Efficient Querying for Cooperative Probabilistic Commitments

Qi Zhang[1], **Edmund H. Durfee**[2], **Satinder Singh**[2]
[1] Artificial Intelligence Institute, University of South Carolina
[2] Computer Science and Engineering, University of Michigan
qz5@cse.sc.edu, durfee@umich.edu, baveja@umich.edu

Abstract

Multiagent systems can use commitments as the core of a general coordination infrastructure, supporting both cooperative and non-cooperative interactions. Agents whose objectives are aligned, and where one agent can help another achieve greater reward by sacrificing some of its own reward, should choose a cooperative commitment to maximize their *joint* reward. We present a solution to the problem of how cooperative agents can efficiently find an (approximately) optimal commitment by querying about carefully-selected commitment choices. We prove structural properties of the agents' values as functions of the parameters of the commitment specification, and develop a greedy method for composing a query with provable approximation bounds, which we empirically show can find nearly optimal commitments in a fraction of the time methods that lack our insights require.

1 Introduction

Commitments are a proven approach to multiagent coordination (Singh 2012; Cohen and Levesque 1990; Castelfranchi 1995; Mallya and Huhns 2003; Chesani et al. 2013; Al-Saqqar et al. 2014). Through commitments, agents know more about what to expect from others, and thus can plan actions with higher confidence of success. That said, commitments are generally uncertain: an agent might abandon a commitment if it discovers that it cannot achieve what it promised, or that it prefers to achieve something else, or that others will not uphold their side of the commitment (Jennings 1993; Xing and Singh 2001; Winikoff 2006).

One way to deal with commitment uncertainty is to institute protocols so participating agents are aware of the status of commitments through their lifecycles (Venkatraman and Singh 1999; Xing and Singh 2001; Yolum and Singh 2002; Fornara and Colombetti 2008; Baldoni et al. 2015; Günay, Liu, and Zhang 2016; Pereira, Oren, and Meneguzzi 2017; Dastani, van der Torre, and Yorke-Smith 2017). Another has been to qualify commitments with conditional statements about what must (not) be true in the environment for the commitment to be fulfilled (Singh 2012; Agotnes, Goranko, and Jamroga 2007; Vokrínek, Komenda, and Pechoucek 2009). When such conditions might not be fully observable to all agents, agents might summarize the likelihood of the conditions being satisfied in the form of a *probabilistic commitment* (Kushmerick, Hanks, and Weld 1994; Xuan and Lesser 1999; Witwicki and Durfee 2007).

Our focus is the process by which agents choose a probabilistic commitment, which serves as a probabilistic promise from one agent (the *provider*) to another (the *recipient*) about establishing a precondition for the recipient's preferred actions/objectives. We formalize how the space of probabilistic commitments for the precondition captures different tradeoffs between timing and likelihood, where in general the recipient gets higher reward from earlier timing and/or higher likelihood, while the provider prefers later timing and/or lower likelihood because these leave it less constrained when optimizing its own policy. Thus, when agents agree to work together (e.g., (Han, Pereira, and Lenaerts 2017)), forming a commitment generally involves a negotiation (Kraus 1997; Aknine, Pinson, and Shakun 2004; Rahwan 2004).

Sometimes, however, a pair of agents might have objectives/payoffs that are aligned/shared. For example, they might be a chef and waiter working in a restaurant (see Section 6.2). In a commitment-based coordination framework, such agents should find a *cooperative* probabilistic commitment, whose timing and likelihood maximizes their *joint* (summed) reward, in expectation. Decomposing the joint reward into local, individual rewards is common elsewhere as well, like in the Dec-POMDP literature (Oliehoek, Amato et al. 2016) and multi-agent reinforcement learning (Zhang et al. 2018). This optimization problem is complicated by two main factors: i) the information relevant to optimization is distributed, and thus the agents need to exchange knowledge, preferably with low communication cost; and ii) the space of possible timing/probability combinations is large and evaluating a combination (requiring each agent to compute an optimal policy) is expensive, and thus identifying a desirable probabilistic commitment is computationally challenging even with perfect centralized information.

The main contribution of this paper is an approach that addresses both challenges for cooperative agents to efficiently converge on an approximately-optimal probabilistic commitment. To address i), our approach adopts a decentralized, query-based protocol for the agents to exchange knowledge effectively with low communication cost. To get the efficiency for ii), we prove the existence of structural properties in the agents' value functions, and show that these can be provably exploited by the query-based protocol.

2 Related Work

Commitments are a widely-adopted framework for multi-agent coordination (Kushmerick, Hanks, and Weld 1994; Xuan and Lesser 1999; Singh 2012). We build on prior research on probabilistic commitments (Xuan and Lesser 1999; Bannazadeh and Leon-Garcia 2010), where the timing and likelihood of achieving a desired outcome are explicitly specified. Choosing a probabilistic commitment thus corresponds to searching over the combinatorial space of possible commitment times and probabilities. Prior work on such search largely relies on heuristics. Witwicki et al. (2007) propose the first probabilistic commitment search algorithm that initializes a set of commitments and then performs local adjustments on time and probability. Later work (Witwicki and Durfee 2009; Oliehoek, Witwicki, and Kaelbling 2012) further incorporates the commitment's feasibility and the best response strategy (Nair et al. 2003) to guide the search. In contrast to these heuristic approaches, in this paper we analytically reveal the structure of the commitment space, which enables efficient search that provably finds the optimal commitment.

Because the search process is decentralized, it will involve message passing. The message passing between our decision-theoretic agents serves the purpose of preference elicitation, which is typically framed in terms of an agent querying another about which from among a set of choices it most prefers (Chajewska, Koller, and Parr 2000; Boutilier 2002; Viappiani and Boutilier 2010). We adopt such a querying protocol as a means for information exchange between the agents. In particular, we draw on recent work that uses value-of-information concepts to formulate multiple-choice queries (Viappiani and Boutilier 2010; Cohn, Singh, and Durfee 2014; Zhang, Durfee, and Singh 2017), but as we will explain we augment prior approaches by annotating offered choices with the preferences of the agent posing the query. Moreover, we prove several characteristic properties of agents' commitment value functions, which enables efficient formulation of near-optimal queries.

3 Decision-Theoretic Commitments

The provider's and recipient's environments are modeled as two separate Markov Decision Processes (MDPs). An MDP is defined as $M = (S, A, P, R, H, s_0)$ where S is the finite state space, A the finite action space, $P : S \times A \to \Delta(S)$ the transition function ($\Delta(S)$ denotes the set of all probability distributions over S), $R : S \times A \to \mathbb{R}$ the reward function, H the finite horizon, and s_0 the initial state. The state space is partitioned into disjoint sets by the time step, $S = \bigcup_{h=0}^{H} S_h$, where states in S_h only transition to states in S_{h+1}. The MDP starts in s_0 and ends in S_H. Given a policy $\pi : S \to \Delta(A)$, a random sequence of transitions $\{(s_h, a_h, r_h, s_{h+1})\}_{h=0}^{H-1}$ is generated by $a_h \sim \pi(s_h), r_h = R(s_h, a_h), s_{h+1} \sim P(s_h, a_h)$. 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 $s \in S_h$. The optimal policy π_M^* maximizes V_M^{π} for all $s \in S$, with value function $V_M^{\pi_M^*}$ abbreviated as V_M^*.

Superscripts p and r denote the provider and recipient, respectively. Thus, the provider's MDP is M^{p}, and the recipient's MDP is M^{r}, sharing the horizon $H = H^{\mathrm{p}} = H^{\mathrm{r}}$. We assume that the two MDPs are weakly-coupled in one direction in the sense that the provider's action might affect certain aspects of the recipient's state but not the other way around. As one way to model such an interaction, we adopt the Transition-Decoupled POMDP (TD-POMDP) framework (Witwicki and Durfee 2010). Formally, both the provider's state s^{p} and the recipient's state s^{r} can be factored into state features. The provider can fully control its state features. The recipient's state can be factored as $s^{\mathrm{r}} = (l^{\mathrm{r}}, u)$, where l^{r} is the set of all the recipient's state features locally controlled by the recipient, and u is the set of state features uncontrollable by the recipient but shared with the provider, i.e. $u = s^{\mathrm{p}} \cap s^{\mathrm{r}}$. Formally, the dynamics of the recipient's state is factored as $P^{\mathrm{r}} = (P_l^{\mathrm{r}}, P_u^{\mathrm{r}})$:

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

where the dynamics of u, P_u^{r}, is controlled only by the provider's policy (i.e., it is not a function of a_h^{r}). Prior work refers to P_u^{r} as the *influence* (Witwicki and Durfee 2010; Oliehoek, Witwicki, and Kaelbling 2012) that the provider exerts on the recipient's environment. In this paper, we focus on the setting where u contains a single binary state feature, $u \in \{u^-, u^+\}$, with u initially taking the value of u^-. Intuitively, $u^+ (u^-)$ stands for an enabled (disabled) precondition needed by the recipient, and the provider commits to enabling the precondition. Further, we focus on a scenario where the flipping is permanent (Hindriks and van Riemsdijk 2007; Witwicki and Durfee 2009; Zhang et al. 2016). That is, once feature u flips to u^+, the precondition is permanently established and will not revert back to u^-.

The provider's commitment semantics. Borrowing from the literature (Witwicki and Durfee 2007; Zhang et al. 2016), we define a probabilistic commitment w.r.t. the shared feature u via a tuple $c = (T, p)$, where T is the commitment time and p is the commitment probability. The provider's commitment semantics is to follow a policy π^{p} that, starting from initial state s_0^{p} (in which u is u^-), sets u to u^+ by time step T with at least probability p:

$$\Pr\left(u^+ \in s_T^{\mathrm{p}} | s_0^{\mathrm{p}}, \pi^{\mathrm{p}}\right) \geq p. \tag{1}$$

For a commitment c, let $\Pi^{\mathrm{p}}(c)$ be the set of all possible provider policies respecting the commitment semantics (Eq. (1)). We call commitment *c feasible* if and only if $\Pi^{\mathrm{p}}(c)$ is non-empty. For a given commitment time T, there is a maximum feasible probability $\overline{p}(T) \leq 1$ such that commitment (T, p) is feasible if and only if $p \leq \overline{p}(T)$. This $\overline{p}(T)$ can be computed by solving the modified reward function: +1 where the commitment is realized at T, and 0 otherwise.

Given a feasible c, the provider's optimal policy maximizes the value with its original reward function of its initial state while respecting the commitment semantics:

$$v^{\mathrm{p}}(c) = \max_{\pi^{\mathrm{p}} \in \Pi^{\mathrm{p}}(c)} V_{M^{\mathrm{p}}}^{\pi^{\mathrm{p}}}(s_0^{\mathrm{p}}). \tag{2}$$

We call $v^{\rm p}(c)$ the provider's commitment value function, and $\pi^{\rm p}(c)$ denotes the provider's policy maximizing Eq. (2).

The recipient's commitment modeling. Abstracting the provider's influence using a single time/probability pair reduces the complexity and communication between the two agents, and prior work has also shown that such abstraction, by leaving other time steps unconstrained, helps the provider handle uncertainty in its environment (Zhang et al. 2016; Zhang, Durfee, and Singh 2020c). Specifying just a single time/probability pair, however, increases the uncertainty of the recipient. Given commitment c, the recipient creates an approximation $\widehat{P}^{\rm r}_u(c)$ of influence $P^{\rm r}_u$, where $\widehat{P}^{\rm r}_u(c)$ hypothesizes the flipping probabilities at other timesteps. Formally, given $\widehat{P}^{\rm r}_u(c)$, let $\widehat{M}^{\rm r}(c)$ be the recipient's approximate model that differs from $M^{\rm r}$ only in terms of the dynamics of u. The recipient's value of commitment c is defined to be the optimal value of the initial state in $\widehat{M}^{\rm r}(c)$:

$$v^{\rm r}(c) = \max\nolimits_{\pi^{\rm r}\in\Pi^{\rm r}} V^{\pi^{\rm r}}_{\widehat{M}^{\rm r}(c)}(s^{\rm r}_0). \tag{3}$$

We call $v^{\rm r}(c)$ the recipient's commitment value function, and $\pi^{\rm r}(c)$ the recipient's policy maximizing Eq. (3) .

Previous work (Witwicki and Durfee 2010; Zhang, Durfee, and Singh 2020a) has chosen an intuitive and straightforward strategy for the recipient to create $\widehat{P}^{\rm r}_u(c)$, which models the flipping with a single branch at the commitment time with the commitment probability. In this paper, we adopt this commitment modeling strategy in Eq. (3) for the recipient, where the strategy determines the transition function of $\widehat{M}^{\rm r}(c)$ through $\widehat{P}^{\rm r}_u(c)$.

The optimal commitment. Let $\mathcal{T} = \{1, 2, ..., H\}$ be the space of possible commitment times, $[0, 1]$ be the continuous commitment probability space, and $v^{\rm p+r} = v^{\rm p} + v^{\rm r}$ be the joint commitment value function. The optimal commitment is a feasible commitment that maximizes the joint value, i.e.

$$c^* = \arg\max\nolimits_{\text{feasible } c\in\mathcal{T}\times[0,1]} v^{\rm p+r}(c). \tag{4}$$

Since commitment feasibility is a constraint for all our optimization problems, for notational simplicity we omit it for the rest of this paper. A naïve strategy for solving the problem in Eq. (4) is to discretize the commitment probability space, and evaluate every feasible commitment in the discretized space. The finer the discretization is, the better the solution will be. At the same time, the finer the discretization, the larger the computational cost of evaluating all the possible commitments. Next, we prove structural properties of the provider's and the recipient's commitment value functions that enable us to develop algorithms that *efficiently* search for the exact optimal commitment.

4 Commitment Space Structure

4.1 Properties of the Commitment Values

We show that, as functions of the commitment probability, both commitment value functions are monotonic and piecewise linear; the provider's commitment value function is concave, and the recipient's is convex. We provide proof sketches for Theorems 1 and 2 on these properties. Formal proofs of all the theorems and the lemmas are included in the full version of this paper (Zhang, Durfee, and Singh 2020b).

Theorem 1. Let $v^{\rm p}(c) = v^{\rm p}(T, p)$ be the provider's commitment value as defined in Eq. (2). For any fixed commitment time T, $v^{\rm p}(T, p)$ is monotonically non-increasing, concave, and piecewise linear in p.

Proof of monotonicity. By the commitment semantics of Eq. (1), $\Pi^{\rm p}(c) = \Pi^{\rm p}(T, p)$ is monotonically non-increasing in p for any fixed T, i.e. $\Pi^{\rm p}(T, p') \subseteq \Pi^{\rm p}(T, p)$ for any $p' > p$. Therefore, $v^{\rm p}(T, p)$ is monotonically non-increasing in p. □

Proof of concavity. Consider the linear program (LP), patterned on the literature (Altman 1999; Witwicki and Durfee 2007), solving the provider's problem in Eq. (2):

$$\max\nolimits_x \ \sum\nolimits_{s^{\rm p},a^{\rm p}} x(s^{\rm p}, a^{\rm p}) R^{\rm p}(s^{\rm p}, a^{\rm p}) \tag{5}$$

$$\text{s.t.}\ \forall s^{\rm p}, a^{\rm p} \quad x(s^{\rm p}, a^{\rm p}) \geq 0; \tag{6}$$

$$\forall s^{\rm p\prime} \quad \sum\nolimits_{a^{\rm p\prime}} x(s^{\rm p\prime}, a^{\rm p\prime}) \tag{7}$$

$$= \sum\nolimits_{s^{\rm p},a^{\rm p}} x(s^{\rm p}, a^{\rm p}) P^{\rm p}(s^{\rm p\prime}|s^{\rm p}, a^{\rm p}) + \delta(s^{\rm p\prime}, s^{\rm p}_0);$$

$$\sum\nolimits_{s^{\rm p}\in s^+_T} \sum\nolimits_{a^{\rm p}} x(s^{\rm p}, a^{\rm p}) \geq p \tag{8}$$

where $\delta(s^{\rm p\prime}, s^{\rm p}_0)$ is the Kronecker delta that returns 1 when $s^{\rm p\prime} = s^{\rm p}_0$ and 0 otherwise, and $s^+_T = \{s^{\rm p} : s^{\rm p} \in S^{\rm p}_T, u^+ \in s^{\rm p}\}$ is the subset of the provider's states at commitment time T in which $u = u^+$. If x satisfies constraints (6) and (7), then it is the occupancy measure of policy $\pi^{\rm p}$, $\pi^{\rm p}(a^{\rm p}|s^{\rm p}) = x(s^{\rm p}, a^{\rm p})/\sum_{a^{\rm p\prime}} x(s^{\rm p}, a^{\rm p\prime})$, where $x(s^{\rm p}, a^{\rm p})$ is the expected number of times action $a^{\rm p}$ is taken in state $s^{\rm p}$ by following policy $\pi^{\rm p}$. Constraint (8) is the commitment semantics of Eq. (1). The expected cumulative reward is in the objective function (5). Therefore, $v^{\rm p}(c)$ is the optimal value of this LP.

For a fixed commitment time T and any two commitment probabilities p and p', let $x^*_p, x^*_{p'}$ be the optimal solutions to the LP, respectively. For any $\eta \in [0, 1]$, let $p_\eta = \eta p' + (1 - \eta)p$. Consider x_η that is the η-interpolation of $x^*_p, x^*_{p'}$,

$$x_\eta(s^{\rm p}, a^{\rm p}) = \eta x^*_{p'}(s^{\rm p}, a^{\rm p}) + (1 - \eta) x^*_p(s^{\rm p}, a^{\rm p}).$$

Note that x_η satisfies constraints (6) and (7), and so it is the occupancy measure of policy $\pi^{\rm p}_\eta$ defined as $\pi^{\rm p}_\eta(a^{\rm p}|s^{\rm p}) = x_\eta(s^{\rm p}, a^{\rm p})/\sum_{a^{\rm p}} x_\eta(s, a^{\rm p})$. Since the occupancy measure of $\pi^{\rm p}_\eta$ is the η-interpolation of x^*_p and $x^*_{p'}$, it is easy to verify that $\pi^{\rm p}_\eta$ is feasible for commitment probability p_η. Therefore, the concavity holds. □

Proof sketch of piecewise linearity. It is well known (Luenberger and Ye 1984) that an optimal solution for a linear program can always be found in the extreme points (or basic feasible solutions). Intuitively, the extreme points move linearly with the commitment probability so that the optimal objective moves piecewise linearly. □

We introduce Assumption 1 that formalizes the notion that u^+, as opposed to u^-, is the value of u that is desirable for the recipient, and then state the properties of the recipient's commitment value function in Theorem 2.

Assumption 1. Let $M^{\rm r+}(M^{\rm r-})$ be defined as the recipient's MDP identical to $M^{\rm r}$ except that u is always set to $u^+(u^-)$. For any $M^{\rm r}$ and any locally-controlled feature $l^{\rm r}$, letting $s^{\mathrm{r}+} = (l^{\mathrm{r}}, u^+)$ and $s^{\mathrm{r}-} = (l^{\mathrm{r}}, u^-)$, we assume $V^*_{M^{\mathrm{r}-}}(s^{\mathrm{r}-}) \leq V^*_{M^{\mathrm{r}+}}(s^{\mathrm{r}+})$.

Theorem 2. Let $v^{\mathrm{r}}(c) = v^{\mathrm{r}}(T,p)$ be the recipient's commitment value as defined in Eq. (3). For any fixed commitment time T, under Assumption 1, $v^{\mathrm{r}}(T,p)$ is monotonically non-decreasing, convex, and piecewise linear in p.

Proof sketch of monotonicity. We fix the commitment time T. For any recipient policy π^{r}, let $v^{\pi^{\mathrm{r}}}_{T,1}$ be the initial state value of π^{r} when u is flipped from u^- to u^+ with probability 1 at T, and let $v^{\pi^{\mathrm{r}}}_{T,0}$ be the initial state value of π^{r} when u never flips to u^+. It is useful to notice that

$$V^{\pi^{\mathrm{r}}}_{\widehat{M}^{\mathrm{r}}(c)}(s^{\mathrm{r}}_0) = p v^{\pi^{\mathrm{r}}}_{T,1} + (1-p) v^{\pi^{\mathrm{r}}}_{T,0} \tag{9}$$

In words, the initial state value can be expressed as the weighted sum of the two scenarios, with the weight determined by the commitment probability. Consider the optimal policy $\pi^*_{\widehat{M}^{\mathrm{r}}(c)}$ for $\widehat{M}^{\mathrm{r}}(c)$. It is guaranteed that $v_{T,1}^{\pi^*_{\widehat{M}^{\mathrm{r}}(c)}} \geq v_{T,0}^{\pi^*_{\widehat{M}^{\mathrm{r}}(c)}}$ because, intuitively, u^+ is more desirable than u^- to the recipient, which leads to $v^{\mathrm{r}}(T,p) \leq v^{\mathrm{r}}(T,p')$ if $p' > p$. □

Proof of convexity and piecewise linearity. Let Π^{r}_D be the set of all the recipient's deterministic policies. It is well known (Puterman 2014) that the optimal value can be attained by a deterministic policy,

$$v^{\mathrm{r}}(T,p) = \max_{\pi^{\mathrm{r}} \in \Pi^{\mathrm{r}}_D} V^{\pi^{\mathrm{r}}}_{\widehat{M}^{\mathrm{r}}(c)}(s^{\mathrm{r}}_0) = \max_{\pi^{\mathrm{r}} \in \Pi^{\mathrm{r}}_D} p v^{\pi^{\mathrm{r}}}_{T,1} + (1-p) v^{\pi^{\mathrm{r}}}_{T,0}$$

which indicates that $v^{\mathrm{r}}(T,p)$ is the maximum of a finite number of value functions that are linear in p. Therefore, $v^{\mathrm{r}}(T,p)$ is convex and piecewise linear in p. □

4.2 Efficient Optimal Commitment Search

As an immediate consequence of Theorems 1 and 2, the joint commitment value is piecewise linear in the probability, and any local maximum for a fixed commitment time T can be attained by a probability at the extremes of zero and $\overline{p}(T)$, or where the slope of the provider's commitment value function changes. We refer to these probabilities as the provider's *linearity breakpoints*, or breakpoints for short. Therefore, one can solve the problem in Eq. (4) to find an optimal commitment by searching only over these breakpoints, as formally stated in Theorem 3.

Theorem 3. Let $\mathcal{P}(T)$ be the provider's breakpoints for a fixed commitment time T. Let $\mathcal{C} = \{(T,p) : T \in \mathcal{T}, p \in \mathcal{P}(T)\}$ be the set of commitments in which the probability is a provider's breakpoint. We have

$$\max\nolimits_{c \in \mathcal{T} \times [0,1]} v^{\mathrm{p+r}}(c) = \max\nolimits_{c \in \mathcal{C}} v^{\mathrm{p+r}}(c).$$

Further, the property of convexity/concavity assures that, for any commitment time, the commitment value function is linear in a probability interval $[p_l, p_u]$ if and only if the value of an intermediate commitment probability $p_m \in (p_l, p_u)$ is the linear interpolation of the two extremes. This enables us to adopt the binary search procedure in Algorithm 1 to efficiently identify the provider's breakpoints. For any fixed commitment time T, the strategy first computes the maximum feasible probability $\overline{p}(T)$. Beginning with the entire interval of $[p_l, p_u] = [0, \overline{p}(T)]$, it recursively checks the linearity of an interval by checking the middle point, $p_m = (p_l + p_u)/2$. The recursion continues with the two halves, $[p_l, p_m]$ and $[p_m, p_u]$, only if the commitment value function is verified to be nonlinear in interval $[p_l, p_u]$. Stepping through $T \in [H]$ and doing the above binary search for each will find all probability breakpoint commitments $\mathcal{C}$.

```
Algorithm 1: Binary search for breakpoints
  Input: The provider's $M^{\mathrm{p}}$, commitment time $T$.
  Output: $\mathcal{P}(T)$: the provider's breakpoints for $T$.
  $\overline{p}(T) \leftarrow$ the maximum feasible probability for $T$
  q $\leftarrow$ A FIFO queue of probability intervals
  q.push($[0, \overline{p}(T)]$)
  Compute and save the provider's commitment value
    for $p = 0, \overline{p}(T)$, i.e. $v^{\mathrm{p}}(T, 0)$ and $v^{\mathrm{p}}(T, \overline{p}(T))$
  Initialize $\mathcal{P}(T) \leftarrow \{\}$
  while q not empty do
      $[p_l, p_u] \leftarrow$ q.pop(); $\mathcal{P}(T) \leftarrow \mathcal{P}(T) \cup \{p_l, p_u\}$
      $p_m \leftarrow (p_l + p_u)/2$; compute and save $v^{\mathrm{p}}(T, p_m)$
      if $v^{\mathrm{p}}(T, p_m)$ is not the linear interpolation of
        $v^{\mathrm{p}}(T, p_l)$ and $v^{\mathrm{p}}(T, p_u)$ then
          q.push($[p_l, p_m]$); q.push($[p_m, p_u]$)
      end
  end
```

This allows for an efficient *centralized* procedure to search for the optimal commitment: construct $\mathcal{C}$ as just described, compute the value of each $c \in \mathcal{C}$ for both the provider and recipient, and return the c with the highest summed value. We will use it to benchmark the decentralized algorithms we develop in Section 5.

5 Commitment Queries

We now develop a querying approach for eliciting the jointly-preferred (cooperative) commitment in a decentralized setting where neither agent has full knowledge about the other's environment. In our querying approach, one agent poses a *commitment query* consisting of information about a set of feasible commitments, and the other responds by selecting the commitment from the set that best satisfies their joint preferences. To limit communication cost and response time, the set of commitments in the query is often small. A query poser thus should optimize its choices of commitments to include, and the responder's choice should reflect joint value. In general, either the provider or recipient could be responsible for posing the query, and the other for responding, and in future work we will consider how these roles could be dynamically assigned. In this paper, though, we always assign the provider to be the query poser and the recipient to be the responder. We do this because the agents must assuredly be able to adopt the responder's selected choice, which means it must be feasible, and per Section 3, only the provider knows which commitments are feasible.

Specifically, we consider a setting where the provider fully knows its MDP, and where its uncertainty about the recipient's MDP is modeled as a distribution μ over a finite set of N candidate MDPs containing the recipient's true MDP. Given uncertainty μ, the Expected Utility (EU) of a feasible commitment c is defined as :

$$EU(c;\mu) = \mathbb{E}_\mu\left[v^{\mathrm{p+r}}(c)\right], \tag{10}$$

where the expectation is w.r.t. the uncertainty about the recipient's MDP. If the provider had to singlehandedly select a commitment based on its uncertainty μ, the best commitment is the one that maximizes the expected utility:

$$c^*(\mu) = \arg\max_c EU(c;\mu). \tag{11}$$

But through querying, the provider is given a chance to refine its knowledge about the recipient's actual MDP. Formally, the provider's commitment query $\mathcal{Q}$ consists of a finite number $k = |\mathcal{Q}|$ of feasible commitments. The provider offers these choices to the recipient, where the provider also annotates each choice with the expected local value of its optimal policy respecting the commitment (Eq. (2)). The recipient computes (using Eq. (3)) its own expected value for each commitment offered in the query, and adds that to the annotated value from the provider. It responds with the commitment that maximizes the summed value (with ties broken by selecting the smallest indexed) to be the commitment the two agents agree on. Therefore, our motivation for a small query size k is two-fold: it avoids large communication cost; and it induces short response time of the recipient evaluating each commitment in the query.

More formally, let $\mathcal{Q} \rightsquigarrow c$ denote the recipient's response that selects $c \in \mathcal{Q}$. With the provider's prior uncertainty μ, the posterior distribution given the response is denoted as $\mu \mid \mathcal{Q} \rightsquigarrow c$, which can be computed by Bayes' rule. When the query size $k = |\mathcal{Q}|$ is limited, the response usually cannot fully resolve the provider's uncertainty. In that case, the value of a query $\mathcal{Q}$ is the EU with respect to the posterior distribution averaged over all the commitments in the query being a possible response, and, consistent with prior work (Viappiani and Boutilier 2010), we refer to it as the query's Expected Utility of Selection (EUS):

$$EUS(\mathcal{Q};\mu) = \mathbb{E}_{\mathcal{Q}\rightsquigarrow c;\mu}\left[EU(c;\mu \mid \mathcal{Q} \rightsquigarrow c)\right].$$

Here, the expectation is with respect to the recipient's response under μ. The provider's *querying problem* thus is to formulate a query $\mathcal{Q} \subseteq \mathcal{T} \times [0,1]$ consisting of $|\mathcal{Q}| = k$ feasible commitments that maximizes EUS:

$$\max\nolimits_{\mathcal{Q}\subseteq\mathcal{T}\times[0,1],|\mathcal{Q}|=k} EUS(\mathcal{Q};\mu). \tag{12}$$

Importantly, we can show that $EUS(\mathcal{Q};\mu)$ is a submodular function of $\mathcal{Q}$, as formally stated in Theorem 4. Submodularity serves as the basis for a greedy optimization algorithm (Nemhauser, Wolsey, and Fisher 1978), which we describe after Theorem 5.

Theorem 4. For any uncertainty μ, $EUS(\mathcal{Q};\mu)$ is a submodular function of $\mathcal{Q}$.

Submodularity means that adding a commitment to the query can increase the EUS, but the increase is diminishing with the size of the query. An upper bound on the EUS of any query of any size k can be obtained when $k \geq N$ such that the query can include the optimal commitment of each candidate recipient's MDP, i.e.

$$\overline{EUS} = \mathbb{E}_\mu\left[\max\nolimits_{c\in\mathcal{T}\times[0,1]} v^{\mathrm{p+r}}(c)\right]. \tag{13}$$

As the objective of Eq. (12) increases with size k, in practice the agents could choose size k large enough to meet some predefined EUS. We will empirically investigate the effect of the choice of k in Section 6.

Structure of the Commitment Query Space. Due to the properties of individual commitment value functions proved in Section 4, the expected utility $EU(c;\mu)$ defined in Eq. (10), as calculated by the provider alone, becomes a summation of the non-increasing provider's commitment value function and the (provider-computed) weighted average of the non-decreasing recipient's commitment value functions. With the same reasoning as for Theorem 3, the optimality of the breakpoint commitments can be generalized to any uncertainty, as formalized in Lemma 1.

Lemma 1. Let $\mathcal{C}$ be defined as in Theorem 3. We have $\max_{c\in\mathcal{T}\times[0,1]} EU(c;\mu) = \max_{c\in\mathcal{C}} EU(c;\mu)$.

As a consequence of Lemma 1, for EUS maximization, there is no loss in only considering the provider's breakpoints, as formally stated in Theorem 5.

Theorem 5. For any query size k and uncertainty μ, we have

$$\max_{\mathcal{Q}\subseteq\mathcal{T}\times[0,1],|\mathcal{Q}|=k} EUS(\mathcal{Q};\mu) = \max_{\mathcal{Q}\subseteq\mathcal{C},|\mathcal{Q}|=k} EUS(\mathcal{Q};\mu).$$

Theorem 5 enables an efficient procedure for solving the query formulation problem (Eq. (12)). The provider first identifies its breakpoint commitments $\mathcal{C}$ and evaluates them for its MDP and each of the N recipient's possible MDPs. Due to the concavity and convexity properties, $\mathcal{C}$ can be identified and evaluated efficiently with the binary search strategy we described in Section 4.2. Finally, a size k query is formulated from commitments $\mathcal{C}$ that solves the EUS maximization problem either exactly with exhaustive search, or approximately with greedy search (Viappiani and Boutilier 2010; Cohn, Singh, and Durfee 2014). The greedy search begins with $\mathcal{Q}_0$ as an empty set and iteratively performs $\mathcal{Q}_i \leftarrow \mathcal{Q}_{i-1} \cup \{c_i\}$ for $i = 1, ..., k$, where $c_i = \arg\max_{c\in\mathcal{C},c\notin\mathcal{Q}_{i-1}} EUS(\mathcal{Q}_{i-1} \cup \{c\};\mu)$. Since EUS is a submodular function of the query (Theorem 4), the greedily-formed size k query $\mathcal{Q}_k$ is within a factor of $1-(\frac{k-1}{k})^k$ of the optimal EUS (Nemhauser, Wolsey, and Fisher 1978).

6 Empirical Evaluation

Our empirical evaluations focus on these questions:

- For EUS maximization, how effective and efficient is the breakpoints discretization compared with alternatives?
- For EUS maximization, how effective and efficient is greedy query search compared with exhaustive search?

To answer these questions, in Section 6.1, we conduct empirical evaluations in synthetic MDPs with minimal assumptions on the structure of transition and reward functions, and we use an environment in Section 6.2 inspired by the video game of Overcooked to evaluate the breakpoints discretization and the greedy query search in this more grounded and structured domain.

6.1 Synthetic MDPs

The provider's environment is a randomly-generated MDP. It has 10 states the provider can be in at any time step, one of which is an absorbing state denoted as s^+, and where the initial state is chosen from the non-absorbing states. Feature u takes the value of u^+ only in the absorbing state, i.e. $u^+ \in s^{\mathrm{p}}$ if and only if $s^{\mathrm{p}} = s^+$. There are 3 actions. For each state-action pair $(s^{\mathrm{p}}, a^{\mathrm{p}})$ where $s^{\mathrm{p}} \neq s^+$, the transition function $P^{\mathrm{p}}(\cdot|s^{\mathrm{p}}, a^{\mathrm{p}})$ is determined independently by filling the 10 entries with values uniformly drawn from $[0, 1]$, and normalizing $P^{\mathrm{p}}(\cdot|s^{\mathrm{p}}, a^{\mathrm{p}})$. The reward $R^{\mathrm{p}}(s^{\mathrm{p}}, a^{\mathrm{p}})$ for a non-absorbing state $s^{\mathrm{p}} \neq s^+$ is sampled uniformly and independently from $[0, 1]$, and for the absorbing state $s^{\mathrm{p}} = s^+$ is zero. Thus, the random MDPs are intentionally generated to introduce a tension for the provider between helping the recipient (but getting no further local reward) versus accumulating more local reward. (Our algorithms also work fine in cases without this tension, but the commitment search is less interesting without it because no compromise is needed.)

The recipient's environment is a one-dimensional space with 10 locations represented as integers $\{0, 1, ..., 9\}$. In locations $1 - 8$, the recipient can move right, left, or stay still. Once the recipient reaches either end (location 0 or 9), it stays there. There is a gate between locations 0 and 1 for which $u = u^+$ denotes the state of open and $u = u^-$ closed. Initially, the gate is closed and the recipient starts at an initial location L_0. A negative reward of -10 is incurred by bumping into the closed gate. For each time step the recipient is at neither end, it gets a reward of -1. If it reaches the left end (i.e. location 0), it gets a one-time reward of $r_0 > 0$. The recipient gets a reward of 0 if it reaches the right end. In a specific instantiation, L_0 and r_0 are fixed. L_0 is randomly chosen from locations $1 - 8$ and r_0 from interval $(0, 10)$ to create various MDPs for the recipient.

To generate a random coordination problem, we sample an MDP for the provider, and N candidate MDPs for the recipient, setting the provider's prior uncertainty μ over the recipient's MDP to be the uniform distribution over the N candidates. The horizon for both agents is set to be 20. Since the left end has higher rewards than the right end, if the recipient's start position is close enough to the left end and the provider commits to opening the gate early enough with high enough probability, the recipient should utilize the commitment by checking if the gate is open by the commitment time, and pass through it if so; otherwise, the recipient should simply ignore the commitment and move to the right end. The distribution for generating the recipient's MDPs is designed to include diverse preferences regarding the commitments, such that the provider's query should be carefully formulated to elicit the recipient's preference.

The principal result from Section 4 was that the commitment probabilities to consider can be restricted to breakpoints without loss of optimality. Further, the hypothesis was that the space of breakpoints would be relatively small, allowing the search to be faster. We now empirically confirm the optimality result, and test the hypothesis of greater efficiency, by comparing the breakpoint commitments discretization to the following alternative discretizations:

- *Even discretization.* Prior work (Witwicki and Durfee 2007) discretizes the probability space up to a certain granularity. Here, the probability space $[0, 1]$ is evenly discretized as $\{p_0, ..., p_n\}$ where $p_i = \frac{i}{n}$.
- *Deterministic Policy (DP) discretization.* This discretization finds all of the probabilities of toggling feature u at the commitment time that can be attained by the provider following a deterministic policy (Witwicki and Durfee 2007, 2010).

For the even discretization, we consider the resolutions $n \in \{10, 20, 50\}$. For DP, we found that the number of toggling probabilities of all the provider's deterministic policies is large, and the corresponding computational cost of identifying and evaluating them is high. To reduce the computational cost and for fair comparison, we group the probabilities in the DP discretization that are within $\frac{i}{n}$ of each other for $n \in \{10, 20, 50\}$. Since the problem instances have different reward scales, to facilitate analyses we normalize for each instance the EUS with the upper bound $\overline{EUS}$ defined in Eq. (13) and the EUS of the optimal and greedy query of the even discretization for $k = 1, n = 10$.

Figure 1 gives the EUS for the seven discretizations over 50 randomly-generated problem instances, for $N = 10$ candidate MDPs for the recipient and $k = 2$ and 5. Figure 1 shows that, coupled with the greedy query algorithm, our breakpoint commitments discretization yields the highest EUS with the lowest computational cost. In Figure 1(left), we see that, for the even and the DP discretizations, the EUS increases with the probability resolution n, and only once we reach $n = 50$ is the EUS comparable to our breakpoints discretization. Figure 1(right) compares the runtimes of forming the discretization and evaluating the commitments in the discretization for the downstream query formulation proce-

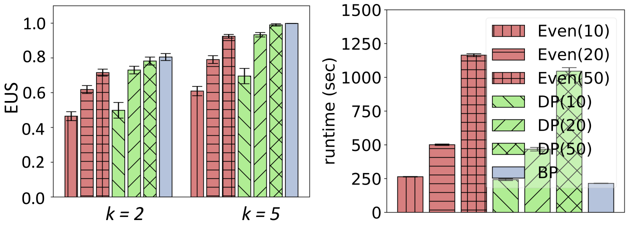

Figure 1: Means and standard errors of the EUS (left) and runtime (right) of the discretizations in Synthetic MDPs.

	$n = 10$	$n = 20$	$n = 50$
Even	7.8 ± 0.1	15.1 ± 0.1	37.1 ± 0.1
DP	6.1 ± 0.2	12.0 ± 0.3	26.5 ± 0.7
Breakpoints	10.0 ± 0.1		

Table 1: Averaged discretization size per commitment time (mean and standard error) in Synthetic MDPs.

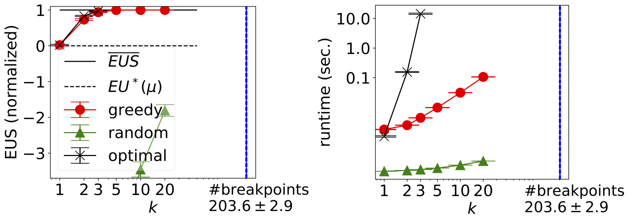

Figure 2: Means and standard errors of the EUS (left) and runtime (right) of the optimal, the greedy, and the random queries formulated from the breakpoints in Synthetic MDPs.

dure, confirming the hypothesis that using breakpoints is faster. Table 1 compares the sizes of these discretizations, and confirms our intuition that the breakpoints discretization is most efficient because it identifies fewer commitments that are sufficient for the EUS maximization.

Next, we empirically confirm the greedy query search is effective for EUS maximization. Given the results confirming the effectiveness and efficiency of the breakpoint discretization, the query searches here are over the breakpoint commitments. Figure 2(left) compares the EUS of the greedily-formulated query with the optimal (exhaustive search) query, and with a query comprised of randomly-chosen breakpoints. The EUS is normalized with $\overline{EUS}$ and the optimal EU prior to querying given uncertainty μ as defined in Eq. (11). We vary the query size k, and report means and standard errors over the same 50 coordination problems. We see that the EUS of the greedy query tracks that of the optimal query closely, while greedy's runtime scales much better.

6.2 Overcooked

We further test our approach in a more grounded domain, Overcooked, introduced by (Wang et al. 2020). We reuse one of their Overcooked settings with two high-level modifications: 1) instead of having global observability, each agent observes only its local environment, and 2) we introduce probabilistic transitions. These modifications induce for the domain a rich space of meaningful commitments, over which the agents should carefully negotiate for the optimal cooperative behavior. Figure 3 illustrates this Overcooked environment. Two agents, the chef and the waiter, together occupy a grid with counters being the boundaries. The chef is supposed to pick up the tomato, chop it, and place it on the plate. Afterwards, the waiter is supposed to pick up the chopped tomato and deliver to the counter labelled by the star. Meanwhile, the chef needs to take care of the pot that can probabilistically begin boiling, and the waiter needs to take care of a dine-in customer (labelled by the plate with fork and knife). This introduces interesting tensions between delivering the food and taking care of the pot and the customer. For coordination, the chef makes a probabilistic commitment that it will place the chopped tomato on the plate, which makes the chef the provider and the waiter the recipient. Crucially, the commitment decouples the agents' planning problems, allowing the agents to only model the MDP in their half of the grid. We repeat the experiments in Section 6.1 that evaluate the breakpoints discretization and the greedy query over 50 problem instances. Table 2 shows that the provider's more structured transition in Overcooked leads to even greater efficiency (fewer breakpoints) than in the synthetic MDPs (Table 1). As before, the Greedy query closely tracks optimal. We give more details in the full version (Zhang, Durfee, and Singh 2020b), along with a comparison between our approach and multiagent planning techniques previously studied in Overcooked, where we saw that, even with small queries ($k = 2$), our decentralized query-based probabilistic commitment approach got within 99.5% of the optimal value achieved through a centrally constructed and executed joint policy.

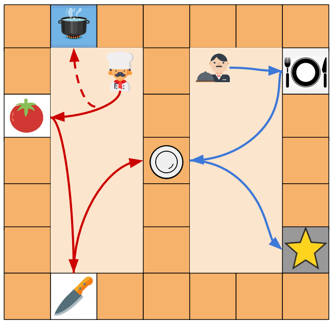

Figure 3: Overcooked.

	$n = 10$	$n = 20$	$n = 50$
Even	6.4 ± 0.1	11.8 ± 0.3	28.0 ± 0.7
DP	5.2 ± 0.2	8.5 ± 0.4	16.4 ± 0.9
Breakpoints	4.9 ± 0.2		

Table 2: Averaged discretization size per commitment time (mean and standard error) in Overcooked.

7 Discussion

Built on provable foundations and evaluated in two separate domains, our approach proves highly appropriate for settings where cooperative agents coordinate their plans through commitments in a decentralized manner, and could provide a good performance/cost tradeoff even compared to coordination that is not restricted to being commitment-based.

For future directions, if the agents can afford the time and bandwidth, querying need not be limited to a single round, which then raises questions about how agents should consider future rounds when deciding on what to ask in the current round. The querying can also be extended to the setting where the query poser is uncertain about both the responder's and its own environments. As dependencies between agents get richer (with chains and even cycles of commitments), continuing to identify and exploit structure in intertwined value functions will be critical to scaling up for efficient multi-round querying of connected commitments.

Acknowledgments

We thank the anonymous reviewers for their valuable feedback. This work was supported in part by the Air Force Office of Scientific Research under grant FA9550-15-1-0039. Opinions, findings, conclusions, or recommendations expressed here are those of the authors and do not necessarily reflect the views of the sponsor.

References

Agotnes, T.; Goranko, V.; and Jamroga, W. 2007. Strategic commitment and release in logics for multi-agent systems. Technical Report IfI-08-01, Clausthal University.

Aknine, S.; Pinson, S.; and Shakun, M. F. 2004. An extended multi-agent negotiation protocol. *Autonomous Agents and Multi-Agent Systems* 8(1): 5–45.

Al-Saqqar, F.; Bentahar, J.; Sultan, K.; and El-Menshawy, M. 2014. On the interaction between knowledge and social commitments in multi-agent systems. *Applied Intelligence* 41(1): 235–259.

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

Baldoni, M.; Baroglio, C.; Chopra, A. K.; and Singh, M. P. 2015. Composing and verifying commitment-based multiagent protocols. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, 10–17.

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.

Boutilier, C. 2002. A POMDP formulation of preference elicitation problems. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, 239–246.

Castelfranchi, C. 1995. Commitments: From Individual intentions to groups and organizations. In *Proceedings of the International Conference on Multiagent Systems*, 41–48.

Chajewska, U.; Koller, D.; and Parr, R. 2000. Making rational decisions using adaptive utility elicitation. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence*, 363–369.

Chesani, F.; Mello, P.; Montali, M.; and Torroni, P. 2013. Representing and monitoring social commitments using the event calculus. *Autonomous Agents and Multi-Agent Systems* 27(1): 85–130.

Cohen, P. R.; and Levesque, H. J. 1990. Intention is choice with commitment. *Artificial Intelligence* 42(2-3): 213–261.

Cohn, R.; Singh, S.; and Durfee, E. 2014. Characterizing EVOI-sufficient k-response query sets in decision problems. In *International Conference on Artificial Intelligence and Statistics*, 131–139.

Dastani, M.; van der Torre, L. W. N.; and Yorke-Smith, N. 2017. Commitments and interaction norms in organisations. *Auton. Agents Multi Agent Syst.* 31(2): 207–249.

Fornara, N.; and Colombetti, M. 2008. Specifying and enforcing norms in artificial institutions. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems*, 1481–1484.

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

Han, T. A.; Pereira, L. M.; and Lenaerts, T. 2017. Evolution of commitment and level of participation in public goods games. *Auton. Agents Multi Agent Syst.* 31(3): 561–583.

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.

Kraus, S. 1997. Negotiation and cooperation in multi-agent environments. *Artificial intelligence* 94(1-2): 79–97.

Kushmerick, N.; Hanks, S.; and Weld, D. 1994. An algorithm for probabilistic least-commitment planning. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, 1073–1078.

Luenberger, D. G.; and Ye, Y. 1984. *Linear and nonlinear programming*. Springer.

Mallya, A. U.; and Huhns, M. N. 2003. Commitments among agents. *IEEE Internet Computing* 7(4): 90–93.

Nair, R.; Tambe, M.; Yokoo, M.; Pynadath, D.; and Marsella, S. 2003. Taming decentralized POMDPs: Towards efficient policy computation for multiagent settings. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*, volume 3, 705–711.

Nemhauser, G. L.; Wolsey, L. A.; and Fisher, M. L. 1978. An analysis of approximations for maximizing submodular set functions. *Mathematical Programming* 14(1): 265–294.

Oliehoek, F. A.; Amato, C.; et al. 2016. A concise introduction to decentralized POMDPs. *Springer Briefs in Intelligent Systems* .

Oliehoek, F. A.; Witwicki, S. J.; and Kaelbling, L. P. 2012. Influence-based abstraction for multiagent systems. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, 1422–1428.

Pereira, R. F.; Oren, N.; and Meneguzzi, F. 2017. Detecting commitment abandonment by monitoring sub-optimal steps during plan execution. In *Proceedings of the 16th Conference on Autonomous Agents and Multiagent Systems*, 1685–1687.

Puterman, M. L. 2014. *Markov Decision Processes.: Discrete Stochastic Dynamic Programming*. John Wiley & Sons.

Rahwan, I. 2004. *Interest-based negotiation in multi-agent systems*. Ph.D. thesis, University of Melbourne, Department of Information Systems Melbourne.

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. London.

Venkatraman, M.; and Singh, M. P. 1999. Verifying compliance with commitment protocols. *Autonomous Agents and Multi-agent Systems* 2(3): 217–236.

Viappiani, P.; and Boutilier, C. 2010. Optimal Bayesian recommendation sets and myopically optimal choice query sets. In *Advances in Neural Information Processing Systems*, 2352–2360.

Vokrínek, J.; Komenda, A.; and Pechoucek, M. 2009. Decommitting in multi-agent execution in non-deterministic environment: experimental approach. In *8th International Joint Conference on Autonomous Agents and Multiagent Systems*, 977–984.

Wang, R. E.; Wu, S. A.; Evans, J. A.; Tenenbaum, J. B.; Parkes, D. C.; and Kleiman-Weiner, M. 2020. Too many cooks: Coordinating multi-agent collaboration through inverse planning. *arXiv preprint arXiv:2003.11778* .

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

Witwicki, S. J.; and Durfee, E. H. 2007. Commitment-driven distributed joint policy search. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems*, 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 *Proceedings of the Twentieth International Conference on Automated Planning and Scheduling*, 185–192.

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

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

Yolum, P.; and Singh, M. P. 2002. Flexible protocol specification and execution: Applying event calculus planning using commitments. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems*, 527–534.

Zhang, K.; Yang, Z.; Liu, H.; Zhang, T.; and Başar, T. 2018. Fully decentralized multi-agent reinforcement learning with networked agents. *arXiv preprint arXiv:1802.08757* .

Zhang, Q.; Durfee, E.; and Singh, S. 2020a. Modeling probabilistic commitments for maintenance is inherently harder than for achievement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 10326–10333.

Zhang, Q.; Durfee, E. H.; and Singh, S. 2020b. Efficient querying for cooperative probabilistic commitments. *arXiv preprint arXiv:2012.07195* .

Zhang, Q.; Durfee, E. H.; and Singh, S. 2020c. Semantics and algorithms for trustworthy commitment achievement under model uncertainty. *Autonomous Agents and Multi-Agent Systems* 34(1): 19.

Zhang, Q.; Durfee, E. H.; Singh, S.; Chen, A.; and Witwicki, S. J. 2016. Commitment semantics for sequential decision making under reward uncertainty. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 3315–3323.

Zhang, S.; Durfee, E.; and Singh, S. 2017. Approximately-optimal queries for planning in reward-uncertain Markov decision processes. In *Proceedings of the Twenty-Seventh International Conference on Automated Planning and Scheduling*, 339–347.