Ankit Shah

Training Autonomous Agents Like Apprentices

My research vision involves developing computational models and algorithms that allow domain-experts to directly train autonomous agents just as they would train a human apprentice. The ideal autonomous apprentice will learn complex tasks through intuitive modalities like demonstrations, natural language instructions, or remote operations; it will reason about its own uncertainty of the expert's belief while performing the task; finally, when the expert's input is available, it will intelligently elicit expert feedback to reduce this uncertainty. The flexibility afforded by these quick-learning apprentices will be invaluable when deployed in domains where rapid adaptation to new tasks is desirable. While I primarily focused on robotics-oriented applications during my Ph.D. research, I am equally motivated towards applications rooted in virtual agents such as digital assistants and decision support systems.

In my Ph.D. research, I developed a Bayesian framework for robot training centered on expressing the expert's intent as a belief over formal specifications. This allows modeling the ambiguities in modalities like demonstrations as a superposition of multiple unambiguous specifications expressed in a formal language. I partnered with Lockheed-Martin to develop an explainable mission analysis and review system for multi-aircraft flight missions [6, 2, 3] by applying my framework towards identifying mission objectives from the mission commander's evaluations. The Bayesian framework also allowed a robot to learn specifications for multi-step manipulation tasks from demonstrations [7], generate execution policies for these tasks that maximally satisfy a belief over logical formula [8], interactively learn a complex task from both an expert's demonstrations and assessments of the robot's own task executions within an active learning framework [9]. We also adopted the Bayesian framework to explain differences between groups of plans within a wide range of symbolic planning domains [4].

Empowering everyday users who may lack programming expertise to teach virtual and embodied autonomous agents that they interact with, will be the defining theme of my future research. I believe that research advances at the intersection of reinforcement learning, abstractions, and cognitive models for users will be crucial towards achieving my research vision.

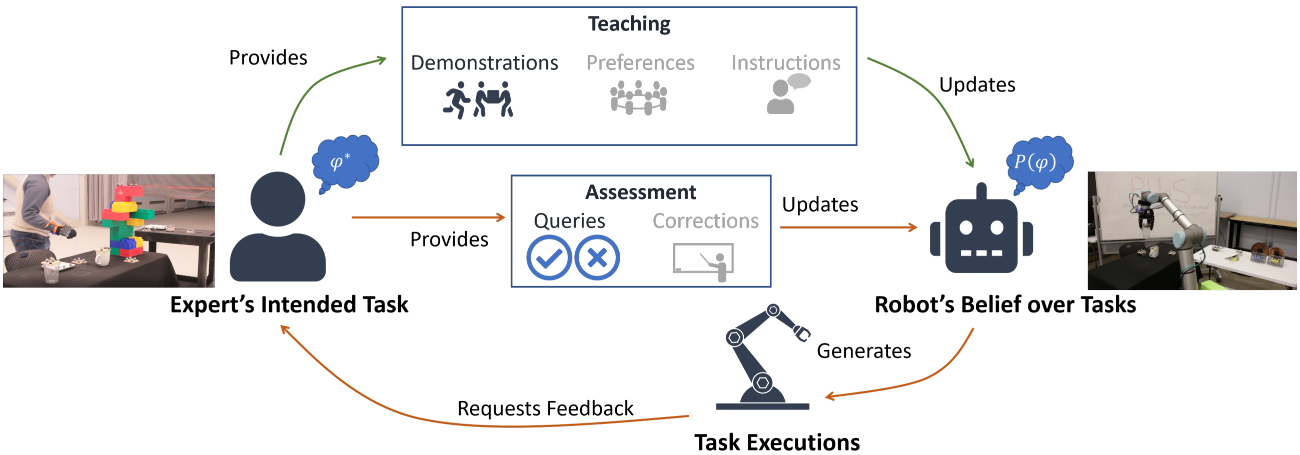

Figure 1: A Bayesian framework for training a robot through multiple intuitive modalities. The expert intends to teach a specific task φ^* to the robot. Meanwhile, the robot maintains a belief $P(\varphi)$ over candidate tasks that it must perform. 'Teaching' is the type of input that is initiated by the user (green arrows), while 'Assessments' are user requested by the robot (orange arrows).

Current Research

During my Ph.D., I developed computational models to interactively learn the user's intended task by leading research in the following directions:

Encoding ambiguities in user provided demonstrations:
Formal languages like linear temporal logic (LTL) [5] are ideal for unambiguously defining non-Markov task specifications. In particular, LTL is expressive enough to capture long-horizon temporal constraints commonly seen in real-world scenarios, such as manufacturing processes and behavior specifications for cars or aircraft. However, formal languages can be unwieldy for a domain-expert compared to providing task demonstrations. Our key contribution was modeling

the epistemic uncertainty caused by demonstrations' ambiguity as a belief over a set of unambiguous logical formulas within a Bayesian formulation.

We used this model to learn task specifications for applications ranging from multi-aircraft flight missions [6] to multi-step manipulation tasks [7]. The work on identifying mission objectives for multi-aircraft flight missions is at the core of the automated mission analysis and review systems developed in collaboration with Lockheed-Martin [2, 3]. Automated evaluations of flight missions generated by our system had over 95% agreement with the evaluation generated by an expert mission commander while learning from only 25 labeled examples. In comparison, widely used sequence classification models struggled to perform better than random predictions. Furthermore, the logical structure of the formulas enabled the mission commander to interpret the model's reasoning in case of disagreements.

Formalizing decision-making with uncertain specifications:
Prior research into planning with formal specifications has considered the task specifications to be known unambiguously. However, the inherent epistemic uncertainty over the true task specifications necessitated a novel approach. I developed planning with uncertain specifications (PUnS) [8], a novel formalism that allows the robot to reason about its own uncertainty over task specifications. Further, I demonstrated that every instance of a PUnS problem is equivalent to a provably minimal Markov decision process making it compatible with planning algorithms for MDPs.

Admitting non-Markov specifications makes PUnS suitable to multi-step tasks with flexible execution policies that are commonly found in manufacturing (a set of parts is assigned to a station, but the worker decides the actual order of installation), household assistance (setting a dinner table, clearing clutter on a table, stocking the pantry) or unstructured domains for example disaster response or space exploration. We chose the task of setting the dinner table, a task that retains the temporal complexities of non-Markov specifications while being amenable to a laboratory setting. The task policies computed using the PUnS formulation based on the specifications inferred through just 30 demonstrations were estimated to have a low error rate of $\approx 0.01\%$ while demonstrating multiple unique sequences for setting the table[1]. We have demonstrated the efficacy of PUnS in generating policies that satisfy beliefs over task specifications for a wide range of specifications [9].

Refining robot's belief through interaction:
During the final phase of my Ph.D., I developed an active learning framework for non-Markov tasks [9]. This framework leverages the original PUnS problem's MDP equivalent to identify and execute queries whose acceptability is most uncertain as per the robot's belief. The user's assessment of this query generated through uncertainty sampling is used to refine the belief optimally. Simulation experiments indicated that our active learning approach leads to better alignment of the posterior belief with the ground-truth specifications compared to learning from an equal number of expert-generated demonstrations. In simulation, the active learning approach inferred the ground-truth specification with near certainty with only eight task executions, while a model that relied purely on demonstrations struggled to achieve equivalent performance with as many as 50 demonstrations.

In a user-study based on teaching a robot to set the dinner table, all our participants were able to teach the robot to correctly set the table without any errors with just five task executions[2]. The use of belief over logical formulas as task specification allowed the participants to train the robot through multiple teaching modalities like physical demonstrations, tele-command of the robot, and binary acceptability assessments of the robot's task execution. Our ability to switch rapidly from physical demonstrations to tele-command – in view of the lab access restrictions resulting from the COVID-19 pandemic – shows the flexibility of our framework can achieve. This study provides evidence that active learning for complex non-Markov tasks in a real-world setting like manufacturing is viable in the near future.

Future Research Agenda

I envision developing autonomous agents capable of learning as flexibly and rapidly as humans. This involves continual improvement after every interaction with the human expert, actively recognizing avenues for improvement, and succinctly explaining its understanding to the human expert. Towards that goal, I am particularly interested in tackling the following research areas adjacent to autonomous systems and reinforcement learning in my future research.

Algorithmic teaching grounded in cognitive models for pedagogy:
Our recent experiments on interactive robot training [10] revealed that the relative performance of various interactive robot training protocols depends on the ground truth specification. Further, we also observed that human teachers are better at training a robot purely through demonstrations than a simulated teacher. This indicates the existence of cognitive biases that promote different behavior while *performing* a task compared to *teaching* a task. Modeling these pedagogical

1 -The video can be viewed at `http://puns.ajshah.info`
2 -The video of example teaching sessions can be viewed at `http://interactive.ajshah.info`

biases remains an open problem that will be a key theme of my future research. Accurately modeling pedagogical biases can not only lead development of autonomous agents that can learn faster by leveraging the pedagogical intent of the teacher, but also result in models for explaining an agent's behavior by generating a representative sample of its policies.

Aligning abstractions between humans and RL agents

Autonomous agents that develop hierarchical abstract task representations can later compose them through simple rules to perform increasingly complex tasks. The ability to develop useful abstractions is key to deploying these agents in new partially defined domains. However, to ensure that the agents can learn from human teachers in these domains, it is key that the robot's abstractions are aligned with those used by human domain experts.

Cognitive science research has a vast body of work characterizing human cognitive biases. Developing learning models for robots that leverage human cognitive biases to achieve data-efficiency will be an active focus area for me. My work on learning temporal specifications from demonstrations [7] was an example of successfully leveraging temporal abstractions along with cognitively inspired Bayesian learning. Closely related to this is the challenge of allowing an agent to accurately and compactly summarize its own representation for the user. More concretely, the agent should be capable of answering questions like: '*when do you consider the bowl to be placed correctly?*' in the table-setting context. Our initial approach to this – Bayes-TrEx [1] – explored generating examples of input data that resulted in requisite outputs as a way of explanation. I believe that advances in active learning; and approaches inspired by theory-of-mind and pragmatics will play a key role in addressing this challenge.

Dynamic environments with multiple decision-makers:

The ability to learn novel tasks through intuitive modalities, jointly construct useful abstractions along with a human teammate, and then convey novel information back to the human teammate are critical enabling technologies for agile robots of the future. However, many real-world environments are dynamic, including multiple decision-making actors and have pre-defined regulations for operations, such as public roads or the airspace systems around the world. There are two key challenges in dynamic environments: first, the expert demonstrator cannot always guarantee the correct execution of tasks due to the non-deterministic nature of the environment and the presence of other decision-making agents; second, the changes in the task state may not be causally linked to the actions of the robot.

The multi-aircraft flight mission domain [3], integration of autonomous aircraft in the national airspace system, and multi-step manipulation problems, and multi-agent task coordination problems – frequently seen in manufacturing and disaster response scenarios – represent instances of these same challenges.

Engendering calibrated trust in automation

Beyond technological hurdles, a key hurdle to widespread deployment of automation is lack of trust in robots leading to misuse or disuse. The technological challenges of characterizing pedagogical priors and aligning a robot's abstractions with that of the human domain-experts are motivated by engendering calibrated trust in the robotic system through data-efficient learning and transparency and explainability of the robot's learned policy. However, trust in a system is not a static phenomenon. Modeling the evolution of an expert's trust in the robot's capabilities from initial deployment to continued improvement post deployment will be a crucial area of my future research.

Conclusion

My vision is to empower domain-experts who may not have programming skills to participate in the development of these technologies. I believe that with my experience in identifying novel research challenges [7, 8] and working with industry partners towards translating my research into practice [2, 3] will enable me to build an impactful research program. I intend to actively seek out applications of my work towards robotics, aerospace, healthcare, and decision support systems for deploying flexible, adaptable and taskable autonomous agents in the real world.

[1] - Serena Booth, Yilun Zhou, Ankit Shah, and Julie Shah. Bayes-trex: a bayesian sampling approach to model transparency by example. In *Proceedings of the AAAI Conference on Artificial Intelligence (Accepted for publication)*, 2021

[2] - Patrick L. Craven, Kevin B. Oden, Kevin J. Landers, Ankit J. Shah, and Julie A. Shah. Man-machine interoperation in training for offensive counter air missions. In *Interservice/Industry Training, Simulation and Education Conference*, 2018

[3] - Patrick L. Craven, Kevin B. Oden, Kevin J. Landers, Ankit J. Shah, and Julie A. Shah. Man-machine interoperation in training for large force exercise air missions. In *Interservice/Industry Training, Simulation and Education Conference*, 2019

[4] - Joseph Kim, Christian Muise, Ankit Shah, Shubham Agarwal, and Julie Shah. Bayesian inference of linear temporal

logic specifications for contrastive explanations. In *International Joint Conference on Artificial Intelligence*, 2019
[5] - Amir Pnueli. The temporal logic of programs. In *Foundations of Computer Science, 1977., 18th Annual Symposium on*, pages 46–57. IEEE, 1977
[6] - Ankit Shah, Pritish Kamath, Shen Li, Patrick Craven, Kevin Landers, Kevin Oden, and Julie Shah. Supervised bayesian specification inference from demonstrations. In *(under review)*, 2019
[7] - Ankit Shah, Pritish Kamath, Julie A Shah, and Shen Li. Bayesian inference of temporal task specifications from demonstrations. In *Advances in Neural Information Processing Systems*, pages 3804–3813, 2018
[8] - Ankit Shah, Shen Li, and Julie Shah. Planning with uncertain specifications (PUnS). *IEEE Robotics and Automation Letters*, 5(2):3414–3421, 2020
[9] - Ankit Shah, Samir Wadhwania, and Julie Shah. Interactive robot training for non-Markov tasks. *arXiv preprint arXiv:2003.02232 (under review)*, 2020
[10] - Ankit Shah, Samir Wadhwania, and Julie Shah. Robot pedagogy for non-Markov tasks. *(in preparation)*, 2020