AFRL-AFOSR-VA-TR-2021-0070

Making and Keeping Informed Commitments in Human-Machine Systems

Durfee, Edmund
REGENTS OF THE UNIVERSITY OF MICHIGAN
503 THOMPSON ST
ANN ARBOR, MI, 48109
USA

07/08/2021
Final Technical Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
Air Force Office of Scientific Research
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
08-07-2021

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*
01 May 2015 - 30 Apr 2021

4. TITLE AND SUBTITLE
Making and Keeping Informed Commitments in Human-Machine Systems

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-15-1-0039

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Edmund Durfee

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
REGENTS OF THE UNIVERSITY OF MICHIGAN
503 THOMPSON ST
ANN ARBOR, MI 48109
USA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AF Office of Scientific Research
875 N. Randolph St. Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR RTA2

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-AFOSR-VA-TR-2021-0070

12. DISTRIBUTION/AVAILABILITY STATEMENT
A Distribution Unlimited: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT

As the tempo, scope, and complexity of their mission environments grow, people will become increasingly reliant on delegating some decisions about situation comprehension and response to computational cognitive systems. When such a symbiotic human-machine system is functioning well, the machine will act as an extension of its human user, making decisions that agree with what the user would have done if the user could have devoted enough time and attention to making them. When a user has trust in the cognitive assistance the system provides, the user can focus on the important and time-critical tasks that require human reasoning, and safely ignore tasks delegated to a cognitive system.

Our approach to reliable human-machine interaction that avoids such automation surprise involves two main thrusts. The first thrust is to enable the machine to recognize for itself when its uncertainty about the user’s expectations under the current circumstances justifies asking the user for clarification, and how to pose queries that efficiently acquire the most consequential information from the user about what expectations the system should commit to fulfilling. The second thrust, then, is to formulate principled computational techniques for explicitly modeling and planning in the context of such commitments to the user, to meet the user’s expectations while nonetheless exercising autonomy when responding to evolving mission conditions. Our project’s objectives have been to make fundamental advances along these thrusts of making and keeping commitments separately, and to integrate them to solve the combined problem of querying about what commitments would be best to form given the goals and uncertainty of both the user and the system.

Across its overall period of performance, the project made substantial contributions to the science and engineering of computational techniques for commitment-based multiagent coordination, for human-agent interaction, and for the combination of these. Our research considered two forms of uncertainty that an artificially-intelligent agent might have about the what its human user would want it to do: 1. Reward uncertainty: Given a choice of outcomes the agent could strive to achieve, which outcome(s) are most preferred by the user? 2. Safety uncertainty: Given a choice of actions the agent could take in pursuing outcomes, and the different side-effects actions might have, which potential side-effects does its user consider benign, and which are considered unsafe?

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	UU	8	HAL GREENWALD

19b. TELEPHONE NUMBER *(Include area code)*
703-588-8441

Final Project Report

May 1, 2015 to April 30, 2021

Project Title:
Making and Keeping Informed Commitments in Human-Machine Systems

Grant Number:
FA9550-15-1-0039

Principal Investigators:
Edmund H. Durfee and Satinder Singh Baveja

Institution:
University of Michigan

Project Objectives

As the tempo, scope, and complexity of their mission environments grow, people will become increasingly reliant on delegating some decisions about situation comprehension and response to computational cognitive systems. When such a symbiotic human-machine system is functioning well, the machine will act as an extension of its human user, making decisions that agree with what the user would have done if the user could have devoted enough time and attention to making them. When a user has trust in the cognitive assistance the system provides, the user can focus on the important and time-critical tasks that *require* human reasoning, and safely ignore tasks delegated to a cognitive system.

Unfortunately, missions often take place in worlds where unexpected circumstances can arise due to chance, or to choices made by adversaries. In such cases, there is danger that a cognitive system might respond to the dynamism of the environment in ways that can surprise its human user. When this can happen, the cognitive system can go from being an asset to a liability, by compounding the user's cognitive load by adding yet more uncertainty in the user's environment, in this case uncertainty about how the cognitive system will behave and whether the user will need to intervene in the pursuit of delegated tasks.

Our approach to reliable human-machine interaction that avoids such automation surprise involves two main thrusts. The first thrust is to enable the machine to recognize for itself when its uncertainty about the user's expectations under the current circumstances justifies asking the user for clarification, and how to pose queries that efficiently acquire the most consequential information from the user about what expectations the system should commit to fulfilling. The second thrust, then, is to formulate principled computational techniques for explicitly modeling and planning in the context of such commitments to the user, to meet the user's expectations while nonetheless exercising autonomy when responding to evolving mission conditions. Our project's objectives have been to make fundamental advances along these thrusts of making and keeping commitments separately, and to integrate them to solve the combined problem of querying about what commitments would be best to form given the goals and uncertainty of both the user and the system.

Summary of Significant Accomplishments

Across its overall period of performance, the project made substantial contributions to the science and engineering of computational techniques for commitment-based multiagent coordination, for human-agent interaction, and for the combination of these. Each of these is summarized below, where fuller descriptions of each contribution are available in the references cited.

Querying for Safely-Optimal Planning

Our research considered two forms of uncertainty that an artificially-intelligent agent might have about the what its human user would want it to do:

1. Reward uncertainty: Given a choice of outcomes the agent could strive to achieve, which outcome(s) are most preferred by the user?
2. Safety uncertainty: Given a choice of actions the agent could take in pursuing outcomes, and the different side-effects actions might have, which potential side-effects does its user consider benign, and which are considered unsafe?

Reward Uncertainty [16]. In the earlier years of this project, our research in this thrust focused on querying under reward uncertainty. In particular, to limit cognitive burden on the human user, we considered "multiple choice" queries with a small number, k, of choices. We proved that no k-ary multiple choice query can acquire more useful knowledge than a query composed of k *policies*, and therefore an agent can restrict its search for an optimal k-ary query to only consider policy queries. Further, by proving submodularity properties in the policy-query space, we were able to utilize a previously-developed greedy construction method for composing queries, where the greedy query has a provable bound on its worst-case suboptimality.

However, the previously-developed greedy method assumed that each potential element to add to the query could be individually evaluated in order to pick the best one, which can work for queries over relatively small numbers of options (e.g., movies or restaurants), but not for the intractably-large space of all possible policies. We devised a mixed-integer linear programming (MILP) approach that would instead *construct* the next policy to greedily add, where we incrementally introduced constraints into the MILP such that the next policy it generates optimally complements the policies already included in the query. The resulting greedily-constructed k-ary policy query has a expected posterior utility (EPU) within $1 - [(k-1)/k]^k$ of the optimal query.

Because expressing and answering a policy query can be difficult (a policy specifies one or more actions to take in every possible state), we also considered trajectory queries, which are a more user-friendly form of query that has been used in the literature. In a k-ary trajectory query, the user is presented with k possible limited-length trajectories (sequences of states and actions), and is asked which best represents a trajectory they would follow given their rewards. Our novel strategy for formulating such queries was for the agent to *project* its greedily-constructed approximately-optimal policy query into a trajectory query, where to realize this strategy we needed to devise a method for identifying the right state from which all of the trajectories would begin such that the trajectories would differ in ways that reflect meaningful differences in the policies.

We empirically compared our greedy policy queries and projected trajectory queries against optimal policy queries and against trajectory queries found by other methods in the literature. Our results showed that, in terms of performance, the greedy policy query is nearly as good as the optimal query, and that while projection into a trajectory query sacrifices some performance, the resultant query still significantly outperforms trajectory queries found by other methods in the literature. Computationally, the greedy policy and projected trajectory queries are found much faster than finding the optimal policy query, as expected [16].

Safety Uncertainty [17, 18, 19, 20]. Our subsequent work in this vein considered the situation where the agent has reward certainty—it knows the goal(s) that the human user wants achieved—but where the agent's possible policies to achieve the goal(s) can have side-effects that change other features of the environment in ways that the user might not expect or want. We call an agent's policy *safe* if all of the possible side-effects that could arise when carrying out the policy are known to be acceptable to the user. In general, then, the goal of the agent is to find and pursue a *safely-optimal policy*: of all of the safe policies, one that maximizes expected cumulative reward.

We began by addressing the problem of how an agent that already has a safe policy can efficiently query about potential side-effects to find a better (closer to optimal) safe policy. To limit cognitive load on the human user, we once again assume queries are k-ary. However, in this case, a query is composed of k features of the environment that could be side-effected, and the user's response is to identify the subset of those that are safe to side-effect (where any not in that subset are now known to be unsafe to side-effect).

In formulating its k-ary query, the agent should carefully select which k of the unknown features (the features that it doesn't yet know are safe or unsafe to change) will in expectation allow it to improve upon its current safe policy the most. We developed a set of techniques for accomplishing this in a computationally tractable way. First, we developed an algorithm that once again exploits an evolving MILP formulation, where this time the algorithm provably finds the set of *dominating policies* (which would be preferred over the current best safe policy if some unknown features were safe to change) and thus the set of *relevant features* to even consider querying about. Then, we developed a Minimax-Regret Query algorithm, MMRQ-k, to find the best k-ary feature query under the assumption that the agent has no knowledge of which features are more or less likely to be considered by the user as being safely changeable. We demonstrated empirically that MMRQ-k indeed finds the optimal k-ary feature query, and does so rapidly [17].

We then turned to the harder problem of querying in the face of safety uncertainty when the agent initially has no safe policy—where every policy the agent can identify for achieving the user's goal(s) involve side-effecting one or more features with unknown safe changeability. For this kind of problem, restricting the agent to a single k-ary query, as we had so far done, is insufficient, if the goal is for the agent to find at least some safe policy to execute. Thus, for this form of querying, the agent asks the user about one feature at a time until one of two things happens. One possibility is that the agent now knows of enough safe side-effects that it can formulate a safe policy, after which it can either execute that policy or try to improve upon that policy using MMRQ-k. The other possibility is that the agent now knows that so many of the features are unsafe to change that in fact it is provably the case that no safe policy exists, so there is no point in asking any more queries.

Thus, the problem that the agent needs to solve is to ask about as few unknown features as possible, in expectation, to either find a safe policy or to prove no safe policy exists. We cast each of these subproblems (minimizing queries to find a safe policy, and minimizing queries to prove no safe policy exists) as a different set-cover problem. The challenge we faced was that, since our agent doesn't know which of these problems it will successfully solve, it needs to ask about unknown features that will make progress in solving *both* of these problems *at the same time*. We developed and evaluated several heuristic strategies for selecting the next feature to ask about. Our results showed that, with reasonable computational overhead, our heuristics could approach the performance of an optimal (but computationally infeasible) approach [20].

Reward and Safety Uncertainty [15]. The above work looked at reward uncertainty and safety uncertainty separately. Our last main effort along this thrust of the project was to study what an agent can do when faced with both types of uncertainty simultaneously. We investigated a myopic strategy that simply runs each of the previous algorithms and selects whichever query is better, terminating when neither can find a query with positive expected gain. However, when a positive gain can only come from asking multiple queries, being myopic might abandon querying too early. We thus developed other strategies that focused on querying about all the safety constraints associated with an entire policy (which has a tendency to ask too many queries), and on forming a batch query for multiple policies at once, where that last strategy has the best empirical performance overall [15].

Probabilistic Commitments.

The techniques just summarized from the querying thrust of this project allow an agent to clarify with the user (or some other agent) what changes to the world it should achieve (if they are goals) or avoid (if they are unsafe side-effects). However, in general, an agent cannot be certain about all the future actions it will need to take in its world. For one thing, its actions might have stochastic outcomes. For example, in attempting to move forward, a robot might instead slip sideways, or in attempting to purchase an item an agent might discover the item is out of stock. It could also be that, in the midst of pursuing the user's goals, the agent might discover more important tasks that it should pursue instead. For these kinds of reasons, our research views a commitment that an agent makes to another as being *probabilistic*.

Tractable Commitment-Constrained Autonomy [2, 3, 4, 10, 12, 14]. An advantage of probabilistic commitments is that they recognize that, despite its efforts, an agent might fail to realize the desired outcome of the commitment. For the commitment to be meaningful, however, the agent pursuing it needs to "try" hard enough to satisfy it. A contribution of our work has been to develop the first prescriptive semantics for probabilistic commitments, where a probabilistic commitment constrains the behavior of the commitment provider such that its actions will respect the commitment. Specifically, the commitment provider formulates (and if its model changes over time, reformulates) its local policy so as to maximize its expected reward under the constraint of "trying hard enough" – in expectation reaching a state that satisfies the commitment with at least the probability promised.

We developed a family of planning algorithms that a commitment provider can follow, all of which provably respect the semantics, but each of which makes a different tradeoff in the time and effort invested in responding to changes, and anticipating future changes, to the agent's model that would lead it to want to revise its policy to improve its expected cumulative reward. At one extreme, our Commitment-Constrained Full Lookahead (CCFL) algorithm allows the agent to anticipate all ways in which its model might change, which leads to optimal behavior but involves an exponential blowup that makes it computationally infeasible for realistic problems. At the other extreme, with our Commitment-Constrained No Lookahead (CCNL) algorithm the agent ignores possible future model changes, and singlemindedly pursues the best policy it formed at the time it made the commitment. We developed a parameterized version, the Commitment-Constrained Lookahead (CCL) algorithm that looks at possible model changes over a tunable horizon. Empirically, the overall best approach is our Commitment-Constrained Iterative Lookahead (CCIL) algorithm, that reinvokes the CCL algorithm periodically: each call to CCL can lookahead to a shorter (and hence more computationally acceptable) horizon, but reinvoking it effectively extends the horizon in the subspace that is actually being encountered [10].

Achievement and Maintenance Commitments [6, 7, 9]. Until our research in this project, it had generally been assumed that the provider of a commitment faced a more difficult problem than the recipient of a commitment. After all, it is the provider that is having to deviate from its optimal behavior in order to satisfy the commitment, so finding a way to balance both objectives (maximizing local reward while meeting the commitment) has seemed hardest. However, our deeper investigation into probabilistic commitments has proven that the recipient's problem is in some ways more challenging! This is because the flexibility that the commitment specification leaves the provider in being able to revise its policy in response to changing circumstances while still meeting the commitment, introduces uncertainty for the recipient in terms of the timing at which the commitment will be probabilistically met. Specifically, in the interval of time between when the agents agree on the commitment, and when the commitment's (probabilistic) satisfaction is promised, the recipient doesn't know whether the probability of satisfaction grows linearly over that interval, or in some non-linear way. In general, the recipient will plan differently depending on, for example, whether or not the commitment has a good chance of being satisfied early.

What our research has shown answers a question that has been vexing some researchers, which is the question of why commitments of maintenance (where the provider promises to maintain some feature of the environment in its current state) are so much harder to coordinate well over than are commitments of achievement (where the provider promises to change some feature to a desired state). Our research has proven that, with an achievement commitment, it is possible for the recipient to model it in a "pessimistic" way, where how the provider ends up satisfying the commitment can never be worse than what the recipient modeled. It turns out, though, that no such "pessimistic" model exists for a maintenance commitment. The implications of this can be far reaching, because it points to fundamental tradeoffs in commitment specifications: to improve the recipient's model means a richer commitment representation, which will reduce the flexibility of the provider [9].

Integrating Querying and Commitments [5, 8, 11] Our work on probabilistic commitment semantics for the provider and recipient of the commitment has developed algorithms and modeling strategies that lead to principled coordination behavior between agents given a commitment between them. The last main component in this thrust of our work was to investigate how agents can efficiently converge on what probabilistic commitment of achievement to agree to, in the first place, if their objective is to maximize their expected joint (summed) reward.

In general, the local reward of a recipient increases with a commitment that is promised earlier and with higher probability, while the local reward of a provider rises with a commitment that is promised later and with lower probability. What our research has proven is that there are structural regularities between the parameters of a probabilistic commitment and the agents' reward functions, such that we can safely prune large portions of the space of possible commitments, to zoom in on a much smaller set of specific time-probability combinations to evaluate.

Furthermore, it will generally be the case that the process for finding the best cooperative agreement will be decentralized: the provider and recipient each knows its own reward function, but not the reward function of the other. To solve this decentralized problem of making an informed decision about their joint commitment, we have integrated into this work results from our thrust on querying. Specifically, the commitment provider reuses the greedy query construction strategy from our earlier work to compose a k-ary commitment query that it sends to the recipient, who uses its local reward and the information contained in the query to identify which of the k commitments proposed is expected to maximize joint reward. Our experiments have shown the efficacy of this integrated approach [5, 11].

Other Accomplishments [1, 13]. Prior to our work just described on using k-ary querying protocols for converging on commitments, we had considered a more general setting of *learning* to communicate between a speaker agent and a listener agent, which we hoped to adapt to the specific problem of learning better communication protocols and representations for commitments. However, while we met with some success, we determined that we were unlikely to succeed in that adaptation within the timeframe of this project. We captured the insights gained into learning to communicate in a paper that appeared at AAAI 2019 [13].

This project also has emphasized how agents should model each other, such as when formulating (and answering) a k-ary commitment query. Issues in mutual modeling, and recursive agent models, was a topic of focus in an article, whose writing was in part supported by this project, that appeared in the journal *Artificial Intelligence* [1].

Project Personnel

The personnel associated with the project over its duration was unusually stable, comprised throughout of the two principal investigators, Professors Ed Durfee and Satinder Singh, and two graduate student research assistants, Qi Zhang and Shun Zhang, who began their doctoral studies at about the same time this project began.

Both of these graduate students successfully defended and submitted their dissertations in the summer of 2020. Qi Zhang joined the University of South Carolina in the fall of 2020 as an assistant professor. Shun Zhang has taken on a post-doctoral position working with IBM Research.

Project Publications

All of the publications in the list of references are based on work from this project since its inception.

References

[1] Prashant Doshi, Piotr J. Gmytrasiewicz, and Edmund H. Durfee. Recursively modeling other agents for decision making: A research perspective. *Artificial Intelligence*, 279:1–44, 2020.

[2] Edmund H. Durfee and Satinder Singh. Commitment semantics for sequential decision making under reward uncertainty. In *Working Notes of the AAAI Fall Symposium on Sequential Decision Making for Intelligent Agents*, pages 13–20, November 2015.

[3] Edmund H. Durfee and Satinder Singh. On the trustworthy fulfillment of commitments. In *Working Notes of the Eigthteenth International Workshop on Trust in Agent Societies*, pages 54–62, May 2016.

[4] Edmund H. Durfee and Satinder Singh. On the trustworthy fulfillment of commitments. In Nadine Osman and Carles Sierra, editors, *Autonomous Agents and Multiagent Systems: AAMAS 2016 Workshops Best Papers*, pages 1–13. Springer Lecture Notes in Artificial Intelligence, 2016.

[5] Qi Zhang. *Making and Keeping Probabilistic Commitments for Trustworthy Multiagent Coordination*. PhD thesis, University of Michigan, 2020.

[6] Qi Zhang, Edmund Durfee, and Satinder Singh. Computational strategies for the trustworthy pursuit and the safe modeling of probabilistic maintenance commitments. In *Working Notes of the IJCAI 2019 Workshop on AI Safety*, August 2019.

[7] Qi Zhang, Edmund H. Durfee, and Satinder Singh. Challenges in the trustworthy pursuit of maintenance commitments under uncertainty. In *Proceedings of the 20th International TRUST Workshop*, pages 75–86, July 2018.

[8] Qi Zhang, Edmund H. Durfee, and Satinder Singh. Efficient querying for cooperative commitments. In *Notes of the AAMAS-2020 Workshop on Optimization and Learning in Multiagent Systems (OptLearn-MAS)*, May 2020.

[9] Qi Zhang, Edmund H. Durfee, and Satinder Singh. Modeling probabilistic commitments for maintenance is inherently harder than for achievement. In *Proceedings of the 34rd Conference of the Association for the Advancement of Artificial Intelligence (AAAI)*, February 2020.

[10] Qi Zhang, Edmund H. Durfee, and Satinder Singh. Semantics and algorithms for trustworthy commitment achievement under model uncertainty. *Autonomous Agents and Multiagent Systems*, 34(19):1–35, 2020.

[11] Qi Zhang, Edmund H. Durfee, and Satinder Singh. Efficient querying for cooperative probabilistic commitments. In *Proceedings of the 35rd Conference of the Association for the Advancement of Artificial Intelligence (AAAI)*, February 2021.

[12] Qi Zhang, Edmund H. Durfee, Satinder Singh, Anna Chen, and Stefan Witwicki. Commitment semantics for sequential decision making under reward uncertainty. In *Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI16)*, pages 3315–3323, July 2016.

[13] Qi Zhang, Richard Lewis, Satinder Singh, and Edmund H. Durfee. Learning to communicate and solve visual blocks-world tasks. In *Proceedings of the 33rd Conference of the Association for the Advancement of Artificial Intelligence (AAAI)*, January 2019.

[14] Qi Zhang, Satinder Singh, and Edmund H. Durfee. Minimizing maximum regret in commitment constrained sequential decision making. In *International Conference on Automated Planning and Scheduling (ICAPS2017)*, pages 348–356, June 2017.

[15] Shun Zhang. *Efficiently Finding Approximately-Optimal Queries for Improving Policies and Guaranteeing Safety*. PhD thesis, University of Michigan, 2020.

[16] Shun Zhang, Edmund H. Durfee, and Satinder Singh. Approximately-optimal queries for planning in reward-uncertain Markov decision processes. In *International Conference on Automated Planning and Scheduling (ICAPS2017)*, pages 339–347, June 2017. Also appeared in Proceedings of the 3rd Multidisciplinary Conference on Reinforcement Learning and Decision Making, June 2017.

[17] Shun Zhang, Edmund H. Durfee, and Satinder Singh. Minimax-regret querying on side effects for safe optimality in factored Markov decision processes. In *Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI18)*, pages 4867–4873, July 2018.

[18] Shun Zhang, Edmund H. Durfee, and Satinder Singh. On querying for safe optimality in factored Markov decision processes (extended abstract). In *International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS18)*, pages 2168–2170, July 2018.

[19] Shun Zhang, Edmund H. Durfee, and Satinder Singh. Querying to find a safe policy under uncertain safety constraints in Markov decision processes. In *Notes of the NeurIPS-2019 Workshop on Safety and Robustness in Decision Making*, December 2019.

[20] Shun Zhang, Edmund H. Durfee, and Satinder Singh. Querying to find a safe policy under uncertain safety constraints in Markov decision processes. In *Proceedings of the 34rd Conference of the Association for the Advancement of Artificial Intelligence (AAAI)*, February 2020.