Varying How We Teach: Adding Contrast Helps Humans Learn about Robot Motions

Tiffany Horter
MIT CSAIL, Wellesley
Cambridge, MA, USA
th105@wellesley.edu

Elena L. Glassman
Harvard University
Cambridge, MA, USA
eglassman@g.harvard.edu

Julie Shah
MIT CSAIL
Cambridge, MA, USA
julie_a_shah@csail.mit.edu

Serena Booth
MIT CSAIL
Cambridge, MA, USA
sbooth@mit.edu

ABSTRACT

Learning how robots move is difficult, but theories of human concept learning can be applied to support humans in this task. We draw insights from the Variation Theory of Learning, a theory that has been validated in the learning sciences through decades of classroom-based studies. Variation Theory prescribes experiencing patterns of structured variation, where some aspects of concepts are held constant while other aspects vary. The result of experiencing these structured patterns is that human learners develop accurate and flexible conceptual models. Through a preliminary study, we show that using insights from Variation Theory improves humans' ability to predict robot motions: accuracy in predicting motions increases from 52.4% using a familiarization-based strategy to 70.2% using a Variation-based strategy. Applying Variation Theory especially increases the human's accuracy in predicting robot motions in novel settings (increasing from 50.0% to 72.4% accuracy).

1 INTRODUCTION

Many works in human-robot interaction focus on improving robots' abilities to model human behaviors [6, 8, 10, 14]. The inverse problem is equally important: building accurate mental models is essential in human-interactive robot learning settings, and is especially helpful for enabling people to work safely in close physical proximity to robots. By understanding the robot's exhibited behaviors and how these differ from the human's intentions and preferences, a human can become a more effective robot teacher [2, 11]. How can humans learn to predict robot motions? We use theories of human concept learning—particularly, the Variation Theory of Learning [9]—to teach humans to form these conceptual models. Through a preliminary study, we show that applying the insights of Variation Theory increases humans' accuracy in predicting robot motions especially in novel settings. Without applying Variation Theory, humans are mostly restricted to collaborating with robots which exhibit predictable and not purely functional motions [4]. By applying the insights of Variation Theory, we find people are better able to learn to predict even functional and non-intuitive robot motions.

2 BACKGROUND & RELATED WORK

Dragan and Srinivasa studied how people learn to predict robot motion [3]: specifically, they studied how familiarizing people to robot motions by showing multiple examples of the robot moving could help people predict how the robot would move in the future, either in previously-experienced or in novel settings. In particular, this work compared humans' ability to learn to predict natural and unnatural motions. Natural motions are designed to mimic human motions, while unnatural motions are designed to seem counter-intuitive and artificial. Intuitively, Dragan and Srinivasa found that

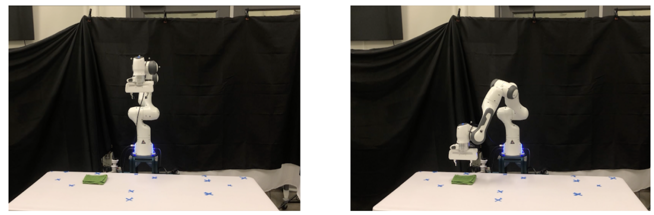

Figure 1: Participants watch videos of the robot moving from start (left) to end (right) positions to learn to predict the robot's motions.

adding a process of familiarization significantly increased humans' accuracy in predicting the natural robot motions compared to a baseline without familiarization. However, as the motion became less human-like, familiarization ceased to improve humans' ability to predict the robot's motion. We build on this work, and ask: can we teach people to better predict unnatural motions, too?

People may become better at predicting robot motions if they learn about the underlying concepts which govern a motion planner's expressed behaviors [1]. To support humans in learning these concepts, we look to the Variation Theory of Learning for guidance [9]. Variation Theory is a classroom-tested theory of human learning [7, 16]; it prescribes a sequence of patterns of variance and invariance to help humans comprehend and establish the bounds of concepts. The first step of Variation Theory involves *repetition*, or exposing the learner to the same concept repeated multiple times in the same setting. The familiarization intervention of Dragan and Srinivasa's study [3] can be interpreted in part as repetition.

The second step of applying Variation Theory is to use *contrast*, in which the human experiences an example of the concept with some aspects held constant alongside out-of-concept examples. Contrast is interesting for the HRI community as it indicates that to build human comprehension, people need to experience counter-examples of robot behaviors—which is rare in practice. The third step of Variation Theory is *generalization*, or exposing the learner to the same concept with varied superficial details. This generalization step is also embedded in Dragan and Srinivasa's familiarization process [3]. Variation Theory has numerous insights and recommendations, but we restrict our analysis to studying the benefits of including the second prescribed—and infrequently incorporated—step of contrast.

3 HYPOTHESES

We formulate hypotheses about how the addition of repetition and contrast improve people's ability to learn to predict robot motions.

- *H1: Experiencing contrast when learning about robot motions improves people's ability to subsequently predict robot motions when compared to a familiarization baseline.* This first hypothesis concerns people's ability to predict future robot motions in all settings: both those previously experienced as well as novel settings.

- *H2: Experiencing contrast when learning about robot motions improves people's ability to subsequently predict robot motions **in novel settings** when compared to a familiarization baseline.* Variation Theory is especially helpful for creating flexible conceptual models supporting generalization of knowledge to new settings. We thus focus on whether people can predict robot motions in novel settings.

4 METHOD

We test humans' ability to predict robot motions, following the protocol of Dragan and Srinivasa [3]. We created three motion planners: one which generates natural motion, and two which generate unnatural motions. For natural motion, we used CHOMP with the standard cost function [13]. For the first unnatural planner, we determined the desired end position using the natural motion planner. We rotated each joint sequentially to its end position and moved joints in order of distance from the end effector—i.e., the "wrist" joint moves first. For the second unnatural motion planner, we inverted the joint order—i.e., the "shoulder" moves first. Since the wrist-first motion controller generates the least intuitive motions, we believe it is the hardest for humans to predict. We aim to support humans in learning to predict the motions of the wrist-first motion planner.

We used a Franka Panda positioned behind a table for this study (Fig. 1). We discretized the robot's reachable space into 28 target locations: 14 on the table and 14 on an elevated plane above the table. We test humans' accuracy in predicting robot motions by asking them, for each scenario, to choose between three videos showing different motions (generated by the three motion planners) which start and end in the same position. The human's objective is to identify which motions are generated by the "correct" motion planner, which exhibits the unnatural wrist-first motions. In the familiarization and contrast interventions, two of these scenarios are designated as Level 1: these start and end in positions which the human experienced. Four other scenarios test their ability to predict motions in novel settings. These scenarios are either designated as Level 2 when the start position is novel or Level 3 when the end position is novel. Every participant answered the same accuracy test consisting of 6 scenarios, 2 for each level. At the end of the study, participants were asked to briefly describe the robot's strategy for how it moves towards the goal.

In this study, we compare the effect of three teaching interventions prior to this accuracy test. These consist of a baseline (i.e., no teaching), familiarization (showing only correct motions) and contrast (showing both correct and incorrect motions). In the familiarization and contrast interventions, participants watched 14 videos of robot

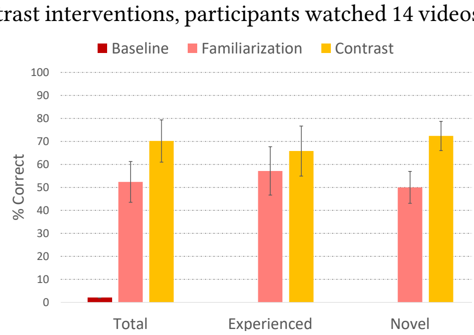

Figure 2: Accuracy increased from 2.0% in the baseline to 52.4% with familiarization to 70.2% with contrast; these differences are further evaluated in novel and experienced settings. Error is standard error.

motions. With familiarization, each video corresponded to the robot moving from a fixed start location to a different target location (i.e., 14 target locations in total). With contrast, the human saw side-by-side videos of both unnatural motion controllers with one indicated as correct. To keep the number of videos the same, each human saw only 7 target locations, equivalent to the first 7 familiarization targets, since they saw two videos per location under contrast.

We ran this IRB-approved study on Prolific with subjects who were 18 or older [12]. In the baseline study, the subjects watched 18 videos; this baseline study took approximately 15 minutes, and participants were paid \$3.75USD. The familiarization and contrast studies each took around 30 minutes with 32 videos, and subjects were paid \$7.50USD. Across all conditions, 8 subjects failed attention checks; we removed these subjects' data. 18 subjects were included in the baseline condition, 19 subjects were included in the familiarization condition, and 21 subjects were included in the contrast condition.

5 RESULTS

With no teaching, users identified robot motions with 2.0% accuracy. This low accuracy is expected; 89.9% of responses incorrectly selected the natural motion. With familiarization, users' accuracy rate increased to 52.4%, reaffirming Dragan and Srinivasa's findings [3]. With contrast, users' accuracy rate further increased to 70.2%. A two-sample T-Test indicates we cannot reject the null hypothesis that the mean accuracies between familiarization and contrast are the same: $t(38)=-1.43$, $p=0.08$. While not statistically significant, the large accuracy improvement suggests the contrast intervention has promise. This accuracy improvement provides trending support for hypothesis 1: *Experiencing contrast when learning about robot motions improves people's ability to predict robot motions.*

Across settings, adding contrast increased accuracy rates, with an especially pronounced difference in novel settings. For settings previously experienced during teaching (Level 1), participants in the familiarization variant were 57.1% accurate compared to 65.8% with contrast. For novel settings (Levels 2 and 3), the contrast participants had higher accuracy (72.4%) than familiarization participants (50.0%). A two-sample T-Test indicates that we can reject the null hypothesis: $t(78)=-2.35$, $p=0.01$. This accuracy improvement may be a consequence of participants using contrast to establish concept boundaries, improving their ability to infer constraints of motion in unseen dimensions of variation. We find support for hypothesis 2: *Experiencing contrast when learning about robot motions improves people's ability to predict robot motions in novel settings.*

6 DISCUSSION & FUTURE WORK

When asked to describe the robot's strategy for how it moved to the goal, one user mentioned they "mimicked the robot's movements" to understand how it moved, which implies they may be using a different form of concept learning, analogical transfer, for comprehension [1, 5, 15]. Future work could further analyze how to best support analogical transfer for learning, alongside or instead of variation. It could also assess whether applying contrast remains helpful for robots with different morphologies. Another direction could be to incorporate the two further steps of Variation Theory (generalization and fusion) to assess whether they equally improve humans' ability to learn the concepts underlying robot motions.

REFERENCES

[1] Serena Booth, Sanjana Sharma, Sarah Chung, Julie Shah, and Elena L Glassman. 2022. Revisiting human-robot teaching and learning through the lens of human concept learning. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 147–156.

[2] Tathagata Chakraborti, Sarath Sreedharan, Sachin Grover, and Subbarao Kambhampati. 2019. Plan explanations as model reconciliation–an empirical study. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Ieee, 258–266.

[3] Anca Dragan and Siddhartha Srinivasa. 2014. Familiarization to robot motion. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. 366–373.

[4] Anca D. Dragan, Shira Bauman, Jodi Forlizzi, and Siddhartha S. Srinivasa. 2015. Effects of Robot Motion on Human-Robot Collaboration. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 51–58.

[5] Dedre Gentner and Linsey A Smith. 2013. Analogical learning and reasoning. *The Oxford handbook of cognitive psychology* (2013), 668–681.

[6] Guy Hoffman and Cynthia Breazeal. 2004. Collaboration in human-robot teams. In *AIAA 1st intelligent systems technical conference*. 6434.

[7] Angelika Kullberg, Ulla Runesson Kempe, and Ference Marton. 2017. What is made possible to learn when using the variation theory of learning in teaching mathematics? *Zdm* 49, 4 (2017), 559–569.

[8] Przemyslaw A Lasota and Julie A Shah. 2015. Analyzing the effects of human-aware motion planning on close-proximity human–robot collaboration. *Human factors* 57, 1 (2015), 21–33.

[9] Ference Marton. 2014. *Necessary conditions of learning*. Routledge.

[10] Stefanos Nikolaidis, Ramya Ramakrishnan, Keren Gu, and Julie Shah. 2015. Efficient model learning from joint-action demonstrations for human-robot collaborative tasks. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 189–196.

[11] Stefanos Nikolaidis and Julie Shah. 2012. Human-robot teaming using shared mental models. *ACM/IEEE HRI* (2012).

[12] Stefan Palan and Christian Schitter. 2018. Prolific. ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance* 17 (2018), 22–27.

[13] Nathan Ratliff, Matt Zucker, J Andrew Bagnell, and Siddhartha Srinivasa. 2009. CHOMP: Gradient optimization techniques for efficient motion planning. In *2009 IEEE International Conference on Robotics and Automation*. IEEE, 489–494.

[14] Vaibhav V. Unhelkar, Przemyslaw A. Lasota, Quirin Tyroller, Rares-Darius Buhai, Laurie Marceau, Barbara Deml, and Julie A. Shah. 2018. Human-Aware Robotic Assistant for Collaborative Assembly: Integrating Human Motion Prediction With Planning in Time. *IEEE Robotics and Automation Letters* 3, 3 (2018), 2394–2401. https://doi.org/10.1109/LRA.2018.2812906

[15] Patrick H Winston. 1980. Learning and reasoning by analogy. *Commun. ACM* 23, 12 (1980), 689–703.

[16] Ngai-Ying Wong, Ference Marton, Ka-Ming Wong, and Chi-Chung Lam. 2002. The lived space of mathematics learning. *The Journal of Mathematical Behavior* 21, 1 (2002), 25–47.