Computational Strategies for the Trustworthy Pursuit and the Safe Modeling of Probabilistic Maintenance Commitments

Qi Zhang , Edmund 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 conditions for other agents. Our work reveals that probabilistic commitments to instead maintain conditions for others are surprisingly different from their achievement counterparts, despite strong semantic similarities. We focus on the question of how the commitment recipient should model the provider's effect on the recipient's local environment, with only imperfect information being provided in the commitment specification. Our theoretic analyses show that we can more tightly bound the inefficiency of this imperfect modeling for achievement commitments than for maintenance commitments. We empirically demonstrate that probabilistic maintenance commitments are qualitatively more challenging for the recipient to model, and addressing the challenges can require the provider to adhere to a more detailed profile and sacrifice flexibility.

1 Introduction

Safe cooperative behavior among humans is often realized via social commitments that constrain people to acting reliably. By making a commitment, a person promises to act in a manner to fulfill it. This form of commitment-based interaction also exists among autonomous agents. To build safe artificial multiagent systems, we need trustworthy and reliable mechanisms that are accountable for pursuing and modeling agent-based commitments. Adopting a decision-theoretic framework, this paper formulates and studies problems that arise in managing commitments in multiagent systems.

In multiagent systems, agents are often interdependent in the sense that what one agent does can be beneficial or harmful to another. If trust means having confidence that another will act so as to reciprocate help in a safe and reliable manner, then being trustworthy—worthy of such trust—constrains the agent to acting thusly. To persuade an agent designer to create trustworthy agents, other agents (individually and/or collectively) can form and share opinions about agents' trustworthiness, and won't act to benefit agents with a bad reputation.

Our work assumes that the designer has been persuaded. Even so, however, it isn't always clear how to create a trustworthy agent, given that an agent often lacks complete control over its environment. Specifically, the form of interdependency we focus on is with respect to a scenario where an agent (the commitment *provider*) makes a social commitment [Singh, 1999; Kalia *et al.*, 2014] to another (the commitment *recipient*). When stochasticity is inherent in the environment, the provider cannot guarantee to bring about the outcomes that the recipient expects [Kwiatkowska *et al.*, 2007; Nuzzo *et al.*, 2019], and in fact could discover after making the commitment that how it planned to try to bring about the outcomes would be more costly or risky than it had previously realized. Given that the recipient is unable to predict precisely the future situations it will face, it's also unclear how it should model the commitment safely and effectively.

There exists work focusing on semantics and mechanisms for an agent to follow such that it is assured of faithfully pursuing its commitments despite the uncertainty [Jennings, 1993; Xing and Singh, 2001; Winikoff, 2006; Durfee and Singh, 2016]. Previous work articulated the perspective that such a *probabilistic* commitment should be considered fulfilled if the provider's actions would have brought about the desired outcome with a high enough expectation, even if in a particular instance the desired outcome was not realized. That is, the provider acted in good faith. Thus, even if the provider changes its course of action as it learns more about costs and risks on the fly, it can still fulfill its commitment if whatever course of action it pursued could be expected to achieve the desired outcome with at least the promised likelihood. With this perspective, previous work has focused largely on commitments of achievement [Xuan and Lesser, 1999; Maheswaran *et al.*, 2008; Witwicki and Durfee, 2009; Zhang *et al.*, 2016; Zhang *et al.*, 2017], which we also call *enablement* commitments, 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 them is open) that it is counting on the provider to enable.

This paper focuses on another form of commitment, which we refer to as a *maintenance commitment*, where instead of committing to some course of action that in expectation will enable conditions the recipient wants, the provider instead commits to courses of action to probabilistically avoid changing conditions that are already the way the recipient wants

them maintained, up until a particular time. After that time, the condition cannot be assumed to remain unchanged, and before that time, there is a (usually small) probability it could be changed at any point. For example, an open door the recipient needs might be initially achieved, but as the provider opens and closes other doors during its housekeeping tasks, a resulting draft could close the door the recipient needs open. The provider could plan its tasks to postpone altering the riskiest doors as long as possible, but an ill-placed breeze could close the door at any time.

Our claim is that decision-theoretic mechanisms for representing and reasoning about enablement commitments cannot straightforwardly apply to maintenance commitments because, despite strong superficial similarities, the two types of commitments are fundamentally different. We will substantiate this claim analytically and empirically. This in turn raises the questions of whether modifications could be made to existing decision-theoretic mechanisms for representing and reasoning about enablement commitments so that they can be applied to maintenance commitments. We will show empirical results that cast doubt on whether this is possible, thus suggesting that in the future a different treatment of maintenance commitments should be considered.

2 Preliminaries

In this section, we describe the decision-theoretic setting we adopt for analyzing probabilistic commitments, including both enablement commitments and maintenance commitments. We review the prior work in the definition and semantics of enablement commitments, which can be extended to maintenance commitments.

The **recipient**'s environment is modeled as an 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 $\mathcal{S}$), $R : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ 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 trajectory is generated by $a_h = \pi(s_h), s_{h+1} \sim P(s_h, a_h), r_h = R(s_h, a_h, s_{h+1})$ for $h = 0, \cdots, H-1$. The value function of π is $V_M^\pi(s) = \mathbb{E}[\sum_{h'=h}^{H-1} r_{h'} | \pi, s_h = s]$ where h is such that $s \in \mathcal{S}_h$. There exists an optimal policy in M, denoted as π_M^*, and its value function maximizes V_M^π for all $s \in \mathcal{S}$ and is abbreviated as V_M^*. The value of the initial state is abbreviated as $v_M^\pi := V_M^\pi(s_0)$.

As one way to model the interaction between the provider and the recipient [Witwicki and Durfee, 2010; Zhang *et al.*, 2016], we assume that the recipient's state can be 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 state feature shared with the provider. The provider and the recipient are weakly coupled in the sense that u is the only shared state feature and is only controllable by the provider. Formally, the dynamics of the recipient's state can be factored as

$$P(s'|s,a) = P\left((l',u')|(l,u),a\right) = P_u(u'|u) P_l\left(l'|(l,u),a\right).$$

We assume the recipient's cumulative reward can be expressed in the trajectory of l:

$$R(s,a,s') = R(s') = R((l',u')) = R(l').$$

Note that though the value of u does not directly determine the reward, it does affect the value of l' at the next time step. Throughout, we refer to P_u as the true *profile* of u, which is fully determined by the provider's policy.

2.1 Commitment Semantics

An enablement or maintenance commitment is concerned with state feature u that is 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. We will refer to u as the commitment feature. In this paper, we focus on the setting where the value of u is binary, letting u^+, as opposed to u^-, be the value of u that is desirable for the recipient. Further, we assume that u can be toggled at most once. Citations with this assumption include [Hindriks and van Riemsdijk, 2007; Witwicki and Durfee, 2009; Zhang *et al.*, 2016]. In transactional settings, a feature (e.g., possession of goods) changing once is common. It is also common in multiagent planning domains where one agent establishes 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 the extreme case where toggling reliably repeats frequently (e.g., a traffic light) there may be no need for explicit commitments. More generally, while removing this assumption can complicate the specification of a commitment (e.g., a compound commitment to enable and then maintain a condition over a time interval), we think the fundamental difference between modeling enablement and maintenance commitments can be best theoretically explained and conceptually understood without such complications. We next formally give the definition and semantics of enablement commitments and maintenance commitments, respectively.

Enablement Commitments

Let the initial state be factored as $s_0 = (l_0, u_0)$. For enablement 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 probability. Formally, an enablement commitment is defined by tuple $c_e = (T_e, p_e)$, where T_e is the enablement commitment time, and p_e is the enablement commitment probability. The provider's commitment semantics is to follow a policy μ that sets u to u^+ by time step T_e with at least probability p_e, i.e.

$$\Pr(u_{T_e} = u^+ | u_0 = u^-, \mu) \geq p_e.$$

Maintenance Commitments

As a reminder, our maintenance commitment is motivated by 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 *et al.*, 2014]), but where the provider could 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 μ 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, it is equivalent to guaranteeing $u = u^+$ at T_m, i.e.

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

2.2 The Recipient's Approximate Profile

The commitment semantics provides partial information on P_u by specifying the profile at the single time step, rather than in a potentially more gradual manner. Why would we choose to do that? The most important reason is that it opens up even more latitude for the provider to evolve its policy as it learns more about its environment: instead of needing to meet probabilistic expectations at multiple time steps, it can modify its policy much more flexibly as long as in the long run (by the commitment time) it hits the target probability. Prior work has shown the value of having such flexibility [Zhang *et al.*, 2017]. Here, we are interested in the problem when the recipient creates a profile $\widehat{P}_u$ with this partial information as an approximation of P_u, and plans accordingly. Specifically, we are interested in the quality of the plan computed from approximate profile $\widehat{P}_u$ when evaluated in (true) profile 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 profile 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.

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

Note that when the support of P_u is not fully contained in the support of $\widehat{P}_u$, the recipient could end up in un-modelled states when executing $\pi^*_{\widehat{M}}$ in M, which makes $V^{\pi^*_{\widehat{M}}}_M$ ill-defined. In this paper, we fix this by re-planning: during execution of $\pi^*_{\widehat{M}}$, the recipient re-plans from un-modelled states.

Previous work chooses an intuitive and straightforward approximate profile for enablement commitments that models a single branch (at the commitment time) for when u^- probabilistically toggles to u^+. This strategy reduces the complexity of the recipient's reasoning; the inefficiency caused by imperfect modelling is easily outweighed by computational benefits [Witwicki and Durfee, 2010; Zhang *et al.*, 2016]. This profile takes a pessimistic view in the sense that u is (stochastically) enabled at the latest possible time consistent with the commitment, and, if it was not enabled at that time, will never be enabled after the commitment time. Therefore, we refer to it as the pessimistic profile, as formalized in Definition 1. For maintenance commitments, the pessimistic profile should probabilistically disable u at the earliest time, and should determinisically disable u after the commitment time, as formalized in Definition 2.

Definition 1. Given enablement commitment $c_e = (T_e, p_e)$, its *pessimistic* profile $\widehat{P}^{\text{pessimistic}}_{u,c_e}$ toggles u in the transition from time step $t = T_e - 1$ to $t = T_e$ with probability p_e, and does not toggle u at any other time step.

Definition 2. Given maintenance commitment $c_m = (T_m, p_m)$, its *pessimistic* profile $\widehat{P}^{\text{pessimistic}}_{u,c_m}$ toggles u in the transition from time step $t = 0$ to $t = 1$ with probability $1 - p_m$, and from $t = T_m$ to $t = T_m + 1$ with probability one. It does not toggle u at any other time step.

In a previous, unpublished workshop paper [Zhang *et al.*, 2018], we have addressed the general topic of using an approximate profile, especially the pessimitisc profile, for the recipient to model probabilistic commitments, but didn't provide any results. This paper presents both theoretical analysis (Section 3) and empirical results (Section 4) that reveal the fundamental difference between enablement and maintenance commitments.

3 Theoretical Analysis

In this section, we derive bounds on the suboptimality of the pessimistic profiles. Our analysis makes the following two assumptions. Assumption 1 intuitively says that u^+ establishes a condition for an action that would be irrational, or even unsafe, 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 analysis stating the true profile agrees with the pessimistic profile after the commitment time, so that the suboptimality is caused by the imperfect modeling by 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 profile 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 pessimistic profile for $h \geq T$, where T is the commitment time.

To derive bounds on enablement 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 profile 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 the other way around, 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 $s^+ = (l, u^+)$, we have $\pi_M(s^+) = \pi_M(s^-) = \pi^*_{M^-}(s^-)$. By Assumption 1, π_M in M yields the same trajectory distribution of l as $\pi^*_{M^-}$ in M^-, and therefore $v^{\pi_M}_M = v^*_{M^-}$ since value 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 $s^- = (l, u^-)$, $s^+ = (l, u^+)$.

For the case in which P_u toggles u at $K > 1$ time steps, we can decompose the value function P_u as the weighted average of K value functions corresponding to the K profiles 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. □

3.1 Bounding Suboptimality for Enablement

Here, we derive a bound on the suboptimality for enablement commitments. From the two assumptions, we also make use of Lemma 2 to prove Theorem 1 that bounds the suboptimality for enablement commitments as the difference between $v^*_{M^-}$ and $v^*_{M^+}$. Lemma 2 states that, for enablement commitments, the possible imperfect modeling of the pessimistic profile can only improve the expected value.

Lemma 2. Given enablement commitment $c_e = (T_e, p_e)$, let $\widehat{P}_u = \widehat{P}^{\text{pessimistic}}_{u,c_e}$, then we have $v^{\pi^*_{\widehat{M}}}_M \geq v^{\pi^*_{\widehat{M}}}_{\widehat{M}}$ where profile P_u in M respects the commitment semantics of c_e.

Proof. For enablement commitments, the initial value of u is u^-. Let $P_u(t)$ be the probability that u is enabled to u^+ at t in profile P_u, $\overline{v}^{\pi}_t$ be the initial state value of 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

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

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 recipient's replanning 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. $\overline{v}^{\pi^*_{\widehat{M}}}_t \geq \overline{v}^{\pi^*_{\widehat{M}}}_{T_e}$. Therefore,

$$\begin{aligned} v^{\pi^*_{\widehat{M}}}_M =& \textstyle\sum_{t=1}^{T_e} P_u(t)\overline{v}^{\pi^*_{\widehat{M}}}_t + (1-p_e)v^{\pi^*_{\widehat{M}}}_{M^-} \\ \geq& \textstyle\sum_{t=1}^{T_e} P_u(t)\overline{v}^{\pi^*_{\widehat{M}}}_{T_e} + (1-p_e)v^{\pi^*_{\widehat{M}}}_{M^-} \\ \geq& p_e\overline{v}^{\pi^*_{\widehat{M}}}_{T_e} + (1-p_e)v^{\pi^*_{\widehat{M}}}_{M^-} \qquad \text{commitment semantics} \\ =& v^{\pi^*_{\widehat{M}}}_{\widehat{M}}. \end{aligned}$$

□

Theorem 1. Given enablement commitment c_e, let $\widehat{P}_u = \widehat{P}^{\text{pessimistic}}_{u,c_e}$. The suboptimality can be bounded as

$$v^*_M - v^{\pi^*_{\widehat{M}}}_M \leq v^*_{M^+} - v^*_{M^-} \tag{1}$$

where profile P_u in M respects the commitment semantics of c_e. Further, there exists an enablement commitment for which the equality is attained.

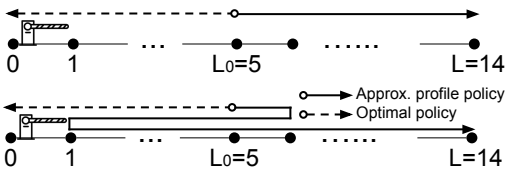


Figure 1: 1D Walk. *Up:* Example in the proof of Theorem 1. *Down:* Example in the proof of Theorem 2.

Proof. The derivation is straightforward from Lemma 2:

$$v^*_M - v^{\pi^*_{\widehat{M}}}_M \leq v^*_{M^+} - v^{\pi^*_{\widehat{M}}}_{\widehat{M}} \leq v^*_{M^+} - v^*_{M^-}.$$

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

Example: An Enablement Commitment in 1D Walk
Consider the example of a 1D walk on $[0, L]$, as illustrated in Figure 1(top), in which 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^- denotes closed. The gate toggles stochastically according to P_u. For each step until the recipient reaches either end, a -1 reward is given. Therefore, 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 trivial case of $v^*_{M^-} = v^*_{M^+}$. A negative reward is incurred when bumping into the closed gate, which makes Assumption 1 hold.

Here, we derive an enablement commitment for which the bound in Theorem 1 is attained. Consider $L = 14, L_0 = 5, H = 15$, enablement commitment $(T_e = L - L_0 = 9, p_e = 1)$, and the true profile P_u in M that toggles the gate to open at $t = 4$ with probability $p_e = 1$. The optimal policy in M is to move left to 0. Therefore, $v^*_M = v^*_{M^+} = -L_0 = -5$. Given the pessimistic profile, moving right to L (arriving at time 9) is faster than waiting for the gate to toggle at $T_e = 9$ 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 going to L is the faster choice. Therefore $v^{\pi^*_{\widehat{M}}}_M = v^*_{M^-} = -(L - L_0) = -9$, and bound (1) is attained. □

3.2 Bounding Suboptimality for Maintenance

We next ask if the bound in Equation (1) on suboptimality in enablement commitments also holds for maintenance commitments. Unfortunately, as stated in Theorem 2, the optimal policy of the pessimistic profile for maintenance commitments can be arbitrarily bad when evaluated in the true profile, incurring a suboptimality exceeding (1). An existence proof is given with an example.

Theorem 2. There exists an MDP M with nonnegative rewards, and a maintenance commitment c_m, such that the profile P_u in M respects the commitment semantics of c_m, $v^*_M = v^*_{M^+}$, $v^{\pi^*_{\widehat{M}}}_M = 0$, and therefore the suboptimality is

$$v^*_M - v^{\pi^*_{\widehat{M}}}_M = v^*_{M^+} \tag{2}$$

where $\widehat{P}_u = \widehat{P}^{\text{pessimistic}}_{u,c_m}$ is the profile in $\widehat{M}$.

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

Consider 1D Walk with the same $L = 14, L_0 = 5, H = 15$ as in the example for Theorem 1. Here we offset the rewards by $+1$ such that the rewards are nonnegative. Consider maintenance commitment $(T_m = 7, p_m = 0)$, and P_u toggles the gate from open to closed at $t = 6$ with probability $1 - p_m = 1$. As illustrated in Figure 1(bottom), the optimality policy should take 5 steps to move directly to 0, for which the value is $v^*_{M^+}$. With probability p_m, the gate is kept open through T_m, and $\pi^*_{\widehat{M}}$ takes 7 steps to reach 0. With probability $1 - p_m$, the gate is closed at $t = 6$, and $\pi^*_{\widehat{M}}$ takes $19 > H$ steps to reach $L = 14$. Therefore, $v_M^{\pi^*_{\widehat{M}}} = 0$. □

Comparing bound (1) in Theorem 1 with bound (2) in Theorem 2 reveals a fundamental difference between enablement and maintenance commitments: maintenance commitments are inherently less tolerant to an unexpected change in the commitment feature. For enablement commitments, it is easy to construct a pessimistic profile, such that any unexpected changes to the feature, if they impact the recipient at all, can only improve the expected value. Thus, if despite the pessimistic profile, a recipient has chosen to follow a policy that exploits the commitment, it can never experience a true profile that would lead it to regret having done so. The same cannot be said for maintenance commitments. The easily-constructed pessimistic profile does not guarantee that any deviations from the profile can only improve the expected value. As our theoretical results show, the pessimistic profile of assuming toggling from u^+ to u^- right away can still lead to negative surprises, since if the toggling doesn't occur the profile 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. The poor performance of the pessimistic model for maintenance is because it is not reliably pessimistic enough: in the example for Theorem 2, the worst time for toggling to u^- is not right away, but right before the condition would be used (the gate shutting just as the recipient was about to pass through).

4 Empirical Results

Our analyses suggest the pessimistic profile might not be the best approximate profile for a recipient to adopt for maintenance commitments. In this section, we identify several alternative heuristics to create approximate profiles for the recipient, and evaluate them for both maintenance and enablement commitments. We conduct our evaluations in two domains. The first is the same 1D Walk domain as in our theoretical analysis, and the second is a Gate Control problem with a more interesting transition profile (violating Assumption 2).

4.1 1D Walk

As previously defined, the 1D Walk domain restricts the set of profiles 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 profiles as $\mathcal{P}^1_u$ from which $P_u, \widehat{P}_u$ are chosen. Besides using the pessimistic profile to approximate the true profile, we consider the following three heuristics for generating approximate profile $\widehat{P}_u \in \mathcal{P}^1_u$:

Optimistic. As opposed to the pessimistic, the optimistic profile toggles u right after the initial time step for enablement commitments, and at the commitment time for maintenance commitments.

Minimum Value. The toggling time minimizes the optimal value over all possible profiles in $\mathcal{P}^1_u$, i.e., $\arg\min_{\widehat{P}_u \in \mathcal{P}^1_u} v^*_{\widehat{M}}$, where $\widehat{P}_u$ is the profile of u in $\widehat{M}$.

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

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

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

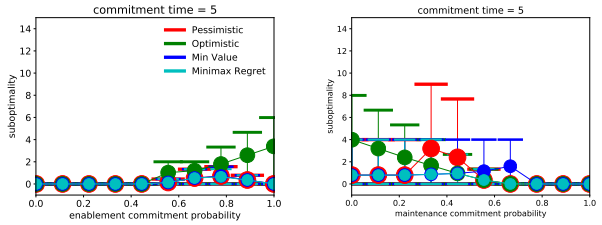
The four heuristics include two simple, inexpensive heuristics (Pessimistic and Optimistic), and two more complex and expensive heuristics (Minimum Value and Minimax Regret). Recall that our theoretical analysis suggests, for maintenance, the worst time for toggling to u^- is not right away, but right before the recipient uses the condition, and this causes the poor performance of the pessimistic profile. We hypothesize that the latter two heuristics can improve the pessimistic profile by identifying the worst toggling time.

Results Here we evaluate the suboptimality of our candidate heuristics for both enablement 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 $P_u \in \mathcal{P}^1_u$ for commitment time $T_e, T_m \in \{5, 7, 10\}$. We see that for enablement commitments, the suboptimality of the pessimistic profile is comparable to the two more sophisticated strategies, and the optimistic profile incurs most suboptimality overall. For maintenance commitments, however, the two expensive strategies incur overall less suboptimality than the pessimistic and the optimistic, yet it is difficult to identify a single best heuristic that reliably reduces the suboptimality for all the maintenance commitments.

4.2 Gate Control

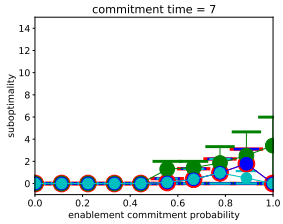
In this domain, we are concerned with the more general situation in which $P_u \notin \mathcal{P}^1_u$ can toggle u at more than one time step by the commitment time, and even can toggle u after the commitment time. We also consider approximate profiles $\widehat{P}_u$ that are not elements of $\mathcal{P}^1_u$.

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 enablement 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 ending up in cell A. For a given commitment, the provider adopts

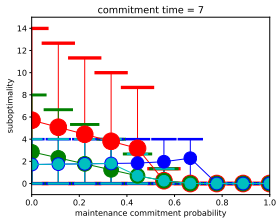


(a) enablement, $T_e = 5$

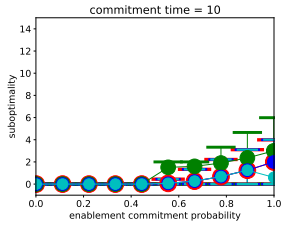
(b) maintenance, $T_m = 5$



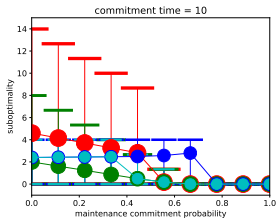
(c) enablement, $T_e = 7$



(d) maintenance, $T_m = 7$



(e) enablement, $T_e = 10$



(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.

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 profile outside of the set $\mathcal{P}_u^1$:

Constant. This profile 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 pessimistic profile after the commitment time.

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

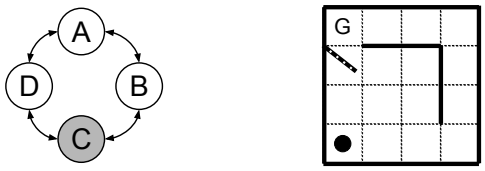
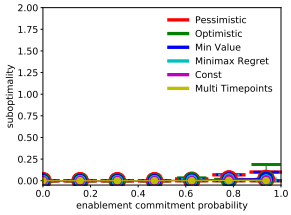
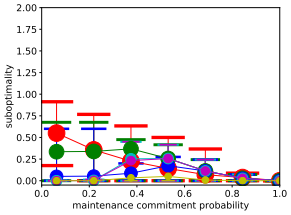


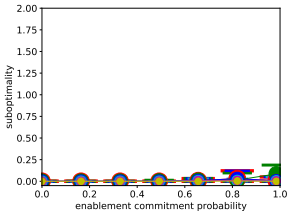
Figure 3: Gate Control. *Left:* The provider. Cell C toggles the gate. *Right:* The recipient.



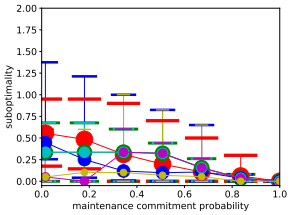
(a) enablement, $T_e = 4$



(b) maintenance, $T_m = 4$



(c) enablement, $T_e = 6$



(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.

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) profiles P_u. Figure 4 shows the mean, maximum, and minimum suboptimality for $T_e, T_m \in \{4, 6\}$ over the 12 profiles. Similar to 1D Walk, the results show that the pessimistic profile is among the best for enablement 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 pairs we consider. Using the multi-timepoints profile that is more aligned with the true profile, the suboptimality can be dramatically reduced for maintenance commitments, but it has a less significant impact for enablement commitments. This suggests that, unlike enablement commitments where the cost is low of the provider retaining 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 committing to a more detailed profile, sacrificing flexibility in order to improve the quality of the recipient's expectations to reduce the frequency and costs of negative surprises.

5 Conclusion

We have shown that we cannot straightforwardly extend the semantics and algorithms for trustworthy fulfillment of enablement commitments to maintenance commitments. Our theoretical and empirical results suggest that, despite their similarities in describing the toggling of conditions over time,

maintenance commitments are fundamentally different from enablement commitments. We have theoretically shown that the easily-constructed pessimistic profile can only improve the expected value in the face of unexpected changes for enablement commitments but not for maintenance commitments. Empirically, we have seen that an inexpensive pessimistic approximation of the profile works comparably to more sophisticated approximations for enablement commitments, but not for maintenance commitments.

The fact that approximating profiles well is harder for maintenance commitments could mean that agents engaged in maintenance commitments might need to make a different tradeoff. That is, for enablement, we could give the provider a lot of flexibility by only constraining it to meet the probability at the commitment time and so can unilaterally change the profile before then. The gain in flexibility for the provider is worth the relatively small value loss to the recipient from using the pessimistic profile. However, for maintenance commitments, the potential for the recipient to lose more value with a bad (simple or sophisticated) approximate profile could mean that the provider should commit to a more detailed profile—the loss of flexibility for the provider in this case is warranted because the recipient makes much better decisions. In other words, our theoretical and empirical work suggests that maintenance commitments could require providers and recipients to inherently be more tightly coupled than they need to be for enablement commitments.

Acknowledgments We thank the anonymous reviewers for great suggestions. This work was supported in part by the Air Force Office of Scientific Research under grant FA9550-15-1-0039. Any opinions, findings, conclusions, or recommendations expressed here are those of the authors and do not necessarily reflect the views of the sponsors.

References

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

[Durfee and Singh, 2016] Edmund H. Durfee and Satinder Singh. On the trustworthy fulfillment of commitments. In Nardine Osman and Carles Sierra, editors, *AAMAS Workshops*, pages 1–13. Springer, 2016.

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

[Jennings, 1993] Nick R Jennings. Commitments and conventions: The foundation of coordination in multi-agent systems. *The knowledge engineering review*, 8(3):223–250, 1993.

[Kalia *et al.*, 2014] A. K. Kalia, Z. Zhang, and M. P. Singh. Estimating trust from agents' interactions via commitments. In *21st Euro. Conf. on AI (ECAI)*, pages 1043–1044, 2014.

[Kwiatkowska *et al.*, 2007] Marta Kwiatkowska, Gethin Norman, and David Parker. Stochastic model checking. In *International School on Formal Methods for the Design of Computer, Communication and Software Systems*, pages 220–270. Springer, 2007.

[Maheswaran *et al.*, 2008] Rajiv Maheswaran, Pedro Szekely, Marcel Becker, Stephen Fitzpatrick, Gergely Gati, Jing Jin, Robert Neches, Narges Noori, Craig Rogers, Romeo Sanchez, et al. Predictability & criticality metrics for coordination in complex environments. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 2*, pages 647–654, 2008.

[Nuzzo *et al.*, 2019] Pierluigi Nuzzo, Jiwei Li, Alberto L. Sangiovanni-Vincentelli, Yugeng Xi, and Dewei Li. Stochastic assume-guarantee contracts for cyber-physical system design. *ACM Trans. Embed. Comput. Syst.*, 18(1):2:1–2:26, January 2019.

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

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

[Witwicki and Durfee, 2009] Stefan J. Witwicki and Edmund H. Durfee. Commitment-based service coordination. *Int.J. Agent-Oriented Software Engineering*, 3:59–87, 01 2009.

[Witwicki and Durfee, 2010] Stefan J. Witwicki and Edmund H. Durfee. Influence-based policy abstraction for weakly-coupled Dec-POMDPs. In *Int. Conf. Auto. Planning Sys. (ICAPS)*, pages 185–192, 2010.

[Xing and Singh, 2001] Jie Xing and Munindar P Singh. Formalization of commitment-based agent interaction. In *Proceedings of the 2001 ACM symposium on Applied computing*, pages 115–120. ACM, 2001.

[Xuan and Lesser, 1999] Ping Xuan and Victor R Lesser. Incorporating uncertainty in agent commitments. In *International Workshop on Agent Theories, Architectures, and Languages*, pages 57–70. Springer, 1999.

[Zhang *et al.*, 2016] Qi Zhang, Edmund H Durfee, Satinder Singh, Anna Chen, and Stefan J Witwicki. Commitment semantics for sequential decision making under reward uncertainty. In *Int J. Conf. on Artificial Intelligence (IJCAI)*, pages 3315–3323, 2016.

[Zhang *et al.*, 2017] Qi Zhang, Satinder Singh, and Edmund Durfee. Minimizing maximum regret in commitment constrained sequential decision making. In *Int. Conf. Automated Planning Systems (ICAPS)*, pages 348–356, 2017.

[Zhang *et al.*, 2018] Qi Zhang, Edmund H. Durfee, and Satinder P. Singh. Challenges in the trustworthy pursuit of maintenance commitments under uncertainty. In *Proceedings of the 20th International Trust Workshop co-located with AAMAS/IJCAI/ECAI/ICML 2018, Stockholm, Sweden, July 14, 2018.*, pages 75–86, 2018.