# SELECTIVE ACQUISITION OF OPERATOR KNOWLEDGE FOR SAFER SEMI-AUTONOMOUS ROBOT OPERATIONS

**Edmund Durfee, David Karmol, Michael Maxim, and Satinder Singh**
Computer Science and Engineering
University of Michigan
Ann Arbor, MI 48109

**ABSTRACT**

*We have developed techniques for a robot to compute its expected myopic gain in performance from asking its operator specific questions, such as questions about how risky a particular movement action is around pedestrians. Coupled with a model of the operator's costs for responding to inquiries, these techniques form the core of a new algorithm that iteratively allows the robot to decide what questions are in expectation most valuable to ask the operator and whether their value justifies potentially interrupting the operator. We have performed experiments in simple simulated robotic domains that illustrate the effectiveness of our approach.*

## INTRODUCTION

Safe operations in environments that include potentially dangerous artifacts as well as humans whose intent and capacities are unknown pose challenges both to robots as well as to their human operators. In such environments, the multitude of demands on an operator's attention necessitate periods of autonomous behavior by the robot. To operate safely, such a robot must be able to reason about its knowledge so that it can decide when it has confidence in its ability to act autonomously, and when it should seek help from the operator.

The work we briefly describe in this paper concentrates on how an autonomous robotic vehicle can make well-founded decisions about when to seek operator input and what input to ask for, given what it knows about its environment and the value of the operator's attention. In the following sections, we first summarize how the robot models its environment and its uncertainty. We then describe a general technique for computing expected myopic gain (EMG) in reducing particular points of uncertainty, and how EMG allows decisions about the desirability of seeking information from the operator. Afterward, we show how EMG selectively acquires operator knowledge for several different types of uncertainty, and summarize its empirical performance. We conclude by describing our ongoing work.

## ENVIRONMENT MODELS AND UNCERTAINTY

A robot models its interaction with its environment as a Markov decision process (MDP), which defines the possible states the robot can be in, actions it can take, transition probabilities (likelihood of being in state $s'$ if action $a$ is taken in state $s$, for all combinations of states and actions), and rewards for each combination of state and action taken. A very simple example is shown in Figure 1, where the states are numbered 0 through 6 (the robot starts in 0), and the actions are either "solid" or "dashed." Simple example transition probabilities are that the solid action moves to the rightside next state with probability .9 (.1 to the left), while the dashed moves to the leftside with probability .9 (.1 to the right). More generally, the actions could have different transition probabilities at each of the states. The rewards of states 3-6 are shown, and are zero for states 0-2.

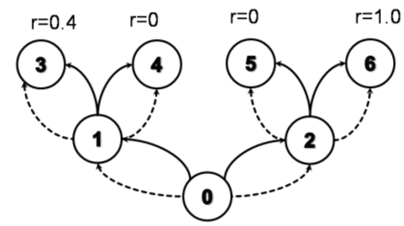


**Figure 1:** Simple Example

With all of this information, the robot could compute the optimal policy using standard MDP techniques (e.g., Bellman backup), to find the optimal policy of choosing the solid action in state 0, solid in state 2, and dashed in state 1 (which it has a probability of .1 of reaching). The expected reward of following this policy beginning in state 0 is 0.846 (reaching state 6 with probability .81 and state 3 with probability .09).

While this example is very small, most interesting environments require large models, such that asking an operator to articulate the full model to the robot is impractical. Instead, the robot will have only (what the operator believes is the most useful) partial model information. The robot thus captures its uncertainty about the complete environment as a probability distribution over full MDPs.

## EXPECTED MYOPIC GAIN

Given this formulation, the robot can infer the expected gain of getting more information to improve its model. We define the expected myopic gain (EMG) as follows:

$$E\left[Gain_{[\theta]}(s',Q)\right] = \sum_{Q=x} Gain_{[\theta]}(s',Q=x)Pr_{[\theta]}(Q=x) \quad (1)$$

Here, $\theta$ represents the state of partial knowledge, so the expected gain in reward if the agent asks question $Q$ about state $s'$ is a sum of the gain for each of the possible answers it might get back times the probability of getting that answer. The (expected) gain if the agent is currently in state $s$ for an answer $x$ is:

$$Gain_{[\theta]}(s',Q=x) = E_{[\theta,x]}\left[V^{\pi^*_{[\theta,x]}}(s)\right] - E_{[\theta,x]}\left[V^{\pi^*_{[\theta]}}(s)\right] \quad (2)$$

That is, the gain of learning answer $x$ is the difference between the expected value of following the optimal policy with that additional knowledge given the more complete model, and the expected value of following the previous (less informed) optimal policy if the world is as the more complete model specifies.

Applying equations (1) and (2) requires the development of the underlying machinery to tractably compute the equations' components. For example, efficiently computing the posterior probability distribution over possible MDP models given an answer to a question in general requires using compact representations [1] (e.g., a factored Dirichlet parameterization). Similarly, computing the optimal policy for a distribution of MDPs is computationally challenging, and we have experimented with several approaches.

The EMG strategy is to ask the question for the state that, in expectation, has the largest gain. The approach is myopic because it does not reason about how a follow-up question could affect the expected gain of a first question [2]. Nonetheless, we can approximate a good sequence of questions by repeatedly applying EMG.

Figure 2 shows example results of applying EMG to the problem in Figure 1, where the robot initially does not know the transition probabilities of any of the actions, meaning that it has 6 possible questions (about "solid" or "dashed" in states 0-2). The graph shows how the robot's expected utility grows as it asks more of these questions. The lower curve corresponds to picking which question to ask next randomly, while the upper curve is where the robot uses EMG. In this case, notice how EMG picks better questions first, achieving near maximum utility after just 2 questions.

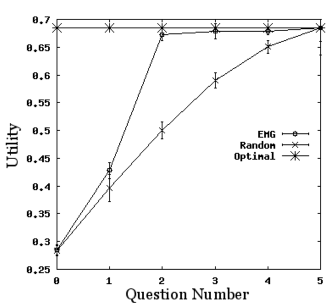


**Figure 2:** EMG for Simple Example

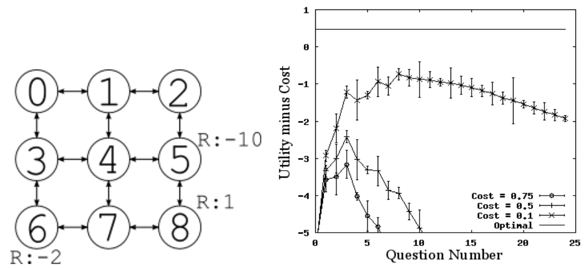


**Figure 3:** Grid-World Performance with Question Costs

## OPERATOR COST

Figure 2 illustrates that, if selected well, asking successive questions might have diminishing returns, suggesting that if asking a question incurs cost (distracting the operator from other duties), then the robot can take this into account and decide to stop asking questions when cost exceeds benefit.

Figure 3, left, shows another simple domain, where the robot can move around a grid-like environment. Some locations are rewarding, while others have severe penalties (such as blundering into a crowd of people). Under the same assumption that the robot lacks knowledge of transition probabilities, EMG for this problem focuses its first questions on states next to the penalizing state, in essence selectively acquiring information most critical to safe operations. As shown in Figure 3, right, different costs associated with asking questions would lead the robot to limit its questions (stopping as soon as a curve's expected utility stops climbing). That is, the robot will always find it useful to ask about risky locations, but as the costs of asking the operator rise it becomes more willing to wander more ignorantly in benign areas of the world.

## CONCLUSIONS

Our past and ongoing research has been extending the approaches in this paper to wider varieties of problems, including more complex environments and worlds where operator responses might be more ambiguous. Our goal is to introduce these techniques into real robots that can autonomously determine when to seek operator assistance.

## ACKNOWLEDGEMENTS

This work is supported in part by the Ground Robotics Reliability Center of the University of Michigan, a Center of Excellence sponsored by the U.S. Army TARDEC.

## REFERENCES

[1] Richard Dearden, Nir Friedman, and David Andre. Model based Bayesian exploration. In *Proc. Fifteenth Conf. on Uncertainty in AI*, pages 150–159, 1999.

[2] Michael O. Duff. Design for an optimal probe. In Tom Fawcett and Nina Mishra, editors, *ICML*, pages 131–138. AAAI Press, 2003.