about to pass through).

Empirical Results

Our analyses suggest the minimal enable duration influence might not be the best approximate influence for a recipient to adopt for maintenance commitments. In this section, we identify several alternative heuristics to create approximate influences for the recipient, and evaluate them for both maintenance and achievement commitments. We conduct our evaluations in two domains: the same 1D Walk domain as in our theoretical analysis, and a Gate Control problem with a more interesting influence (violating Assumption 2).

1D Walk

As previously defined, the 1D Walk domain restricts the set of influences to toggle u only at a single time step no later than the commitment time, and agree with Assumption 2 thereafter. We denote the set of such influences as $\mathcal{P}_u^1$ from which $P_u, \widehat{P}_u$ are chosen. Besides using the minimal enable duration influence to approximate the true influence, we consider the following three heuristics for generating approximate influence $\widehat{P}_u \in \mathcal{P}_u^1$:

Maximal enable duration. As opposed to the minimal enable duration influence, the maximal enable duration influence optimistically toggles u right after the initial time step for achievement commitments, and at the commitment time for maintenance commitments.

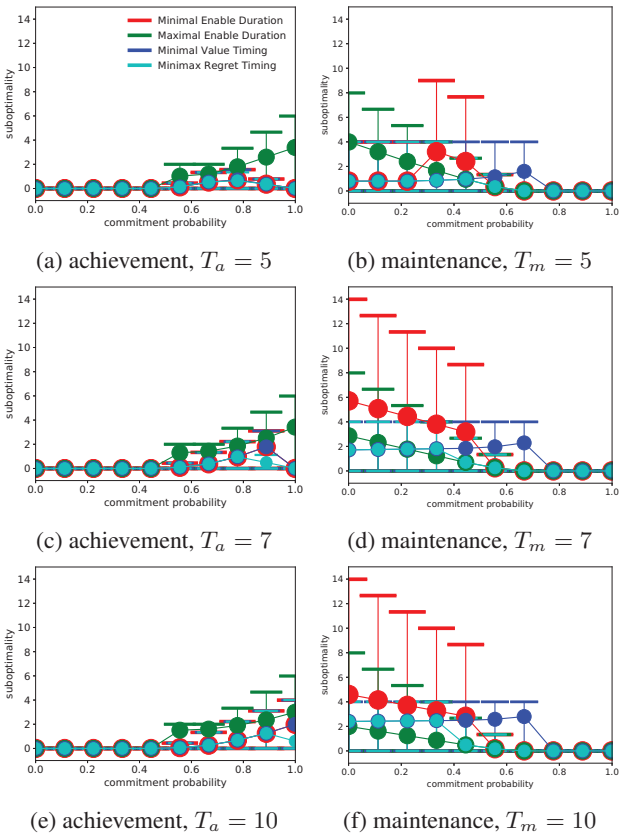
Minimal value timing. The toggling time minimizes the optimal value over all possible influences in $\mathcal{P}_u^1$, i.e. $\arg\min_{\widehat{P}_u \in \mathcal{P}_u^1} v^*_{\widehat{M}}$ where $\widehat{P}_u$ is the influence in $\widehat{M}$.

Minimax regret timing. The toggling time is chosen based on the minimax regret principle. Formally,

$$\arg\min_{\widehat{P}_u \in \mathcal{P}_u^1} \max_{P_u \in \mathcal{P}_u^1} v^*_M - v^{\pi^*_{\widehat{M}}}_M$$

where $P_u, \widehat{P}_u$ are the influences in $M, \widehat{M}$, respectively.

The four heuristics include two simple, computationally inexpensive heuristics (minimal and maximal enable duration), and two more complex and expensive heuristics (minimal value and minimax regret timing). Recall that our theoretical analysis suggests, for maintenance, the pessimistic time for toggling to u^- is not right away, but right before the recipient uses the precondition, and this causes the poor performance of the minimal enable duration influence. We hypothesize that the latter two heuristics can be more pessimistic (and thus better) than the minimal enable duration influence by identifying the worst toggling time.



(a) achievement, $T_a = 5$ (b) maintenance, $T_m = 5$

(c) achievement, $T_a = 7$ (d) maintenance, $T_m = 7$

(e) achievement, $T_a = 10$ (f) maintenance, $T_m = 10$

Figure 2: Suboptimality in 1D Walk. Markers on the curves show the mean suboptimality over possible time steps of toggling, $P_u \in \mathcal{P}_u^1$. Bars show the minimum and maximum.

Results. Here we evaluate the suboptimality of our candidate heuristics for both achievement commitments and maintenance commitments. The setting is the same as the example for Theorem 1 except that the horizon is longer, $L = 14, L_0 = 5, H = 30$. Figure 2 shows the mean, minimum, and maximum suboptimality over all realizations of $P_u \in \mathcal{P}_u^1$ for commitment time $T_a, T_m \in \{5, 7, 10\}$. We see that for achievement commitments, the suboptimality of the minimal enable duration influence is comparable to the two more sophisticated strategies, and the maximal enable duration influence incurs the most suboptimality overall. For maintenance commitments, however, the two expensive strategies incur overall less suboptimality than the minimal and the maximal enable duration, yet it is difficult to identify a single best heuristic that reliably reduces the suboptimality for all the maintenance commitments.

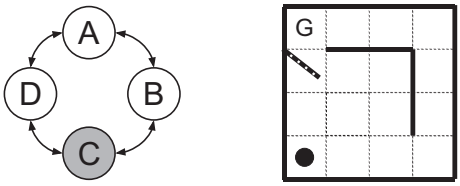


Figure 3: Gate Control. *Left:* The provider. Cell C toggles the gate. *Right:* The recipient.

Gate Control

In this domain, we are concerned with the more general situation in which $P_u \notin \mathcal{P}_u^1$ can toggle u at more than one time step by the commitment time, and even can toggle u after the commitment time. We also consider approximate influences $\widehat{P}_u$ that are not elements of $\mathcal{P}_u^1$.

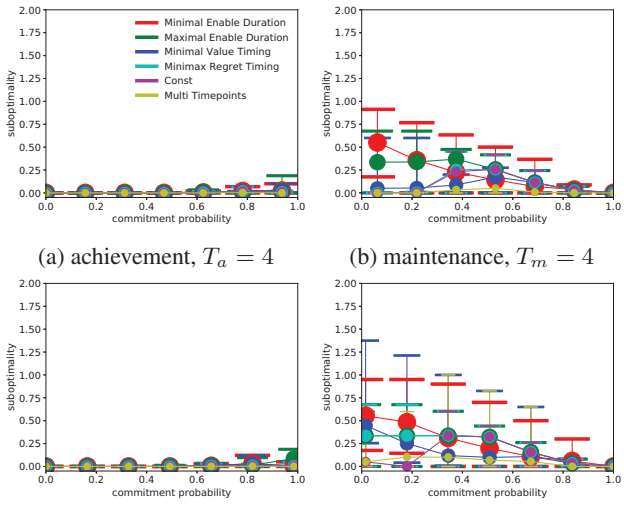
As illustrated in Figure 3, the provider's environment contains four cells, A $\leftrightarrow$ B $\leftrightarrow$ C $\leftrightarrow$ D $\leftrightarrow$ A, that are connected circularly. The provider can deterministically move to an adjacent cell or stay in the current cell. Upon a transition, the gate could toggle with probability 0.5 if the provider ends up in cell C. In the achievement commitment scenario, the provider gets a +1 reward if it ends up in cell C, and in the maintenance commitment scenario it gets a +1 reward if it ends up in cell A. For a given commitment, the provider adopts a policy that aims to maximize its cumulative reward while respecting the commitment semantics. The recipient gets a -0.1 reward each time step. Upon reaching cell G, the recipient gets a +1 reward and the episode ends.

Besides the four heuristics we considered for the 1D Walk, we further consider the following two that choose an approximate influence outside of the set $\mathcal{P}_u^1$:

Constant. This influence toggles u at every time step up to the commitment time with a constant probability, and the probability is chosen such that the overall probability of toggling by the commitment time matches the commitment probability. It agrees with the minimal enable duration influence after the commitment time.

Multi-timepoints. Besides the commitment time, the provider also provides the recipient with the toggling probabilities for other time steps $\mathcal{T}$. Here, we consider $\mathcal{T} = \left\{1, \lfloor \frac{T}{2} \rfloor\right\}$, and the minimal enable duration heuristic is then used to match the toggling probabilities at these three time steps.

Results. We consider the combination of the following scenarios: the provider can start in any one of the four cells; and the toggling can happen in even, odd, or all time steps. The time horizon is $H = 10$ for both the provider and the recipient. This yields in total 12 (true) influences P_u. Figure 4 shows the mean, maximum, and minimum suboptimality for $T_a, T_m \in \{4, 6\}$ over the 12 influences. Similar to 1D Walk, the results show that the minimal enable duration influence is among the best for achievement commitments, but it is difficult for maintenance commitments to identify a best heuristic, besides the multi-timepoints, that reliably reduces the suboptimality for all commitment time/probability



(a) achievement, $T_a = 4$ (b) maintenance, $T_m = 4$

(c) achievement, $T_a = 6$ (d) maintenance, $T_m = 6$

Figure 4: Suboptimality in Gate Control. Markers on the curves show the mean suboptimality over possible time steps of toggling. Bars show the minimum and maximum.

pairs we consider. Using the multi-timepoints influence that is more aligned with the true influence, the suboptimality can be dramatically reduced for maintenance commitments, but it has a less significant impact for achievement commitments. This suggests that, unlike achievement commitments where the cost is low for the provider to retain considerable flexibility by only committing to a single time-probability pair (leaving itself freedom to change its policy dynamically so long as it meets or exceeds that target), maintenance commitments greatly benefit from a provider planning with more constraints on its influence, sacrificing flexibility in order to improve the quality of the recipient's expectations to reduce the frequency and costs of negative surprises.

Conclusion

We have explained why the application of algorithmic and representational strategies that have succeeded for probabilistic commitments of achievement have failed for those of maintenance, despite the fact that the two types of commitments are identical except in their directions of precondition toggling. Contrary to intuitions, the difficulty lies not on the provider's side, but as we have analytically and empirically shown it lies in the recipient's uncertainty in how to approximate the provider's influence that probabilistically changes the precondition over time. Though this uncertainty is present for both commitment types, we have proven that the suboptimality it induces is effectively unbounded only in the case of maintenance commitments, for which we have also empirically shown that even sophisticated heuristic approximations of the influences fall short.

Knowing that approximating influences well is harder for maintenance commitments can affect how multiagent systems are designed. While the customarily terse commitment specification can be beneficial for achievement, in that

the flexibility it engenders to the provider outweighs the bounded suboptimality it imposes on the recipient, it is a liability for maintenance, where the recipient's suboptimality is unbounded. As our experiments showed, performance with maintenance commitments can rise with a more expressive specification, where the provider and recipient adhere to more narrowly-constraining influences. While our immediate plans are to study alternative heuristics for single- and multi-step approximate influences more deeply, our hope is that our results so far might prove illuminating to the broader community designing specifications and protocols for applying commitment-based coordination to domains involving both achievement and maintenance.

Acknowledgments This work was supported in part by the Air Force Office of Scientific Research under grant FA9550-15-1-0039, and the Open Philanthropy Project to the Center for Human-Compatible AI. Opinions, findings, conclusions, or recommendations expressed here are those of the authors and do not necessarily reflect the views of the sponsors.

References

Altman, E. 1999. *Constrained Markov decision processes*, volume 7. CRC Press.

Bannazadeh, H., and Leon-Garcia, A. 2010. A distributed probabilistic commitment control algorithm for service-oriented systems. *IEEE Transactions on Network and Service Management* 7(4):204–217.

Clement, B. J., and Schaffer, S. R. 2008. Exploiting C-TÆMS models for policy search. In *ICAPS Workshop on Multiagent Planning*.

Desai, N.; Narendra, N. C.; and Singh, M. P. 2008. Checking correctness of business contracts via commitments. In *AAMAS*, 787–794.

Duff, S.; Thangarajah, J.; and Harland, J. 2014. Maintenance goals in intelligent agents. *Computational Intelligence* 30(1):71–114.

Durfee, E. H., and Singh, S. 2016. On the trustworthy fulfillment of commitments. In Osman, N., and Sierra, C., eds., *AAMAS Workshops*, 1–13. Springer.

Goldman, R. P.; Musliner, D. J.; Durfee, E. H.; and Boddy, M. S. 2008. Coordinating highly contingent plans: Biasing distributed MDPs towards cooperative behavior. In *ICAPS Workshop on Multiagent Planning*.

Günay, A.; Liu, Y.; and Zhang, J. 2016. Promoca: Probabilistic modeling and analysis of agents in commitment protocols. *JAIR* 57:465–508.

Hiatt, L. M. 2009. *Probabilistic Plan Management*. Ph.D. Dissertation, Carnegie Mellon University.

Hindriks, K. V., and van Riemsdijk, M. B. 2007. Satisfying maintenance goals. In *5th Int. Workshop Declarative Agent Languages and Technologies (DALT)*, 86–103.

Jennings, N. R. 1993. Commitments and conventions: The foundation of coordination in multi-agent systems. *The Knowledge Engineering Review* 8(3):223–250.

Kalia, A. K.; Zhang, Z.; and Singh, M. P. 2014. Estimating trust from agents' interactions via commitments. In *ECAI*, 1043–1044.

Maheswaran, R.; Szekely, P.; Becker, M.; Fitzpatrick, S.; Gati, G.; Jin, J.; Neches, R.; Noori, N.; Rogers, C.; Sanchez, R.; et al. 2008. Predictability & criticality metrics for coordination in complex environments. In *AAMAS*, 647–654.

Oliehoek, F. A.; Spaan, M. T.; and Witwicki, S. J. 2015. Influence-optimistic local values for multiagent planning. In *AAMAS*, 1703–1704.

Oliehoek, F. A.; Witwicki, S. J.; and Kaelbling, L. P. 2012. Influence-based abstraction for multiagent systems. In *AAAI*, 1422–1428.

Pereira, R. F.; Oren, N.; and Meneguzzi, F. 2017. Detecting commitment abandonment by monitoring sub-optimal steps during plan execution. In *AAMAS*, 1685–1687.

Sandholm, T., and Lesser, V. R. 2001. Leveled commitment contracts and strategic breach. *Games and Economic Behavior* 35:212–270.

Singh, M. P. 1999. An ontology for commitments in multiagent systems. *Artificial Intelligence and Law* 7(1):97–113.

Singh, M. P. 2008. Semantical considerations on dialectical and practical commitments. In *AAAI*, 176–181.

Singh, M. P. 2012. Commitments in multiagent systems: Some history, some confusions, some controversies, some prospects. In *The Goals of Cognition. Essays in Honor of Cristiano Castelfranchi*, 601–626.

Steinmetz, M.; Hoffmann, J.; and Buffet, O. 2016. Goal probability analysis in probabilistic planning: exploring and enhancing the state of the art. *JAIR* 57:229–271.

Vokrínek, J.; Komenda, A.; and Pechoucek, M. 2009. Decommitting in multi-agent execution in non-deterministic environment: experimental approach. In *AAMAS*, 977–984.

Winikoff, M. 2006. Implementing flexible and robust agent interactions using distributed commitment machines. *Multiagent and Grid Systems* 2(4):365–381.

Witwicki, S. J., and Durfee, E. H. 2007. Commitment-driven distributed joint policy search. In *AAMAS*, 480–487.

Witwicki, S. J., and Durfee, E. H. 2009. Commitment-based service coordination. *Int.J. Agent-Oriented Software Engineering* 3:59–87.

Witwicki, S. J., and Durfee, E. H. 2010. Influence-based policy abstraction for weakly-coupled Dec-POMDPs. In *ICAPS*, 185–192.

Xing, J., and Singh, M. P. 2001. Formalization of commitment-based agent interaction. In *Proc. of the 2001 ACM Symposium on Applied computing*, 115–120.

Xuan, P., and Lesser, V. R. 1999. Incorporating uncertainty in agent commitments. In *International Workshop on Agent Theories, Architectures, and Languages*, 57–70. Springer.

Zhang, Q.; Durfee, E. H.; Singh, S.; Chen, A.; and Witwicki, S. J. 2016. Commitment semantics for sequential decision making under reward uncertainty. In *IJCAI*, 3315–3323.

Zhang, Q.; Singh, S.; and Durfee, E. 2017. Minimizing maximum regret in commitment constrained sequential decision making. In *ICAPS*, 348–356.