Helping People Predict Agent Behaviors by Operationalizing the Variation Theory of Learning

Tiffany Horter
Wellesley College
horter@csail.mit.edu

Serena Booth
Brown University

Vinitha Gadiraju
Wellesley College

Abstract

To stay safe and effective when collaborating with a cobot or an AI agent, people must be able to predict the future behaviors of their automated partners. We propose using the Variation Theory of Learning, a theory of how humans learn new concepts, to allow people to predict agent behaviors by building conceptual models of agent policies. In this work, we explore the space of design decisions needed to operationalize Variation Theory and how to best to scaffold peoples' experiences of interacting with agents to inform their conceptual model development. We study this operationalization by analyzing two domains: a pick-and-place robot arm task and a simulated highway driving environment. We find evidence that operationalizing Variation Theory can assist people in identifying a given agent's behavior in novel settings, an intermediary task en route to measure the promise of applying Variation Theory to people predict new agent behaviors.

1 Introduction and Background

Imagine working in close physical proximity to a robot. If you cannot predict how the robot will pick something up, there is a risk that you will collide with the robot [6]. Alternatively, imagine a radiologist examining scans with the help of a faulty prediction model. The consequences of misunderstanding where likely failures exist could harm patients and lead to unequal health outcomes. When we interact with an agent, we face the problem described in philosophy as the "problem of other minds" [5]. We cannot analyze the other mind's internal decision-making policy and instead determine agents' internal reasoning based on observed behaviors [5]. In its most general form, being able to predict agent behaviors implies the person maintains an accurate belief distribution over the agent's policy that determines which actions the agent might take.

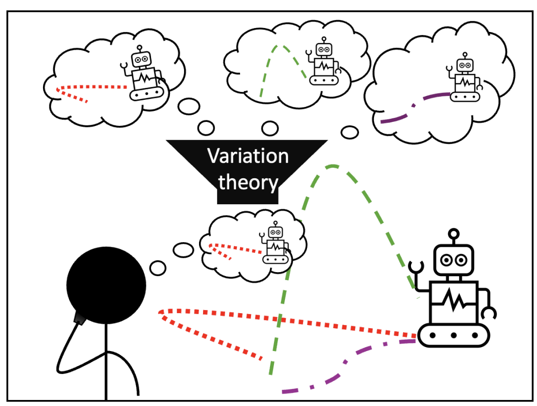

Figure 1: People start with many hypotheses about how an agent might behave. For example, a robot might move in many different ways, here abstractly represented by three trajectories. Variation Theory can be applied to help people *narrow* the space of possible policies towards the true behavior.

We present an approach to addressing the problem of how to learn about the behaviors of a particular agent based on a theory of human learning—the Variation Theory of Learning [9]. Variation Theory builds up people's concept models through a sequence of steps that exposes "patterns of variance and invariance" to help the person isolate the core behaviors from superficial details [9]. Variation Theory can help the person predict future agent behaviors by sequentially reducing the

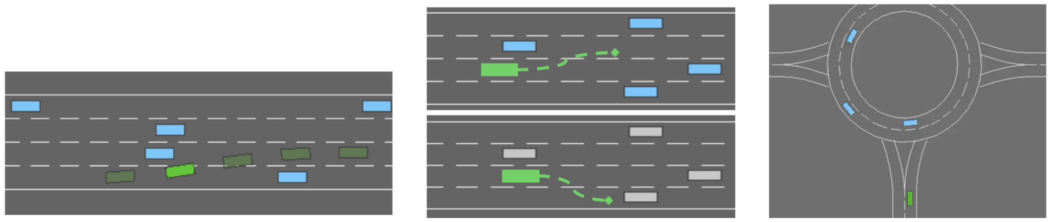

Figure 2: The steps of variation theory applied to the HighwayEnv environment.

hypothesis space of possible agent policies. Variation Theory helps people develop a conceptual model that matches the outputs of that correct policy. This technique scaffolds a person's model of the agent by revealing information-rich examples of agent behaviors. While our larger goal is to help people—whether expert model-designers or end users—predict agents' behaviors, we study the intermediary task of identifying a given agent's behavior from a set of alternative behaviors. This tests whether Variation Theory can help people narrow down the space of possible agent policies to ultimately help them predict the agent's future behaviors. Variation theory is a theory of how humans learn concepts, where each concept is made of "aspects" that distinguish one version of the concept from another. We apply these broad learning theories to assist people in learning what to expect from a machine's behavior; this space of expected behaviors constitutes, in our setting, the concept. Variation theory, as described by Marton [9], is a sequential procedure that helps people learn each aspect by isolating its effect in the larger concept. The sequential steps of Variation Theory are:

1. *Familiarization:* Familiarization shows examples with all aspects maintained the same, which helps people find the essential aspects of the policy. When people see an example of the behavior, this causes them to generate a set of possible explanations for what they have just seen: their hypothesis space of policies.
2. *Contrast:* Contrast varies the aspect being learned but keeps all other environmental factors the same. This helps create the policy's concept boundaries by showing what is not a part of the concept, eliminating it from people's space of possible policies.
3. *Generalization:* Here, the aspect being learned remains the same while external aspects are changed. Generalization helps define the policy concept by finding the unchanging aspects that are key to the policy regardless of the experienced environment, and combats overfitting in the contrast step if irrelevant aspects were coincidentally held constant and interpreted as essential.

Booth et al. first proposed applying Variation Theory to the human-robot interactions setting and identified its use for guiding people's learning of robot models [3]. However, this work did not implement Variation Theory or any human subject studies. Though they did not use the language of Variation Theory, Dragan and Srinivasa's work showed that using familiarization helps people better predict the motion of a robot [4]. Their study showed users a robot reach its arm to an object placed on the table, then tested whether the users could correctly identify the motion from several possible motion options. Familiarization alone improved users' ability to predict future robot motions, but users still found this task challenging, especially when predicting unnatural robot motions. Hence, adding in the other steps of Variation Theory has promise for improved behavior prediction, an essential component of improved human-robot or human-AI collaboration.

2 Domains

We consider two motivating example domains: a robot arm in a pick and place task, and a reinforcement learning agent for a highway environment. The pick and place domain was modeled after Dragan and Srinivasa's setup [4], though with a different robot (a Franka Panda). This robot reaches for different objects before it, but there is ambiguity in how the robot performs that motion—for example, the robot might use a policy based on optimizing the distances from obstacles like in CHOMP [10] or it might follow a policy that rotates each joint in a specific order, like sequentially from wrist-to-shoulder. The human observing this behavior has many possible hypotheses of policies that could exhibit this behavior. Some people might imagine the robot always taking direct paths; some people might imagine the robot being cautious and slow. This work attempts to scaffold people's learning to narrow down that space of possible policies to be the one moving wrist to shoulder joint-wise. To learn to be able to predict motions for this particular policy, there are two distinct aspects that must be understood: (a) it moves joint-wise and (b) it moves sequentially from the smallest joint to the largest.

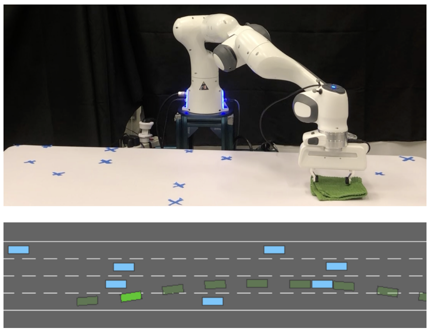

Figure 3: Pick and place domain above highway driving domain.

To test the resilience of the operationalization in multiple settings, we also study a reinforcement learning agent in the HighwayEnv [8] environment, a simple simulator approximating a self-driving car. There were several possible environments ranging from a simple highway with several lanes of traffic moving in one direction to a round-about and a highway entrance. We test multiple policies, based loosely on the reward functions described by Amitai et al [1]. All the policies have the same task (avoid other cars while moving forward), but each has a distinguishing aspect: moving in the right lane quickly, avoiding other cars, or driving parallel with the car in front. Here, correctly being able to predict future car motions is important for giving targeted feedback and, for other drivers in its proximity, avoiding the vehicle and maintaining safe driving procedures.

Both these domains were challenging for people to predict or identify the correct robot behavior without the guidance of Variation Theory. For the robot arm, when people were simply exposed to examples of robot behavior, they had 52.4% accuracy. For the cars, those trained by just showing the car moving in the environment had 48% accuracy. Without improving training, people cannot reliably identify the policies.

3 Design Decisions for Operationalizing Variation Theory

To design our scaffolding in line with Variation Theory, we must find aspects of the policy that would change the behavior of the agent and group each aspect into either "environmental" or "agent-related". Is this aspect changing the "environment" (making the agent's behavior change in response to changes in the world around it such as the type of road it is on) or is it changing the agent itself (the policy)? Showing only a small sample of the behaviors is desirable as Dragan and Srinivasa showed thatpredictability does not increase with the number of examples [4], and too many examples risks overwhelming the user.

3.1 Familiarization

The first step of Variation Theory, familiarization, is relatively straightforward. This step repeatedly shows the minimal typical behavior of the agent. The objective is to communicate the agent's task to the user by providing a basic example of it achieving the task.

- Pick-and-place: Show an example of robot motion that shows the arm moving from point A to B, moving without obstacles. This step establishes the concept space but extrapolation of the behavior remains challenging. The current concept space might consist, for example, of the set: {up then down, joints move one at a time, gripper followed by body, gets best grip}
- Highway Driving: Show a video of the car slowing and moving to the left. This sets up the concept space as being all possible policies that don't hit others and move forward. The current concept space might consist, for example, of the set: {moving to the left-most lane, slowing, avoiding other cars, passing other cars, fastest in lane}

Key Choices The key choices in this step are using the variables identified earlier to find what should be held constant throughout the examples shown. Another important decision is to show an example of successful behavior at achieving the task—showing what it cannot do is not yet important if the people watching don't know what it should be doing.

3.2 Contrast

The contrast step contrasts different possible policies' reaction to the same environment. Even if everything else were the same, how could this agent plausibly behave differently? This step now tries to reduce the space by eliminating valid but incorrect possible policy generalizations from the first step. By showing alternatives to each of the distinguishing aspects, the concept space is corralled.

- Pick-and-place: Show examples of an alternate policy (shoulder-to-wrist) next to correct policy. This highlights order of the joints' movement as an essential consideration in the policy. The concept space is now limited to concepts that produce the right behavior in this scenario - but this still leaves alternative policies which exhibit identical behavior. A possible current concept space could be: {joints move one at a time, moves for best grip, rotates instead of straight line}
- Highway Driving: Compare with an agent moving aggressively near other vehicles to pass, therefore eliminating policy options that involve getting close to other vehicles like being the fastest. Someone's concept space might update to: {moving to the left-most lane, slowing, avoiding others}

Key Choices To eliminate possible but incorrect options, these alternatives should be:
Plausible Consider alternate possible policies that could, on first glance, be viable - this disrupts people's malformed priors. Show the user that these are not actually how YOUR agent will behave. Pick-and-place: Show the users policies that are still optimal in terms of robot movement - gibberish trajectories that are not optimal in any way would usually violate people's optimality assumption [4], so are likely not in the space of concepts they are thinking about. Highway Driving: Do not show alternate trajectories that weren't initially in the concept space, like hitting cars.

Maximally visually distinct In selecting these, pick moments of greatest difference in behavior between policies to maximize the effect. All demonstrations must be visibly different in behavior for humans to learn from it. Showing the direct opposite of the distinguishing aspects, not changing anything else about the scenario, ensures that the difference is highlighted. For example, the smooth motion contrasts with joint-wise motion, eliminating any smooth motion from the possibilities.

Aligned When demonstrating these, show the two policies in an aligned representation to make the differences clearer, as suggested by Analogical Transfer Theory [2]. For example, robot arms were displayed side by side simultaneously for the differences to be immediately comparable.

3.3 Generalization

Demonstrations for this step should help users become able to predict the possible changes from the "environment" variables, like obstacles or a different object to pick up. Essentially, now that the user has a rough idea of the policy, we show examples that break any incorrect models they built. Ideas from the concept space that could be disguised in the previous steps as causing the same behavior are elicited by showing different environments where those policies would behave differently.

- Pick-and-place: By showing an environment where the item has a different grip but the robot's policy remains the same would eliminate a grip-based policy from the considered space.
- Highway Driving: Display scenarios where the car moves in a new environment. The limited number of exposures mean that there are other possible policies which could cause those behaviors. Here, show alternate environments where the car maintains avoidance of other vehicles but moves into different lanes, or is forced to speed up to avoid others approaching from behind.

Key Choices The environmental changes should be prioritized by what would cause the largest deviation from the prior behaviors seen. To do this, consider the simplest model that could have been generated from what has been seen so far, examine where that model might fail, then generate examples of that kind of "out-of-distribution" behavior. Generalization can not only narrow the concept space, it can also widen it if the contrast step only showed policies with a red herring shared aspect. For example, if locations during the contrast step only showed the robot going left, it is possible someone might interpret that as a key aspect - generalization could then show an example of it going right to counteract that false aspect. By the end of the training process, the concept space has been scaffolded in such a way that the policy to expect is clear.

4 Evidence of Conceptual Model Formation

Pick and Place For the robot arm, we ran a first user study to [7] investigate the impact of solely using the contrast step on scaffolding people's mental model formation of how a robot would move. Using a similar setup to Dragan and Srinivasa [4], we showed participants training videos beforehand followed by six test scenes, four of which they had not been exposed to in training in two groups - familiarization and contrast. Familiarization participants saw videos of a robot arm moving to a place on the table, while contrast saw the same video side-by-side to a contrasting video of the same goal but with a different policy of motion.

Throughout this study and the next, prediction accuracy was defined as the intermediate measure of correctly selecting the example policy that matched the policy they were trained on out of 3 shown options with visibly different policies.

Though the mean accuracy results of this preliminary study were not statistically significant, the accuracy showed an absolute increase in prediction ability moving from 52.4% with familiarization to 70.2% under contrast. Unseen settings showed contrast had a significantly higher accuracy at 72.4% compared to familiarization's 50.0% with p=0.01. Users exposed to the first step, familiarization, seemed to not have coalesced to a single explanation - one person noting their assessment of the robot's policy being: "it moves chaotically". Participant's descriptions of the policy varied widely but included "Each arm segment adjusts its angulation to arrive at the final result." Here we see that the actual policy is within the concept space that people come up with, but that there are many alternate policies people believe could cause the behavior seen so far. On the other hand, most contrast users had discerned the important aspects of the policy - rotating joint-wise and moving wrist to shoulder.

Highway Driving For the Highway Driving domain, we ran a second user study to test the impacts of our chosen decisions for each step of variation theory and the combination of all the steps. There were four branches - familiarization, contrast, generalization, and those combined. The study featured a training phase where human participants were shown examples of car behaviors as described above and a testing phase which asked them to predict which of the three options matched the behavior of the car they had been trained with. The tests were drawn from both seen and unseen environments. For all scenarios, the policies of moving fast in the right lane and maintaining distance from others had a baseline mean accuracy for familiarization of 48%, contrast with 52%, generalization at 63%, and the combined condition increasing to 58% accuracy, though no statistically significant effect was observed. These absolute increases in accuracy support further work to test the intervention.

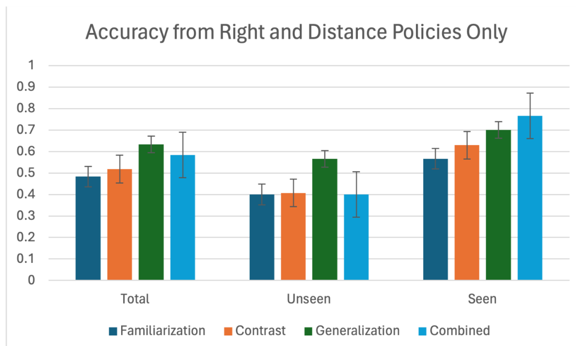

Figure 4: Accuracies for each step of variation theory in the highway domain: this demonstrates the overall help of using variation theory steps in the right and distance policies.

Highway Driving revealed a possible limitation when using Variation Theory - we theorize that it narrows down the concept space users have, but if a user does not start with a certain concept as a possibility (here likely due to pre-exposure to how cars usually drive) that concept needs to be added as a possibility in their mental model before Variation Theory can make an impact. Showing Variation Theory steps with a car whose policy was to move parallel to its nearest neighbor resulted in not statistically significant decreases in prediction ability. The qualitative responses reveal the critical aspect of the right or distance policies (which lane or how much distance) were immediately discerned by the users (appearing in 87% of responses) while the critical aspect of the parallel policy was only discerned by 1 respondent out of 20, most likely because this is not a typical goal people have while driving. This implies that if we have a robot that can go against our innate biases of what we think it can do, it will take more scaffolding to get there.

5 Conclusion

In this paper, we have explored operationalizing Variation Theory in two domains and proposed design guidelines for how to do so in others. We have found support for Variation Theory helping people to predict machine behavior, though there are limitations: policies that are excluded from people's original ideas of how the machine should behave seem to be not be supported by this operationalization. This is a first attempt to make these theories of learning concrete - more work should be attempted to do the challenging work of operationalizing the generality of the theory in different domains. Further work would help explore how to expand, rather than narrow, people's concept space to include these unexpected policies.

References

[1] Yotam Amitai and Ofra Amir. "I Don't Think So": Summarizing Policy Disagreements for Agent Comparison, 2021.

[2] Miriam Bassok. "Analogical Transfer in Problem Solving.". In Janet E Davidson and Robert J. Sternberg, editors, *The Psychology of Problem Solving*, pages 343–370. Cambridge University Press, Cambridge, 2003.

[3] Serena Booth, Sanjana Sharma, Sarah Chung, Julie Shah, and Elena L. Glassman. Revisiting Human-Robot Teaching and Learning Through the Lens of Human Concept Learning. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '22, pages 147–156, Sapporo, Hokkaido, Japan, 2022. IEEE Press.

[4] Anca Dragan and Siddhartha Srinivasa. Familiarization to robot motion. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '14, page 366–373, New York, NY, USA, 2014. Association for Computing Machinery.

[5] David J. Gunkel and Jordan Joseph Wales. Debate: What is personhood in the age of AI? *AI & SOCIETY*, 36(2):473–486, June 2021.

[6] Bradley Hayes and Julie A. Shah. Improving robot controller transparency through autonomous policy explanation. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, pages 303–312, 2017.

[7] Tiffany Horter, Elena L. Glassman, Julie Shah, and Serena Booth. Varying How We Teach: Adding Contrast Helps Humans Learn about Robot Motions. In *2023 HRI Workshop on Human-Interactive Robot Learning*.

[8] Edouard Leurent. An Environment for Autonomous Driving Decision-Making, 2018.

[9] Ference Marton. *Necessary Conditions of Learning.* Routledge., 2014.

[10] Matt Zucker. CHOMP: Covariant Hamiltonian Optimization for Motion Planning.